The Theory of Scientific Reasoning and Argumentation in Practice

Claudia SCHULZ [a], Jan KIESEWETTER [b], Michael SAILER [c], Elisabeth BAUER [c], Martin R. FISCHER [b], Frank FISCHER [c], Iryna GUREVYCH [a],

[a] *Ubiquitous Knowledge Processing (UKP) Lab, Technische Universität Darmstadt, Germany*

[b] *Institut für Didaktik und Ausbildungsforschung in der Medizin, Klinikum der Universität München, LMU München, Germany*

[c] *Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie, LMU München, Germany*

Motivation

Analysing human argumentative reasoning behaviour can be advantageous in many applications, such as automatic feedback on students' essays [10,9], the identification of most helpful or deceptive reviews [8,2], and aiding eRulemaking [7]. To analyse the argumentative process in text or speech, an understanding of what makes an argument or which steps are involved in argumentation is needed. Some works present their own definition of argument, whereas others try to apply existing formalisations of argumentation (see [5]) such as Toulmin's model of argument [11] or Freeman's theory of argumentation [4]. Arguably, using existing formalisation of argumentation is beneficial as it facilitates a common understanding when analysing argumentation. It furthermore provides an evaluation of the theory as to its applicability to, and representativeness of, human argumentation. In addition, and most importantly, it provides a bridge between the *formalisation* of argumentation and *actual human* argumentation.

The Theory of Scientific Reasoning and Argumentation (SRA)

Humans use argumentative reasoning not only for persuasion but also for problem-solving. One form of problem-solving crucial in many professions is *diagnosis*: physicians determine a patient's disease based on clinical tests, teachers recognise behavioural disorders in children based on observations, and engineers debug errors in machines or programs based on their analyses of log files or flight recorders. The analysis of the reasoning underlying diagnosis is thus of importance across disciplines for educational applications aiming to understand and improve students' diagnostic reasoning skills.

Building upon findings in education and psychology, Fischer et al. [3] propose a *theory of scientific reasoning and argumentation* (SRA). In contrast to formalisations of argumentation in terms of components such as premise and conclusion used in the context of persuasion, SRA formalises *epistemic activities* involved in problem-solving: problem identification, questioning, hypothesis generation, construction and redesign of artefacts, evidence generation, evidence evaluation, drawing conclusions, and communicating and scrutinising.

Bridging the gap between theory and human argumentation, we choose SRA to analyse students' argumentation when diagnosing. This provides a unified theory to study argumentative reasoning when diagnosing in different disciplines. We here focus on teacher and medical education.

Analysing Argumentation in Diagnostic Reasoning Texts with SRA

To simulate professional diagnosis, various professional scenarios are outlined to the students, detailing both relevant and irrelevant information about a virtual patient (medicine) or pupil (teacher education). The students' task in each scenario is to decide on a diagnosis and to then write an explanation on how they came up with this diagnosis. These *(diagnostic) reasoning texts* are the object of our argumentation analysis, i.e. we aim to identify epistemic activities in the reasoning texts.

Since the texts contain highly domain specific terminology, we recruited domain-experts for the identification of epistemic activities. These experts simultaneously identified epistemic activity segments and their type. However, the original definitions of epistemic activities could not be applied one-to-one in the context of reasoning texts.

Building the Bridge between Theory and Human Argumentation

We find that four of the epistemic activities rarely occur in reasoning texts and thus focus on the four frequently used ones: hypothesis generation (HG), evidence generation (EG), evidence evaluation (EE), and drawing conclusions (DC). Their general-purpose definitions furthermore had to be interpreted in our context of reasoning texts stimulated by scenario simulations as follows. See Figure 1 for an example of epistemic activities identified based on our interpretations thereof.

EG: Due to the scenario setup, students cannot generate evidence in the original sense, i.e. by performing tests and analyses, since such evidence is already given in the scenario's information. We thus interpret EG as statements describing the explicit activity of obtaining evidence from the scenario information or by recalling own knowledge.

EE: Many students do not explicitly evaluate evidence concerning its degree of relevance in supporting or refuting a potential answer. We thus interpret the mentioning of evidence as an active selection of information considered relevant and define EE in this manner. Compared to the original definition, we also drop the restriction that EE is targeted at supporting or refuting an answer, since not all students state an answer (hypothesis or conclusion) in their reasoning texts.

First I wanted to see if the problem was new, so I checked the teacher's observations.
As it was the same back then, I ruled out a trauma or another dramatic event.
I was then undecided between autism and ADHD, *since his social behaviour seems to be problematic and that's a sign for both diagnoses*.
In the end, I settled on ADHD *since his script seems chaotic and unorganised* and *because he seems to have some friends despite his difficult behaviour*.

Figure 1. Exemplary diagnostic reasoning text from the teacher education domain, annotated with epistemic activity segments: evidence generation, *evidence evaluation*, drawing conclusions, hypothesis generation.

We also found that some of the epistemic activities were difficult to distinguish based on their definition by Fischer et al. [3].

HG versus DC: In theory, HG is the identification of *possible* solutions often not based on evidence, whereas DC involves aggregating evidence to come to a *final* decision. However, in practice the distinction is less clear. Some students state a possible diagnosis based on evidence at the beginning of the reasoning text or a certain diagnosis without any evidence, other students state a final decision without explicit reference to evidence or with uncertainty. We therefore define the difference between HG and DC based on the role they play in the reasoning process: HG *initiates* whereas DC *terminates* (a part of) the reasoning.

EE versus DC: When students generate new knowledge by evaluating given information, it is often difficult to distinguish EE and DC. We thus define DC as an evaluation leading to knowledge that forms an answer to the problem (diagnosis), whereas EE as an evaluation that may lead to knowledge about certain aspects of the problem, e.g. more information about evidence.

Results and Insights

Using these interpretations of the theory of SRA, we find that the domain-experts can reliably identify epistemic activities in diagnostic reasoning texts (agreement of 0.67 and 0.65 Krippendorff's α_U [6] between the experts in medicine and teacher education, respectively). This indicates that the argumentation formalism chosen, that is SRA, is suitable for analysing argumentative reasoning in diagnostic reasoning texts across domains.

Having built a bridge between the theory and actual human argumentation in one way (from theory to human argumentation), our context-specific interpretations of the definitions of epistemic activities may be useful to go the opposite direction in the future. In other words, the further development of the theory of SRA may be informed by the findings of our analysis of human argumentative reasoning. It is interesting to note that SRA, and in particular the epistemic activities HG and DC, resemble abductive reasoning and inference to the best explanation [1]. A detailed comparison of these theories is part of future work.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) under the reference 16DHL1040 (FAMULUS).

References

[1] Daniel G. Campos. On the distinction between peirce's abduction and lipton's inference to the best explanation. *Synthese*, 180(3):419–442, 2011.

[2] Oana Cocarascu and Francesca Toni. Detecting deceptive reviews using Argumentation. In *Proceedings of the 1st International Workshop on AI for Privacy and Security (PrAISe'16)*, pages 1–8, 2016.

[3] Frank Fischer, Ingo Kollar, Stefan Ufer, Beate Sodian, Heinrich Hussmann, Reinhard Pekrun, Birgit Neuhaus, Birgit Dorner, Sabine Pankofer, Martin R. Fischer, Jan-Willem Strijbos, Moritz Heene, and Julia Eberle. Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 4:28–45, 2014.

[4] James B. Freeman. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. De Gruyter, 1991.

[5] Ivan Habernal and Iryna Gurevych. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1):125–179, 2017.

[6] Klaus Krippendorff. On the Reliability of Unitizing Continuous Data. *Sociological Methodology*, 25:47–76, 1995.

[7] John Lawrence, Joonsuk Park, Katarzyna Budzynska, Claire Cardie, Barbara Konat, and Chris Reed. Using argumentative structure to interpret debates in online deliberative democracy and erulemaking. *ACM Transactions on Internet Technology*, 17(3):25:1–25:22, 2017.

[8] Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, and Hao Wang. Using Argument-based Features to Predict and Analyse Review Helpfulness. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17)*, 2017.

[9] Yi Song, Michael Heilman, Beata Beigman Klebanov, and Paul Deane. Applying Argumentation Schemes for Essay Scoring. In *Proceedings of the 1st Workshop on Argumentation Mining (ArgMining'14)*, pages 69–78, 2014.

[10] Christian Stab and Iryna Gurevych. Recognizing Insufficiently Supported Arguments in Argumentative Essays. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL'16)*, pages 980–990, 2016.

[11] Stephen E. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.