Dealing with Unknowability in Formal Argumentation

Pietro Baroni[1], Massimiliano Giacomin[1], and Beishui Liao[2]

[1] Dip. Ingegneria dell'Informazione, Univ. of Brescia, Brescia, Italy
{pietro.baroni,massimiliano.giacomin}@unibs.it
[2] Center for the Study of Language and Cognition, Zhejiang Univ., Hangzhou, China
baiseliao@zju.edu.cn

Abstract. In this position paper we discuss the importance of giving a proper account of unknowability in argument-based reasoning, suggest that existing formal tools are not fully adequate in this respect and lay down the basics of a research program in this direction.

Keywords: Unknowability, Structured Argumentation

1 Introduction

Human knowledge is obviously limited and anyone has to admit that there are things whose knowledge is beyond the present and possibly also future capabilities of any human being, due to various kinds of limitations. For instance, the fact that a new drug is effective for a given disease cannot be known in absence of an adequate testing. While such a limitation appears to be temporary and can be overcome with adequate actions and resources, other can be more radical. For instance, the answer to some computational problems in the worst case may require an amount of resources which overcomes Bremermann's limit[4], namely the maximum computational speed of a self-contained system in the material universe. Even more, some things are formally known to be unknowable based on undecidability proofs, like the famous Gödel theorems.

Theoretical limitations apart, unknowability plays a significant role in many practical debates about important questions.

Consider discussions about very complex global issues like climate change or world-level economic policies. Questions like "Are the causes of global warming natural or human related?" or "Are free trade policies good or bad for developing countries?" may receive both quite bold answers on either side, but one can also put forward the standpoint that any definite position on such issues is unacceptable, since it is actually impossible to provide well-founded answers to such questions. This standpoint, while possibly unsatisfactory from a psychological point of view, might turn out to be the most reasonable one, in the light of the currently available knowledge and evidences, and of the fact that such problems are "untamable" due to their intrinsic huge complexity.

In this paper, expanding a preliminary discussion in [2], we argue that unknowability gives rise to a special kind of undecidedness, called *epistemological*

undecidedness, and that it requires a proper formal treatment, which calls for a non-standard use and revision of existing argumentation formalisms. Then, we discuss the main ideas of a preliminary proposal aimed at overcoming these limitations.

The paper is organised as follows. Section 2 discusses the concept of unknowability and provides some motivation for dealing with it in argumentation, and Section 3 sketches a solution by adapting the well-known $ASPIC^+$ formalism. The solution is briefly illustrated by means of an example in Section 4. Finally, Section 5 draws some conclusions focusing on future work.

2 Unknowability and Epistemological Undecidedness

In most argumentation systems the language adopted to construct arguments allows one to assert two truth-values for each proposition, i.e. a proposition can either be (asserted as) true or (asserted as) false. Usually this is obtained by equipping the language with classical negation, so that asserting that p is false can be expressed as the assertion that $\neg p$ is true. Accordingly, each argument takes a definite position on its conclusion, supporting the acceptance of a proposition p or the acceptance of its negation $\neg p$.

As a consequence, undecidedness on propositions is a derived concept which arises from the evaluation of relevant arguments. For the sake of explanation, let is consider a Dung's style evaluation of arguments by means of a unique-status semantics [5], like grounded semantics, where arguments can be evaluated as *undefeated*, i.e. justified, *defeated*, i.e. attacked by an undefeated argument, or *provisionally defeated*, corresponding to an intermediate status which is assigned e.g. to arguments that are counter-attacked by equally preferred arguments. Considering a particular proposition p, we can then distinguish four possible justification states for p, the last two ones corresponding to (different kinds of) indecision:

- p is accepted, in case there is an argument for p which is undefeated
- p is rejected, in case there is an argument for $\neg p$ which is undefeated
- p is unknown, in case there are no arguments for p and $\neg p$, or the only arguments for p and $\neg p$ are defeated
- p is contradictory, in case there is an argument for p which is provisionally defeated.

As mentioned in Section 1, a different kind of undecidedness, called epistemological undecidedness, relies on an explicit reason supporting the standpoint that any definite answer on a proposition p is unacceptable, i.e. that p is unknowable. Examples of reasons why a proposition p is unknowable include:

- Universal principles (like Heisenberg's) saying that you cannot know everything at the same time. For instance, if you know the speed of a particle with a given precision then its momentum is unknowable with "the same" precision.

- Evidence from common knowledge: if there is a question which is known to be very debated with many disagreements about it, or a question that is still unsolved despite many attempts, one can infer that it is unknowable to everybody (e.g. the Millennium Prize Problems).
- Something which is not unknowable in principle but it is impossible to verify in any reasonable time horizon, e.g. whether there is life on a potentially habitable planet outside the Solar System.

It is important to point out the difference between epistemological undecidedness and the undecided justification states introduced above. The main difference is that epistemological undecidedness is not the result of argument evaluation, but originates from single arguments that can independently support the conclusion that a proposition p is unknowable. Thus, it clearly differs both from the status of "contradictory" and from that of "unknown". More specifically, differently from the case where p is contradictory, it does not arise from contradictory information: a single argument may support unknowability for a specific non contradictory reason, and in case this argument has no counterarguments then unknowability of p is accepted. Moreover, differently from the case where p is unknown, there is an explicit valid argument supporting unknowability, which is able to attack all those arguments that say that p is true or p is false.

In a nutshell, epistemological undecidedness and derived undecidedness are orthogonal concepts: whereas epistemological undecidedness concerns the level of assertions (language level), the justification states mentioned above concern the evaluation of arguments and their conclusions. Thus, it may well be possible that e.g. unknowability of p is contradictory since an argument supporting that p is unknowable is counterattacked by an equally preferable argument supporting that p is true.

As a final comment, it is worth noting that epistemological undecidedness cannot be encompassed by Pollock's undercutting attack [9]. The latter amounts to infer that a specific defeasible rule is not applicable because a particular exceptional condition holds. For instance, the fact that an object looks red is a defeasible reason for believing that the object is red, however this derivation is undercut in the specific circumstance that the object is illuminated by a red light. Similarly to the case where an argument supports unknowability, undercutting attack does not support a definite truth value (in the example above, the possibility that the object is red for other reasons remains open). However, epistemological undecidedness does not refer to a specific application of a defeasible rule, rather it corresponds to stating that a proposition is unknowable independently of the way a definite truth value for it is derived. Continuing Pollock's example, having a picture showing that the object is red under normal light would not be undercut, since a different rule of inference would be applied. On the other hand, if it is known that the colour is unknowable since it randomly changes continuously, then every argument saying that the object is red (including the one based on the picture) is attacked.

While the above considerations may suggest to model unknowability as a sort of universal undercut, i.e. attacking all relevant rules, it has to be noted that also defeasible premises of an argument can be affected by unknowability.

3 Unknowability in an ASPIC-like formalism

According to the considerations in the previous section, in order to capture unknowability one may exploit the general model of structured argumentation *ASPIC*$^+$, which abstracts from the specific language used to construct arguments and generalizes classical negation between formulas to a generic, possibly asymmetric, relation of contrariness $^-$. In this model, arguments are built from a knowledge base, which includes a set of language elements partitioned into axioms and ordinary premises, by applying two kinds of rules, i.e. strict and defeasible rules (depicted as single arrows $\rightarrow$ and double arrows $\Rightarrow$, respectively). Arguments can be represented as trees with roots corresponding to their conclusions and leaves to premises. The model is able to capture different kinds of conflict, including undercutting, and to deal with a preference order between arguments: we refer the reader to [7] for further details.

In order to capture unknowability, we can require that for each language symbol a there are also two other symbols, i.e. $\neg a$ and $\otimes a$, where $\otimes a$ means that a is unknowable. Then a strict or defeasible rule has a set of premises each of them of the kind p, $\neg p$ or $\otimes p$, where p ranges over the considered language, and similarly for its consequent. As a consequence, chaining rules of inference yields arguments that may have conclusions of the kind $\otimes c$ (besides c and $\neg c$), i.e. they can support unknowability.

The adoption of a third truth value makes it necessary to handle the relevant contradictions between the language elements. One option would be to include in the set of rules for any language element a the strict rules $\otimes a \rightarrow \neg a$ and $\otimes a \rightarrow \neg(\neg a)$, i.e. if a proposition is unknowable then it cannot be true and it cannot be false. However, since $\neg a$ and $\neg(\neg a)$ are in contradiction, this would mean that any argument for unknowability would give rise to a contradiction.

Another option is to exploit the generic relation of contrariness provided in *ASPIC*$^+$, by establishing that for any language element a the relations $\overline{a} = \{\neg a, \otimes a\}$, $\overline{\neg a} = \{a, \otimes a\}$ and $\overline{\otimes a} = \{a, \neg a\}$ hold, i.e. that a, $\neg a$ and $\otimes a$ are mutually contradictory. However, a technical difficulty in *ASPIC*$^+$ prevents a direct application of the formalism with a contrariness relation like this. More specifically, in order to satisfy a set of rationality postulates, a closure requirement on strict rules is required such that whenever there is a strict rule $p \rightarrow c$ then $-c \vdash -p$, where $-a$ denotes a language element a' such that $a' \in \overline{a}$ and $a \in \overline{a'}$, and $\vdash$ denotes derivation under strict rules only. In our context, this entails that whenever there is a strict rule $p \rightarrow c$ then $\neg c$ strictly entails $\neg p$ and $\otimes p$, leading to a contradiction.

Another way to model unknowability is to exploit modal operators, i.e. expressing the fact that p is known as $K(p)$, the fact that $\neg p$ is known as $K(\neg p)$ and the fact that p is unknown as $\neg K(p) \wedge \neg K(\neg p)$ (i.e. it is not known that p is true

and it is not known that $\neg p$ is true). It is worth noticing that this would not avoid the need for a non binary relation of contrariness, e.g. $\overline{K(p)} = \{K(\neg p), \neg K(p)\}$. As a consequence, the technical difficulty mentioned above should still be dealt with.

A solution to this problem can be obtained by closing the relation of contrariness under strict rules: a technical treatment of the resulting formalism is however outside the scope of the present paper, and the reader is referred to [1] for a preliminary proposal.

4 Application example

In order to illustrate the approach, let us consider a simple football example where a set of defeasible rules model the relationship between players quality and the plausibility of winning football matches. In particular, a simple modelling may include a defeasible rule $GPi \Rightarrow STi$, i.e. if a team Ti has good players (GPi) then normally it is strong (STi), a defeasible rule $BPi \Rightarrow WTi$, i.e. if a team Ti has bad players (BPi) then normally it is weak (WTi), a rule $STi, WTj \Rightarrow TiWINS$ prescribing that if a team Ti is strong and another team Tj is weak then Ti tends to win the match against Tj ($TiWINS$), and another defeasible rule $\otimes STi, WTj \Rightarrow \otimes TiWINS$ stating that the result of a match between a team with unknown strength and a weak team is unknowable.

Now, assume there is a match between two teams $T1$ and $T2$ that are known to have good and bad players, respectively. We can then build the arguments $\alpha_1 : GP1$, $\alpha_2 : \alpha_1 \Rightarrow ST1$, $\beta_1 : BP2$, $\beta_2 : \beta_1 \Rightarrow WT2$, $\gamma_1 : \alpha_2, \beta_2 \Rightarrow T1WINS$. Since there are no attacks between arguments, all of them are justified and we may conclude that $T1$ should win the match.

Assume now that $T1$ suffers some injuries ($INJ1$), and there is a defeasible rule stating that in this case the strength of a strong team becomes unknowable: $GPi, INJi \Rightarrow \otimes(STi)$. Then we can construct the following arguments in addition to the above ones: $\alpha_3 : INJ1$, $\alpha_4 : \alpha_1, \alpha_3 \Rightarrow \otimes(ST1)$, $\gamma_2 : \alpha_4, \beta_2 \Rightarrow \otimes T1WINS$. According to the relation of contrariness, α_2 and α_4 have contradictory conclusions and the same holds for γ_1 and γ_2. If we adopt a specificity criterion to compare arguments, α_4 attacks α_2 since it is based on a greater set of premises, thus it also attacks γ_1. As a consequence, γ_2 emerges as justified as well as its conclusion that the result of the match is unknowable.

5 Conclusions

In this position paper we have discussed the importance of dealing with unknowability in formal argumentation. Several research steps may then be performed.

First, a complete formal treatment based on $ASPIC^+$has to be carried out. This may also require to distinguish between the concept of "unknowability to a single agent" vs "general unknowability", which may affect the relation of attack between arguments.

A further development may account for different attitudes concerning unknowable facts, i.e. one may accept or refuse an argument based on them according to pragmatic considerations on values, risks, and so on. In line with this, it might be interesting to generalize the approach to continuous truth-values, i.e. fuzzy propositions.

From a more conceptual point of view, similar issues concerning unknowability have been investigated in formal epistemology and epistemic logics, while recent work discusses some connections between abstract argumentation and epistemic logics [6, 10]. The relationships between these works and our proposal will be a topic for further investigation, as well as the relation with formalisms to deal with uncertainty. Finally, the distinction between different kinds of undecidedness discussed in Section 2 may be handled by explicitly identifying different abstraction levels, thus raising interesting relationships with the work on meta-argumentation [8, 3].

Acknowledgments

The research reported in this paper was partially supported by the University of Brescia under the WAT_CHALLENGE project.

References

1. P. Baroni, M. Giacomin, and B. Liao. Dealing with generic contrariness in structured argumentation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 2727–2733, 2015.
2. P. Baroni, M. Giacomin, and B. Liao. I don't care, I don't know ... I know too much! on incompleteness and undecidedness in abstract argumentation. In *Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, volume 9060 of *Lecture Notes in Computer Science*, pages 265–280. Springer, 2015.
3. Guido Boella, Dov M. Gabbay, Leendert W. N. van der Torre, and Serena Villata. Meta-argumentation modelling I: methodology and techniques. *Studia Logica*, 93(2-3):297–355, 2009.
4. H. J. Bremermann. Quantum noise and information. In *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4: Biology and Problems of Health*, pages 15–20, Berkeley, Calif., 1967. University of California Press.
5. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artif. Intell.*, 77(2):321–357, 1995.
6. D. Grossi and W. van der Hoek. Justified beliefs by justified arguments. In *Proceedings of the Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning, KR14*, pages 131–140, 2014.
7. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361 – 397, 2013.
8. Sanjay Modgil and Trevor J. M. Bench-Capon. Metalevel argumentation. *J. Log. Comput.*, 21(6):959–1003, 2011.
9. J. Pollock. How to reason defeasibly. *Artif. Intell.*, 57:1–42, 1992.
10. C. Shi, S. Smets, and F.R. Velazquez-Quesada. Argument-based belief in topological structures. In *Proceedings TARK 2017*, pages 489–503, 2017.