Jun 6th, 8:01 AM - 9:00 AM

Automated Argument Analysis – Comment on: Mizrahi & Dickinson: "Argumentation in Philosophical Practice: An Empirical Study"

Christoph Lumer
University of Siena

Automated Argument Analysis – Comment on: Mizrahi & Dickinson: "Argumentation in Philosophical Practice: An Empirical Study"

CHRISTOPH LUMER

DISPOC (Dipartimento di Scienze Sociali, Politiche e Cognitive)

University of Siena

Via Roma, 56

I-53100 Siena

Italy

lumer@unisi.it

1. Mizrahi & Dickinson's analysis of philosophical argument types

In their empirical study "Argumentation in Philosophical Practice: An Empirical Study" (Mizrahi & Dickinson 2020 – abbreviated in the following as: M&D), Moti Mizrahi and Michael Dickinson examine the change in philosophical types of argument over the last 150 years (1867-2017). They searched an extensive corpus of texts from JSTOR with 435,703 (M&D 1; 6; 12) philosophical publications for argument indicators for three types of arguments, namely deductive, "inductive" and abductive ("i.e. arguments in which the conclusion is supposed to be the best explanation for some phenomenon" (Govier 2010, pp. 298-302, quoted in M&D 3)). More precisely, they used 36 argument indicator pairs for each of these three types of arguments, with one component indicating an inferential relationship – e.g. "therefore", "hence" (nine in total) – and the other indicating the type of argument – e.g. "necessarily", "probably", "best explain" (four per argumentation type) –, so that indicator pairs such as "therefore necessarily", "therefore certainly", "hence necessarily", "so definitely" were used to find deductive arguments, indicator pairs such as "therefore probably", "consequently likely" were used to detect inductive arguments, and indicator pairs such as "therefore best explain" and "hence make sense of" were used to determine abductive arguments (M&D 3-5). (There could be up to ten other words between the two components. (M&D 6))

Mizrahi & Dickinson summarize their results with the following hypotheses:

H1: "Deductive arguments were the most common type of argument in philosophy until the end of the twentieth century: [H1.1.] significantly more common than abductive, [1.2] but not inductive, arguments." (M&D 1; 2; similar 8; 10; 11)

H2: "Then, around 2008 a shift in methodology occurred, and inductive arguments took over as the most common type of argument." (M&D 1; 2-3; similar 11)

H3: "In addition, abductive arguments are becoming increasingly more popular in philosophy." (M&D 1; 3; similar 11)

H2r.&3r: "Our results suggest that deductive arguments are giving way to not only inductive arguments but also abductive arguments in philosophical practice." (M&D 1; 3; similar 12)

In addition to the results of the statistical analysis, they also provide graphs (three of them, for up to three words, up to six and up to ten words distance between the two indicator components) with the proportions of texts that, according to the indicators, contained at least one deductive, one inductive, and one abductive argument, respectively, in the total number of philosophical publications of one of the 150 years. These graphs show the increase of all three types of argument over time - since about 1950, the increase in a smoothed curve would probably be almost linear for all three types of argument - and they also show that the deductive arguments have been overtaken by the inductive ones as claimed in H2. (M&D 7; 9; 10) [1]

2. A positive assessment and the course of the further discussion

The effort of the study, the number of texts examined and the degree of statistical elaboration are impressive. The amount of work must have been enormous. And I acknowledge the pioneering work. I also appreciate the great clarity in the presentation of the method used.

I would like to place before everything else my overall assessment of the hypotheses put forward by Mizrahi & Dickinson. I consider the following of their hypotheses to be plausible:

H1 (deductive arguments are dominant until 2000),

H1.1 (deductive arguments are significantly more frequent than abductive ones),

H3 (since about 2008 there are significantly more abductive arguments),

H2r&3r (the relation of inductive and abductive to deductive arguments improves in favor of inductive and abductive)

These hypotheses also coincide with my experiences with the increase in experimental philosophy since about 2008, the strengthening of methodological naturalism and also the discussion of the corresponding studies by authors who do not share these approaches.

The following hypotheses of Mizrahi and Dickinson, on the other hand, are not sufficiently supported by the study:

H1.2 (until 2000 deductive arguments were not significantly more frequent than inductive ones),

H2 (from 2000 on, inductive arguments are more frequent than deductive ones).

The methods used by Mizrahi and Dickinson to prove these theses are, as I will show in a moment, so uncertain, so weak indicators for the actual numerical relationships, that the - even according to their results - small lead of inductive versus deductive (according to the 10-

---

[1] With this reading of the curves, however, I am not completely sure. The authors describe the relations only as e.g. "Ratios of philosophy publications in the JSTORE corpus with deductive, inductive, and abductive arguments" (M&D 7; 9; 10), without specifying what the denominator of this fraction is. However, the most plausible interpretation is: the total number of philosophical publications in a given year contained in the JSTORE database. And I don't really see any other reasonable interpretation.

word distance measurement) in the peak year 2009 of about 31.8% inductive to 30.3% deductive (read from M&D 10, Figure 3) is by no means sufficient to prove the predominance of inductive arguments over deductive ones since 2008 or even the merely insignificant lead of deductive arguments over inductive ones until 2008. The validity of the method of analysis used is, as I will show soon, so imprecise that the numerical result of Mizrahi & Dickinson is, instead of the numerical ratio of inductive to deductive of $1/0.953$ ($=1/(30.3\%/31.8\%)$) in the year 2009 in favor of inductive in my estimation, compatible with an actual numerical ratio of $1/0.2$ or conversely of $1/5$.

In the next section I would like to substantiate this assessment of the merely very imprecise indicator function of the method used by Mizrahi &Dickinson by a more detailed analysis. Secondly, I will address some further weaknesses of the method used. However, I would like to repeat here once again my recognition of the pioneering work. In this respect, the following criticism is to be understood as a constructive indication for future studies. - Section 4 draws a more general theoretical conclusion from this methodological critique on the possibilities of computerized, automatic argument analysis, especially the classification of argument types. - Finally, Section 5 discusses the philosophical significance of Mizrahi & Dickinson's analysis and similar empirical studies in general: What does an empirical philosophy of this kind offer?

3. A critical analysis of the method used by Mizrahi & Dickinson

In the following I discuss some aspects of the method used by Mizrahi & Dickinson.

Text selection, the basic set of analyzed texts:

1. Language: The authors summarize their results as follows: "Overall, the results of our empirical study suggest that deductive arguments were the most common type of argument in philosophy until the end of the twentieth century ...". (M&D 2) And on representativeness they write: "Since our results were obtained from a survey of a large corpus of philosophical texts mined from the JSTOR database (n = 435,703), we can be quite confident that they are representative of philosophical practice." (M&D 12). Now, as indicators for the different types of arguments, only English expressions were used, such as "therefore necessarily", "consequently improbable". Does this mean that arguments in non-English languages are not philosophical arguments per se? (The fact that the non-English languages are underrepresented in JSTOR is problematic enough. That they are completely ignored here, however, is dramatic.) This - presumably unintentional and simply naive - Anglo-Saxon chauvinism may also lead to serious distortions of the quantitative results, if - and there is some evidence for this - the "inductive wave" (see below) in non-English philosophical publications, for example, is by far not as large as in English-speaking ones.

2. Criterion for a philosophical text: By what criteria was it determined what a philosophical text is? Unfortunately, the authors do not say. Is the criterion that it was published in a journal to be considered philosophical? What then applies to journals such as 'American Journal of Theology and Philosophy', 'Philosophy and Rhetoric'? The articles published there are certainly not all philosophical. And do not, conversely, philosophical articles also appear in the general journals of the surrounding disciplines or interdisciplinary journals such as 'Philosophy and Economics' or 'Philosophical Psychology'? What applies to monographs and anthologies? The texts cover a period of 150 years, the oldest of which date from 1867. Can we be sure that the texts in "philosophical" journals were always philosophical? In the past,

philosophical texts also included texts that we would now quite clearly regard as educational or psychological. Different criteria for the philosophical character of a text naturally change the results considerably. Presumably, the more liberal the admission, the greater the proportion of "inductive" arguments.

Different types of arguments: The authors distinguish between deductive, inductive and abductive arguments. What deductive arguments are is probably fairly clear in the argumentation theory community. The other two categories are more problematic.

Inductive arguments: What "inductive arguments" are supposed to be is not defined by the authors. But it is clear from the context that for them inductive arguments are those arguments in which probabilistic words such as "probably", "likely", "improbable", "plausible" occur in the thesis or in the description of the inferential relationship (M&D 3). This criterion is rather un-theoretical and leads to a completely heterogeneous set of arguments. 1. Apart from probabilistic arguments, many of the actual probabilistic indicators also occur in all non-probabilistic arguments, especially to express uncertainty, in particular in practical arguments, but even in deductive ones: "Probably the thesis follows from it", if the final relationship is complicated and has not been formally verified. One can also express a thesis probabilistically out of laziness, if one does not take the trouble to give a more rigorous proof. 2. The set of really probabilistic arguments that want to establish degrees of probability is very heterogeneous. This might have been worth the differentiation. For example, epistemologically reconstructed, even the genesis of knowledge arguments such as arguments from expert opinion, arguments from position to know, source references, which are usually understood as a separate group, belong to the probabilistic arguments. (Lumer 2011a) 3. From probabilism does not follow that these are empirical arguments in the usual sense - as the authors seem to assume, however - namely that from individual observations a general connection is inferred or from there back to an individual case. Example: "We have not seen a real proof of p. Therefore, probably we cannot say that p."

Abductive arguments: 1. Arguments that provide the best explanation can be very different in structure, ranging from simple backward conclusions: "Where smoke rises, there is probably a fire. There is smoke rising. Underneath is probably a fire" to fully developed interpretive arguments with alternative hypothetical explanations of known facts and the Bayesian determination of the probability of these explanations; in the latter case the "best explanation" is interpreted as the most probable. The former are indicative argumentations with simple statistical (inferences) with a special content. The latter are interpretive arguments with a very high complexity. (Cf. Lumer 1990, pp. 221-246) This structural difference is neglected by Mizrahi & Dickinson's classification. But perhaps this is not so serious, and this differentiation can also be omitted. 2. But in any case, arguments to the best explanation, correctly analysed, also belong to the probabilistic arguments, i.e. according to the terminology of Mizrahi and Dickinson to the inductive arguments. Accordingly, all their indicators of "inductive" argumentations can also indicate abductive arguments. However, if abductive arguments are treated as a separate type of argument, then other prominent subtypes of probabilistic arguments would also have to be treated as separate types of argument.

Elementary and molecular arguments: 1. Molecular or complex arguments are composed of several elementary arguments to form a tree structure, whereby the arguments at the ends of the tree establish the premises for the arguments further towards the trunk. Such complex arguments often involve elementary arguments of different types; and yet molecular argumentation can often be assigned to a particular type of argument (see Lumer 2011a). According to the method of Mizrahi and Dickinson, all elements of a molecular argumentation would be included as separate arguments. However, this contradicts the usual

approach, according to which complex arguments are regarded as one argument. 2. The method of Mizrahi & Dickinson can then also lead to a distortion of the numerical relations between the types of arguments, for example if one type of argument is used more often than others in a subordinate function. This is somewhat the case with practical arguments that often use probabilistic arguments to prove their consequential assumptions.

Missing argument types: The triple list of deductive, inductive, abductive arguments does not contain a number of other arguments, which are, however, philosophically relevant:

- practical arguments (for value judgements with the listing of advantages and disadvantages of the value object);

- genuine statistical arguments (with information on correlations and levels of significance; these arguments usually do not speak of "probability");

- intuitionistic arguments (which use certain intuitions as the essential premise);

- genesis of knowledge arguments (which refer to the verification of the thesis by another person: argument from from expert opinion, reference to historical sources, arguments from testimony).

While practical arguments are definitely a separate type of argument alongside deductive and probabilistic ones, descriptive statistical and intuitionistic arguments are mostly deductive and genesis of knowledge arguments are probabilistic (Lumer 2011a). But this is highly controversial. And it should have been addressed.

Empirical determination of the argument types - errors in the recognition of argument types: In the empirical identification of argument types within the selected text corpus, basically two types of errors can occur: *1. false positive, misclassifications*, in which a piece of text is classified as argument of type T although the piece of text is not an argument of type T, and *2. false negatives*, it is not recognized that an argument of type T is present. A special case are false positives, which are also false negatives, i.e. if an argument is classified as being of a different type than it is.

False positives: 1. Some of the argument indicators used by Mizrahi & Dickinson may, despite their first appearance, indicate something different than the assumed type of argumentation: e.g. "accordingly ..." may indicate the execution of a plan or the compliance with criteria: "Accordingly, I definitely took her at her word"; "Accordingly necessarily we first had to check whether...". 2. The indicators may be taken from a literal quotation or a paraphrased text, from which the argumer however disassociates himself: "The author infers that certainly p. But ...". 3. The indicator may be qualified or even negated by the context: "From this we cannot infer that necessarily p holds"; "it seems therefore necessarily that p; however ...".

False negatives: 1. Most arguments use only one indicator and not two. The inference indicator can be missing and only the strength indicator can be used: "p1, ..., pn. Necessarily t." The converse case is much more frequent, however, namely that the inference indicator is present, but the strength indicator or abduction indicator is omitted. 2. Instead of the mentioned indicators, wording variants of them are used: "infer / proves with necessity" (instead of "infer / proves necessarily"); "we have proved definitely" (instead of "proves definitely"). 3. Completely different indicators are used than those that appear in the Mizrahi & Dickinson list, including also those that the authors themselves have mentioned before, especially inferential indicators: "because of p q (holds)"; "because p we can be sure that q", "this makes me think", "with this I have shown that", "this makes it probable", "this implies", "since (then)", "then also holds", "presumably then", "this is strong / weak evidence for", "the

most probable explanation for this is". 4. Not the usual inference indicators are used, but completely free descriptions like: "my thesis is ...; and the reasons for it are", "why should it be the case that? Well, consider". 5. No explicit argument indicators are used at all, which in English, unlike in other European languages, occurs relatively frequently, but the argumentative connection results only from the content. 6. The distance between inference indicator and qualifier is even greater than 10 words.

Much more false negatives than false positives: Mizrahi & Dickinson have chosen extra strong indicators to exclude false positives ("In order to make sure that our indicators for argument types [...] aktucally indicate arguments in the corpus, we anchored them to argument indicators, such as 'therefore' and 'hence'." (M&D 4)). And indeed, the resulting combinations are already relatively safe signs for the respective assumed arguments - even if, as analyzed above, false positives remain. The price for this relatively high degree of certainty in the exclusion of false positives by an indicator overkill is, conversely, the high number of false negatives: If the hurdles for "recognition" as an argument of a certain type are set so high, then they are not met by very many arguments that definitely belong to this type. Many arguments do not use this double indicator, but only a single one (and, in addition, others than specified or none at all, see below). I estimate that with the argument indicators used by Mizrahi & Dickinson at most 10% of the actual arguments of the type in question are covered.[2]

Due to the high proportion of false negatives very large uncertainty, strong distortions of the relations are possible: Mizrahi & Dickinson may have chosen the strong indicators with the hidden idea that the numerical relations among the captured arguments of the three types are the same as among all arguments of these types in their text corpus. But with the high number of false negatives, a relatively small difference in the rates of false negatives for the three argument types makes a great deal of difference to the overall ratio. For example, if the proportion of false negatives in the deductive arguments is 90%, but 95% in the "inductive" arguments, then the measured ratio can change from 1:1 to 1:2.

No check of the reliability of the test instrument: What is most surprising in this situation is that the authors did not check the reliability of their test instrument (i.e. the automated search by computer for the given argument indicators) at all and did not calibrate this instrument. Such a reliability check can be carried out relatively easily by having the test instrument examine some texts, from which all the arguments contained in them have been identified and classified by reading and analysis, and comparing the two results. - I started such a test with an analysis of Derek Parfits article: "Equality and Priority" (1997). The result confirms the criticism that has been made so far from a theoretical standpoint: The search for the 108 indicators of Mizrahi & Dickinson had exactly one result: on page 208 the indicator words "necessarily ... therefore" were found with a sufficiently small distance. Unfortunately, this was a false positive: Parfit only explains the application of the telic egalitarian criterion here; it is therefore, if at all, something like a quoted argument, not one of Parfit's himself. - Parfit's text is not particularly argumentative but explanatory: Parfit describes a new, prioritarian criterion for distributive justice and differentiates it from egalitarian ones. Nevertheless, the text contains a whole range of arguments which cannot be found with the method of Mizrahi & Dickinson, i.e. false negatives: 2 practical arguments (Parfit 1997, pp. 210-211) (in them objections are introduced with the words "objections" ... In addition, the text contains a whole range of deductive arguments, e.g. on page 212 with the argumentation indicator "since" or

---

2   This, my estimate may seem prima facie to be completely wrong if one considers e.g. that 2009, the year with the most hits, the "inductive" arguments are 31.8%, so that there would be room for a tripling (95.4%) but not for a tenfold increase. But this appearance is deceptive: the number of hits does not indicate the proportion of argumentations, but the proportion of texts with this type of argumentation, whether this text contains one or 100 arguments of this type.

on pages 215-216 with the argumentation indicator "then".No check of the reliability of the test instrument: What is most surprising in this situation is that the authors did not check the reliability of their test instrument (i.e. the automated search by computer for the given argument indicators) at all and did not calibrate this instrument. Such a reliability check can be carried out relatively easily by having the test instrument examine some texts, from which all the arguments contained in them have been identified and classified by reading and analysis, and comparing the two results. - I started such a test with an analysis of Derek Parfits article: "Equality and Priority" (1997). The result confirms the criticism that has been made so far from a theoretical standpoint: The search for the 108 indicators of Mizrahi & Dickinson had exactly one result: on page 208 the indicator words "necessarily ... therefore" were found with a sufficiently small distance. Unfortunately, this was a false positive: Parfit only explains the application of the telic egalitarian criterion here; it is therefore, if at all, something like a quoted argument, not one of Parfit's himself. - Parfit's text is not particularly argumentative but explanatory: Parfit describes a new, prioritarian criterion for distributive justice and differentiates it from egalitarian ones. Nevertheless, the text contains a whole range of arguments which cannot be found with the method of Mizrahi & Dickinson, i.e. false negatives: 2 practical arguments (Parfit 1997, pp. 210-211) (in them objections are introduced with the words "objections" ... In addition, the text contains a whole range of deductive arguments, e.g. on page 212 with the argumentation indicator "since" or on pages 215-216 with the argumentation indicator "then".

The empirical determination of argumentation types – methodological distortion of quantities: What the authors count with their method is not the number of deductive, inductive and abductive arguments in the philosophical publications of one year each, but with regard to the publications of one year: the sums of all numbers of the types of deductive, inductive or abductive argument indicators occurring in the individual publications (M&D 6).[3] Thus, behind one unit of the count there can be one argument with a certain indicator, but also 20 or more. In order to determine the weight of e.g. deductive arguments, however, the number of arguments should actually be considered. Of course both quantities might correlate optimally, but this would be coincidence. Distortions also of the relations of the frequencies of the individual argument types can result from the following: 1. Authors use a very limited number of indicators for a certain argument type, while they vary the indicator very much for another argument type. 2. Behind the individual indicator types there are very different numbers of indicator occurrences and thus of arguments. (This leads in particular to distortions if these different frequencies are not compensated by an increase in the richness of variation of the indicator expressions. For example, the author uses three indicator types each for deductive arguments and for "inductive" arguments; but in the text there are 20 deductive and five "inductive" argumentations).

A methodological blemish is that the article does not contain certain information that would have been helpful for the classification of the result: How many texts contained arguments at all and how many arguments of what kind? Which indicators of argumentation occurred how often?

4. Limits of automatic argumentation analysis

---

3   This means for example: In the year n three philosophical articles have been published; the first one does not contain any deductive argument; the second one contains ten deductive arguments for whose indication three different indicators from the list of Mizrahi & Dickinson are used, the third one contains 20 deductive argumentations for whose indicatioin only two different argument indicators are used. Then the year n is led with (3+2=) five deductive arguments.

Mizrahi & Dickinson have used a relatively simple criterion for the automatic identification of arguments and for the analysis of the type of argument, whose considerable methodological weaknesses and quantitative inaccuracies have been sufficiently explained in the previous section. Can the weaknesses described be eliminated with better automated procedures? Perhaps sometime, but there is certainly a long way to go until then.

The problems begin with the fact that up to now there is not even an approximate agreement on the existing types of argumentation, even on a theoretical level. Of course, this does not exclude the possibility of trying to use automatic procedures in cases where there is broad agreement. But the problems mentioned above exist to a not inconsiderable extent for any automatic analysis methods: There is not only a wealth of standardized argument indicators, but also non-standardized, completely individually formulated ones; or the argument indicators are missing altogether. In addition, most of the argument indicators do not indicate the type of argument; the arguing persons themselves are usually too ignorant in theory to be able to determine it consciously. If all these indicators are not sufficiently informative, one has to refer back to the definitions of the argument types themselves and to clues that are based on the original differences (i.e. not on the indicators that were deliberately added). According to the epistemological approach (which I advocate) in the theory of argumentation, the different types of arguments are distinguished according to the epistemological principle on which they are based and on which the structure of the argument is then also oriented (Lumer 2011a). The deductive arguments, for example, try to show that the conditions of the deductive epistemological principle are fulfilled for the respective thesis, namely: 'A proposition p is true if it is deductively implied by true premises'. If one considers the mode of action of deductive inferences, one can, for example, gain from this the indication that certain content-related terms that occur in the conclusion must already be found in the premises (Lumer 2019). However, this only applies to ideal deductive arguments. If certain premises are missing in non-ideal, enthymematic deductive arguments, or if the singular terms for the same subject or even the predicate expressions for the same predicate vary, then the repetition of the content-related terms is difficult to recognize and only after additional operations, especially the premise supplementation. Now, however, the adequate premise supplementation in deductive arguments is, as all those who have ever tried it know, a particularly laborious business. In the case of candidates for possibly suitable additions, it must be checked, especially in between, whether the examined addition is benevolent, i.e. whether it fulfils the conditions of validity and adequacy for the argument in question, especially whether it is true and accepted by the arguer. Computer programs will continue to have difficulty in answering such questions for a long time to come.

What I want to show with this small excursus is that until a halfway reliable automatic recognition of argument types - not to speak of the automatic evaluation of argumentations - many problems still have to be solved, which by far exceed the capabilities of today's machine analysis methods like text recognition. Until then, traditional "manual" analysis is the only method that is in principle suitable for the reasonably reliable identification of argumentation types and for the evaluation of argumentations - also and especially because it is very cumbersome for complex argumentations.

5. The value of empirical philosophy, especially of automatic analysis of philosophical arguments

What has been gained with the insights acquired by Mizrahi & Dickinson? The authors attribute the relative increase of "inductive" arguments to a methodological "shift in methodology" (D&M 1; 2; 11). At the end they specify this and suggest that it is a shift from an a priori philosophy, for which deductive arguments stand and which derives necessary truths in the armchair, to an empirically oriented philosophy, especially experimental philosophy (D&M 11-12).

This would be an important insight in the history of philosophy. But can this be inferred from the findings? For one thing, it is still open whether not at least a considerable part of the – all in all not extremely blatant – changes in quantitative relations is due to a trend away from absolutist to more cautious formulations or to the use of more standardised argument indicators, especially in the "inductive" arguments (in the course of a standardisation of writing style, especially in the "inductive camp"). For another thing, however, if the result is really due to a methodological change, some differentiations are necessary. 1. The change would not be a change in "philosophy", but perhaps a relative increase in those philosophers who also use empirical data. 2. deductive arguments are not only used in a priori philosophy, but more or less ubiquitously, even in empirical studies. They even occur within practical arguments. 3. Conversely, the use of probabilistic terms does not mean that empirical data are involved at all – one can, as I said, use them to express any kind of uncertainty. 4. Nor does the use of empirical information mean that this use is part of an empirical philosophy in the usual sense, i.e. which primarily wants to gain empirical information. Empirical information is also needed in all applied philosophies, especially in applied ethics and applied epistemology, but often also, for example, in philosophies that are interested in developing good instruments, from political philosophy to theory of action (e.g. against weakness of will) to rational decision theory. If the use of relatively more "inductive" argument indicators does indeed indicate a methodological change, then it remains to be seen which one: a change towards methodological naturalism in general or towards empirical philosophy in particular, towards an increase in applied philosophies or instrumental philosophy. 5. In the historical explanation offered by Mizrahi & Dickinson, another philosophical trend is lost, namely the strong spread of methodological intuitionism since the 1970s, especially, but by far not only, in normative and applied ethics. Indeed, this trend is probably largely a change within the "deductivist camp", since intuitions are often used within deductive arguments; and it can only be discovered through a content analysis, namely that authors' intuitions are used as premises. – In short, a more accurate documentation of historical philosophical trends would have required detailed analyses of the content of philosophical texts.

I was astonished that Mizrahi & Dickinson do not at all refer to a much more significant result of their study which, if I have interpreted this correctly, can be read from the enclosed graphs: the strong historical increase in argumentativity of philosophical texts: The share of the examined philosophical texts with deductive arguments (measured, of course, according to the method of Mizrahi & Dickinson) increases (according to the graph for up to ten word distance between the two partial indicators) from about 2-4% in the period 1890-1933 to 30.3% in the peak year 2009 and that of the texts with inductive arguments from 1-2.5% in the period 1890-1933 to 31.8% in the peak year 2009. (M&D 10; my reading from their figure, C.L.) Of course, after the criticism expressed above I would not trust the exact data at all. But the trend in the indicators is so blatant and probably highly significant that it will also indicate a real trend – probably with a different actual scope. The fact that even today about 70% of the English philosophical texts do not contain deductive and inductive arguments and also about 70% do not contain inductive ones is probably wrong again and is due to the false negatives. This proportion should be empirically much smaller, how small, the study does not really reveal. Well, if the historical increase of deductive arguments is so dramatic, then *this* is the main result of the study; and, of course, it would be a very welcome result. If professional philosophers now additionally argue inductively where deductive arguments

cannot be used any more, then that is even better. The fact that the quantitative primacy is scarcely reversed from deductive to inductive arguments is then a rather minor, perhaps even irrelevant result: It is not simply the case that the inductive arguments "suppress" the deductive ones (as e.g. species of living beings do), but that they are additionally introduced where deductive arguments can no longer be used. The explanation of the main trend – strengthening of analytical philosophy?, loss of authority of philosophers and public demands for well-founded results? ... – would again require a content analysis.

How should these trends be evaluated? Mizrahi & Dickinson write several times that philosophers thought that deductive arguments were the best arguments. (M&D 1; 2; 11) And the authors tend to agree with this. (M&D 11) Accordingly, they should actually take a negative view of the advance of "inductive" arguments; de facto, however, they do not comment on them. In fact, the assessment of deductive argumentation as the best argumentation in this generality is wrong. Deductive arguments are the only certain arguments, and certainty is in the first place positive. But deductive arguments have only a small range because of their epistemic prerequisites; many important theses cannot be justified with them, e.g. those about the future cannot. If one wants to substantiate such theses, one must resort to other, weaker arguments. In this respect, the additional growth of "inductive" and abductive arguments would be prima facie a progress. A problem would be their increase, however, if this increase was due to the use of methods not suitable for philosophy. And this is unfortunately true in several cases, for example when one tries to solve philosophical questions with the help of (empirically surveyed) folk intuition.

So what has been gained from Mizrahi & Dickinson's study? For a direct philosophical-historical use, the empirical results are only realtively weak and very imprecise indications of the exact actual developments. Of course, they are not of direct normative significance for philosophical methodology; this would be a clear violation of Hume's law. My suspicion is: In order to gain important philosophical insights, we need to know the questions of philosophy and the kind of answers that are being sought (cf. Lumer 2020; 2011b). However, a clear conception of the philosophical questions is missing in a not inconsiderable part of empirical philosophy. It would therefore be important in general to concern ourselves more with the metaphilosophical question of what good philosophy is. And in philosophical argumentation theory we need more research on this basis on criteria for good arguments.

References

Govier, Trudy (2010). *A Practical Study of Argument*. Belmont, California: Wadsworth [1]1985. 7th revised ed. [7]2010.

Lumer, Christoph (1990). *Praktische Argumentationstheorie. Theoretische Grundlagen, praktische Begründung und Regeln wichtiger Argumentationsarten*. Braunschweig: Vieweg.

Lumer, Christoph (2011a). Argument Schemes – An Epistemological Approach. In: F. Zenker (Ed.), *Argumentation. Cognition and Community. Proceedings of the 9th International Conference of the Ontario Society for the Study of Argumentation (OSSA), May 18-22, 2011* (32 pp.). Windsor, Canada: University of Windsor. CD-ROM. Web-publication, URL = <http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1016&context=ossaarchive>.

Lumer, Christoph (2011b). Tre tipi di teorie filosofiche. In: R. Davies (Ed.), *Analisi. Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA)* (pp. 45-75), Milano / Udine: Mimesis.

Lumer, Christoph (2019). Recognizing Argument Types and Adding Missing Reasons. In: B.J. Garssen, D. Godden; G. Mitchell; J.H.M. Wagemans (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation (ISSA). [Amsterdam, July 3-6, 2018.]* (769-777), Amsterdam: Sic Sat. Web publication, URL = <http://cf.hum.uva.nl/issa/ISSA_2018_proceedings.pdf>, 5.4.19.

Lumer, Christoph (2020). A Theory of Philosophical Arguments. To appear in: *Proceedings of the 12th International Conference of the Ontario Society for the Study of Argumentation (OSSA), June 3-6, 2020* (37 pp.), Windsor, Canada: University of Windsor.

Mizrahi, Moti; Michael Dickinson (2020 = M&D). Argumentation in Philosophical Practice. An Empirical Study. To appear in: *Proceedings of the 12th International Conference of the Ontario Society for the Study of Argumentation (OSSA), June 3-6, 2020* (13 pp.), Windsor, Canada: University of Windsor.

Parfit, Derek (1997). Equality and Priority. *Ratio, New Series* 10, 202-221.