AUSTRALASIAN JOURNAL OF LOGIC

Modeling Deep Disagreement in Default Logic

Frederik J. Andersen, Arché, University of St Andrews

Abstract

Default logic has been a very active research topic in artificial intelligence since the early 1980s, but has not received as much attention in the philosophical literature thus far. This paper shows one way in which the technical tools of artificial intelligence can be applied in contemporary epistemology by modeling a paradigmatic case of deep disagreement using default logic. In §1 model-building viewed as a kind of philosophical progress is briefly motivated, while §2 introduces the case of deep disagreement we aim to model. On the heels of this, §3 defines our formal framework, viz., a refined Horty-style default logic. §4 then uses the framework to model deep disagreement, and finally §5 provides a critical discussion of the result.

Keywords: Model-Building; Deep Disagreement; Default Logic; Social Epistemology; Formal Epistemology; Non-Monotonic Reasoning

1 Motivation

It's uncontroversial that model-building is crucial to the progress of science. When a certain phenomenon cannot be studied directly, for whatever reason, building a (formal) model of it can often lead to progress indirectly.[1] Studying a model in detail may give rise to new insights about the modeled phenomenon, and these insights can eventually result in a better model than the one we started out with.

Curiously, however, the gradual process of model-building is perhaps not as celebrated in philosophy as in science—even though building better and better models of complex phenomena is an integral part of philosophical progress as well (Williamson, 2007, 2013, 2017, 2019). As Williamson observes:

[1] Astrophysics being an illustrative example.

> [I]n philosophy, too, one form of progress is the development of better and better models... The need for model-building is hardest to avoid where the complex, messy nature of the subject matter tends to preclude informative exceptionless universal generalizations. The paradigm of such complexity and mess is the human world. Hence the obvious places to look for model-building in philosophy are those branches most distinctively concerned with human phenomena, such as ethics, epistemology, and philosophy of language. (Williamson, 2017, pp. 160-163)[2]

Social epistemology fits the bill here. The complex, multi-agent dynamics found in core topics of the field such as group rationality, expert testimony, peer disagreement, epistemic injustice etc., naturally lend themselves to systematic and intuition-guiding models.

The aim of this paper is to take a first stab at formally modeling *deep disagreement*, and to this end we'll use the formal machinery of *default logic*. A reason in favor of using this framework for modeling in social epistemology is the common interpretation of default rules (or simply defaults) as *defeasible generalizations* (Horty, 2012)—i.e., exactly the kind of generalizations Williamson takes to be prevalent in those branches of philosophy most distinctively concerned with human phenomena. An example is: *if Tweety is a bird, then Tweety can fly*. Clearly, learning the truth of the antecedent provides a reason to believe the consequent, but additionally learning that Tweety is a penguin defeats it.[3]

Below one benefit of using *formal* modeling in philosophy will be exemplified, as Hansson (2000, p. 166) writes:

> ...[F]ormalization serves to make implicit assumptions visible.

More specifically, §§4-5 of the present paper will make certain assumptions from the deep disagreement-debate come to the foreground of our analysis. We'll, for instance, see clearly how an agent-relative ranking of epistemic principles is

[2]Of course Williamson's metaphilosophy isn't completely uncontroversial. Consider the case of philosophy of language, for example: if model-building in this area means understanding natural language and linguistic phenomena via semantic models, then many philosophers (and linguists) would disagree with Williamson's metaphilosophical view.

[3]This defeasibility makes default logic a *non-monotonic* framework. Monotonic logics—such as classical logic—satisfy the property that for any well-formed formula φ from the language L, if $\varphi \in L$ is a consequence of a set of formulas $\Gamma \subseteq L$ and if $\Gamma \subseteq \Delta \subseteq L$, then φ is also a consequence of Δ. Non-monotonic logics, by contrast, allow conclusions to be withdrawn in the light of new information.

needed to even make rational deep disagreement possible; and the importance of drawing a distinction between object- and metalanguage when specifying which principles are fundamental ones will also be illuminated to us. These assumptions are oftentimes taken for granted, or are underspecified, in the informal epistemology literature.

Before getting down to business, it's worth stressing that our aim is rather modest. Our ambition is merely to construct a provisional model of paradigmatic deep disagreement, which is open to—perhaps even in need of—further innovation. Yet, our modesty should not be confused with a lack of ambition as it encapsulates the spirit of model-building very well. Just like in science; model-building in philosophy is an incremental achievement.

2 Deep Disagreement

Let's now get an intuitive grasp of deep disagreement, which is the phenomenon that we want to model. Consider the **Young Earth Creationist**:

> Henry is an Evangelical young Earth creationist, who accepts that the Earth is no more than 6000 years old and a nexus of conspiratorial claims as evidence of why scientists have been misleading us about the age of the Earth. Henry also rejects the theory of evolution and contemporary cosmology, citing literal readings of the Bible: 'your denial of scripture is unjustified', he says. Henry's neighbor Richard is a proponent of so-called 'New Atheism', and rejects the religious and young Earth creationist views of his neighbor Henry, and asserts that the Earth is much older than 6000 years: 'your denial of geology and evolutionary biology are unjustified', he says. (Ranalli, 2021, p. 984)

This case has been widely discussed in the literature and is considered a paradigmatic case of deep disagreement (Lynch, 2010; Pritchard, 2010; Kappel, 2012; Hazlett, 2014; Ranalli, 2021; Ranalli and Lagewaard, 2022a,b).

Although there are several different ways of understanding the essentials of deep disagreement, we'll focus on the so-called *Fundamental Epistemic Principle Theory* to avoid unnecessary detours.[4] According to this theoretical stance, deep disagreements are *deep* because they are not solely concerned with "surface-level"

[4]According to Ranalli (2021), state of the art research on how to best characterize deep disagreement falls roughly into two theoretical camps. On the one hand we have the *Hinge Propo-*

propositions about, say, a particular weather forecast (Christensen, 2007), but also propositions stating the fundamental epistemic principles we ought to apply when trying to predict the weather in general. In other words, deep disagreements are disagreements over fundamental epistemic principles like those specifying which traditions, institutions, methods, sources of evidence, and patterns of reasoning to rely upon when forming beliefs (Kappel and Andersen, 2019).

Rational irresolvability is often considered a necessary property of deep disagreements because of their dialectical set-up (Wittgenstein, 1969; Fogelin, 2005; Lynch, 2010, 2016; Kappel, 2012). How is one supposed to give a compelling argument for target-proposition *p,* when one's interlocutor asserts not-*p* (or suspends judgement as to whether *p*), and does so by appealing to fundamental epistemic principles that conflict with one's own?[5] In the words of Michael Lynch:

> ...explicit defenses of such principles will always be subject to a charge of circularity. Hume showed that the principle of induction is like this: you can't show that induction is reliable without employing induction. It also seems true of observation or sense perception. It seems difficult, to say the least, to prove that any of the senses are reliable without at some point employing one of the senses. Similarly with the basic principles of deductive logic: I can't prove basic logical principles without relying on them. In each case, I seem to have hit rock bottom... (Lynch, 2016, pp. 250-251)

As should be clear—in the case **Young Earth Creationist**—Henry and Richard disagree about the age of the Earth at surface-level, but their disagreement depends on a much more fundamental disagreement about evidential standards and what justifies beliefs. This is why their story has come to be viewed as a paradigmatic case of *deep* disagreement.

3 Default Logic

Default logic has been a very active research topic in artificial intelligence since the early 1980s (Reiter, 1980; McDermot and Doyle, 1980; Reiter and Criscuolo,

sition Theorists (Wittgenstein, 1969; Feldman, 2005; Fogelin, 2005; Friemann, 2005; Hazlett, 2014); on the other the *Fundamental Epistemic Principle Theorists* (Lynch, 2010; Kappel, 2012; Jønch-Clausen and Kappel, 2015; Lynch, 2016; Kappel, 2021; Lagewaard, 2021).

[5]See (Ranalli, 2020) for a helpful disambiguation of the term 'rational irresolvable'. Consult (Martin, 2021) for a recent argument *against* the rational irresolvability of deep disagreements.

1981; Mccarthy, 1986; Poole, 1988; Brewka, 1989; Baader and Hollunder, 1993; Brewka, 1994a,b; Makinson, 1994; Baader and Hollunder, 1995; Rintanen, 1995; Antoniou et al., 1996; Rintanen, 1998; Antoniou, 1999; Brewka and Eiter, 2000; Antonelli, 2005; Thomason, 2018), but has not received as much attention in the philosophical literature thus far.[6]

Nonetheless, John F. Horty's monograph *Reasons as Defaults* (2012) highlights several promising applications of default logic in philosophy—e.g., modeling the structure and strength of reasons, defeaters, and arguments.[7] This section refines the basic definitions of Horty's default logic such that it can model the multi-agent dynamics of typical deep disagreement scenarios.

3.1 Horty's Framework

In its most basic form Horty's default logic is simply classical propositional logic extended with default rules.

Definition 1. Let Φ be a countable set of atomic propositions and $\mathscr{L}$ a language such that:

$$\varphi := p \mid \top \mid \neg\psi \mid \psi \to \psi'$$

When $\Gamma \subseteq \mathscr{L}$ and $\varphi \in \mathscr{L}$, we write $\Gamma \vdash \varphi$ to express left-to-right classical deducibility. Denote the logical closure of Γ by $Th(\Gamma) := \{\varphi : \Gamma \vdash \varphi\}$. Where $\varphi, \psi \in \mathscr{L}$, a default rule is any expression of the form:

$$(\varphi \rightsquigarrow \psi)$$

It's important to notice that default rules are metalinguistic, so they cannot be expressed in $\mathscr{L}$. Further, the symbol '$\rightsquigarrow$' cannot be nested to generate more complex default rules.

We let $\mathscr{D}$ denote the set of all possible defaults (with typical elements $\delta, \delta' ...$). For a default rule $\delta = (\varphi \rightsquigarrow \psi)$, let $Conclusion(\delta) := \psi$. And for a set of default rules $D \subseteq \mathscr{D}$, let $Conclusions(D) := \{Conclusion(\delta) : \delta \in D\}$.

Consider next a rational agent's basis for default reasoning.

[6] Yet, it's worth flagging that philosophical works involving default logic has become more common in recent years, see, e.g., (Bonevac, 2018; Knoks, 2021a,b, 2022).

[7] See also (Horty, 2007a,b, 2016).

Definition 2. A single agent default theory is a tuple:

$$\Delta = (W, D, \leq, \Gamma)$$

W denotes the agent's set of background information, i.e., hard facts; D refers to the set of default rules which are *available* to the agent (these need not be plausible defaults, just available ones). The order $\leq$ is a non-strict partial order on D with the formal properties transitivity, reflexivity, and antisymmetry.[8] Suggesting the following reading of $\delta \leq \delta'$: "δ' *represents a default of a priority which is at least as high as the one* δ *represents*," where "*higher priority*" means less easily defeasible. We say that $\delta \in D$ is *fundamental* when there is no $\delta' \in D$ such that $\delta < \delta'$, i.e., when there is no other available default δ' of strictly higher priority than δ in D.[9]

A **scenario** S (based on a default theory Δ) is a subset $S \subseteq D \subseteq \mathscr{D}$ contained in Δ. We interpret S as a particularly plausible set of available default rules, i.e., the defaults of which the antecedents provide sufficient support for their conclusions according to the agent in question. We'll assume that if a given agent considers δ fundamental, then $\delta \in S$ also holds for that agent.

The last element of the tuple (i.e., the default theory) is the agent's belief set Γ.

Definition 3. Define a belief set:

$$\Gamma = Th(W \cup Conclusions(S))$$

[8] Relation R is transitive if and only if $\forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow Rxz)$. R is reflexive if and only if $\forall x Rxx$. R is antisymmetric if and only if $\forall x \forall y((Rxy \wedge Ryx) \rightarrow x = y)$.

Note that when it comes to the requirements on the ordering of defaults, Horty argues that transitivity is a natural requirement, that the relation should be irreflexive (i.e., strict) so that "no default can ever have a higher priority than itself" (2012, p. 20), and that the relation should not be strongly connected, i.e., the priority order should not be assumed *connex,* that for any distinct defaults δ, δ', either $\delta < \delta'$ or $\delta' < \delta$, as—although this would help to resolve conflicts between defaults—the requirement would be unreasonable. Horty provides two reasons against an assumption of connex: (1) some defaults are simply incommensurable, and (2) some defaults may have equal priorities.

Notice that reason (2) contrasts with the choice of a *strict* order and suggests using a preorder $\leq$ instead. The order is then *reflexive* instead of irreflexive, with the also natural interpretation that every default is comparable to itself, and to itself it has *the same* priority. As a preorder, it may still be partial, in accordance with Horty's intuitive examples (2012, p. 20).

[9] Notice that this understanding of fundamentality allows a reasoner to have multiple fundamental defaults as long as they are of equal priority.

That is, the logical closure of the hard background information plus the conclusion-set of the plausible defaults available to the agent.

To illustrate, consider default theory Δ such that $W = \{p\}$, $D = \{\delta\}$ with $\delta = (p \rightsquigarrow \neg q)$, and $\leq = \emptyset$. Assuming that δ is plausible, the resulting belief set is $\Gamma = Th(\{p, \neg q\})$.[10]

Another—slightly more sophisticated—example is a classic of non-monotonic reasoning. The example concerns the bird Tweety and its ability to fly. The fact that Tweety is a bird provides a reason to conclude that Tweety can fly. But if Tweety is also a penguin, the reason to conclude that Tweety can fly is defeated. The details of the Tweety-example is captured in Figure 3.1 below.
Since $\delta_1 < \delta_2$ holds true, a rational agent should only endorse the default $\delta_2 = (p \rightsquigarrow \neg f)$ in the Tweety-case (Horty, 2012, pp. 23–25, 32–33); and consequently end up with $\Gamma = Th(\{p, p \rightarrow b, \neg f\})$.

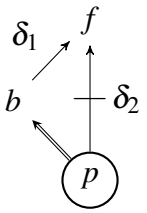
A final example concerns conflicting information about the former US president Nixon. The so-called "Nixon Diamond" is summarized in Figure 3.2. This example illustrates the difficulties of drawing an unambiguous conclusion when one is presented with conflicting information: That Nixon is a Quaker constitutes a reason to believe that he is a pacifist, while Nixon being Republican provides a reason to believe that he is not. Our default logic framework reflects the seemingly insolvable conflict between the two defaults δ_1 and δ_2 since—assuming that both defaults are plausible—the Nixon Diamond has two scenarios that a rational

[10]Horty *doesn't* directly associate extensions of default theories with beliefs (Horty, 2012, pp. 34-40): a default theory Δ may have no extensions or multiple ones, and identifying the Δ-beliefs with *the* extension of Δ is therefore not well-defined. Horty discusses both multiple and empty extensions, but he does not give a clear solution. As we won't be confronted with empty extensions in this paper, we simply ignore that problem. For multiple extensions, we can interpret every extension of a default theory as a possible equilibrium state that an ideal reasoner might arrive at—i.e., as a *possible belief state*.

Horty remarks that belief sets based on arbitrary scenarios are unsatisfactory (2012, p. 23). According to him, satisfactory belief sets are obtained only from what he calls 'proper scenarios'. The definition of a proper scenario requires the auxiliary notions of *triggered*, *conflicted*, and *defeated* defaults. Let $S \subseteq D$ be a scenario based on $\Delta = (W, D, <)$. Define

$$\begin{aligned}
Triggered(\Delta, S) &= \{\delta \in D\colon W \cup Conclusions(S) \vdash Premise(\delta)\}.\\
Conflicted(\Delta, S) &= \{\delta \in D\colon W \cup Conclusions(S) \vdash \neg Conclusion(\delta)\}.\\
Defeated(\Delta, S) &= \{\delta \in D\colon \exists \delta' \in Triggered(\Delta, S) \text{ such that}\\
&\qquad \delta < \delta' \text{ and } W \cup \{Conclusion(\delta')\} \vdash \neg Conclusion(\delta)\}.
\end{aligned}$$

Using these three notions, Horty presents two definitions of a proper scenario, cf. Definitions 5-7 and 26-27 of Reasons as Defaults, respectively.



$\delta_1 < \delta_2$

Propositions:
b : Tweety is a bird. f : Tweety flies.
p : Tweety is a penguin.

All possible scenarios:
$\emptyset$,
$S_1 = \{\delta_1\}$, $S_2 = \{\delta_2\}$,
$S_3 = \{\delta_1, \delta_2\}$

Figure 3.1: **The Tweety Triangle**. Circled propositions constitute the set of hard background information; the double arrow shows that $(p \to b)$ is in the background information. A δ-labeled arrow from one formula φ to another ψ means the default $\delta = (\varphi \rightsquigarrow \psi)$ is among the available defaults. When a δ-labeled arrow from φ to ψ is crossed out, it means that $\delta = (\varphi \rightsquigarrow \neg\psi)$ is available. For the order $\leq$ we omit reflexive loops and links obtainable by transitive closure to ease readability.

agent aiming for consistency could endorse, namely $S_1 = \{\delta_1\}$ and $S_2 = \{\delta_2\}$, resulting in $\Gamma_1 = Th(\{q, r, p\})$ and $\Gamma_2 = Th(\{q, r, \neg p\})$, respectively. Hence, the framework does not specify a unique belief set, but leaves us with an open-ended conclusion.[11]

To conclude our formal framework we refine Definition 2 enabling it to handle cases with multiple agents.

Definition 4. A multi-agent default theory is tuple:

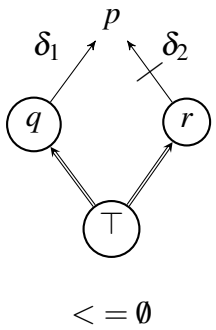
$$\Delta_i = (W_i, D_i, \leq_i, \Gamma_i)_{i \in A}$$

where 'A' denotes a countable set of agents with typical elements $a, b, c...$

4 The Model

Now, let's put our formal framework to use and construct a model of the **Young Earth Creationist** as advertised earlier. The agents disagreeing—i.e., Henry and Richard—are endorsing different fundamental epistemic principles (modeled as fundamental default rules, cf. Definition 2) with incompatible conclusions. More explicitly:

[11] Horty (2012, pp. 34–37) discusses three possible ways to deal with multiple extensions.



Propositions:
q : Nixon is a Quaker.
r : Nixon is Republican.
p : Nixon is a pacifist.

All possible scenarios:
$\emptyset$,
$S_1 = \{\delta_1\}$, $S_2 = \{\delta_2\}$,
$S_3 = \{\delta_1, \delta_2\}$

Figure 3.2: **Nixon Diamond**. The bottom $\top$-node and the double arrows are included to retain the traditional diamond shape, but are superfluous when circling the background information.

- Let p represent the target-proposition of the disagreement, viz., *Planet Earth is no more than 6000 years old*;
- let q denote the proposition *The Bible asserts that Planet Earth is no more than 6000 years old*;
- and finally, let r refer to the proposition *The scientific consensus is that Planet Earth is more than 6000 years old*.

So, Henry endorses fundamental default $\delta = (q \rightsquigarrow p)$ whereas Richard endorses fundamental default $\delta' = (r \rightsquigarrow \neg p)$, which suggests the following three-step logical analysis of their disagreement.[12]

1. *Initial situation*. Henry and Richard's situation can be explicated using a multi-agent default theory $\Delta_i = (W_i, D_i, \leq_i, \Gamma_i)_{i \in A}$ (cf. Definition 4). Let a refer to Henry and b to Richard (such that $a, b \in A$). We can assume that a and b each has internally consistent belief sets, and that q is in a's background information while r is in b's ditto. Given a's endorsement of δ and b's endorsement of δ', the belief set $\Gamma_a \cup \Gamma_b$ is inconsistent (by Definition 3). Hence, a and b, are in a state of potential deep disagreement.

2. *Appreciation*. a and b realize that they are in deep disagreement.

[12]It would actually be a fair objection to claim that realistic instances of *fundamental* epistemic principles should be captured by some much more *schematic* default rules than those suggested here. Yet we allow ourselves to neglect this complication in order to keep things simple.

3. *Update*. a and b exchange information about their respective positions, thus we need an updated multi-agent default theory to capture a state of full disclosure: $\Delta_i' = (W_i', D_i', \leq_i', \Gamma_i')_{i \in A}$, where $W_i' = W_i \cup W_j, i \neq j$; $D_i' = D_i \cup D_j, i \neq j$. As D_a and D_b are disjoint (yet comparable) the ordering $\leq_i'$ can either be specified such that for all fundamental $\delta_i \in D_i$ and all fundamental $\delta_j \in D_j$: $\delta_i >_i' \delta_j$, or $\delta_i =_i' \delta_j$, or $\delta_i <_i' \delta_j$. Each of these corresponds to a specific type of response to deep disagreement.

- **Steadfastness**: $\delta_i >_i' \delta_j$ represents a conservative rationale where the new information is considered of less priority than the old. Hence, the beliefs of both agents will be unaffected by full disclosure.

- **The Equal Weight View**: $\delta_i =_i' \delta_j$ represents a rationale where new and old information is considered equal. This is a strong conciliationist rationale leading each agent to suspend judgement—i.e., each agent would become undecided about what fundamental principle to endorse after the update (on the pain of inconsistency).[13]

- **World View Switching**: $\delta_i <_i' \delta_j$ represents a rationale where the new information is considered of higher priority than the old. Thus, a would adopt b's initial belief set (cf. step 1) and *vice versa*. While this response may seem unrealistic, it neatly captures the drastic nature of deep disagreement, i.e., succumbing to one's opponent means giving up one's fundamental epistemic principle(s).[14]

5 Discussion

So far, so good. The model we have just constructed is both provisional and extremely simple, yet it does quite well in modeling the interpersonal dynamics of typical disagreement cases. It's a first-mover in modeling *deep* disagreement using the tools of default logic and it can easily be augmented to bring about more

[13] This outcome is technically unproblematic for us because default theories allow for multiple extensions (cf. footnote 10).

[14] A potential fourth response to deep disagreement—where one side is steadfast while the other switches their world view—would require us to accept that two rational agents can react differently upon realizing deep disagreement. Whether this is a tenable option depends on our understanding of rationality, see, e.g., (Fogal and Worsnip, 2021), for a useful discussion of *structural* versus *substantive* rationality.

sophisticated models, e.g., by drawing on technical results in Default Logic from artificial intelligence, or from neighboring fields such as AGM Belief Revision (Alchourrón et al., 1985; Hansson, 2017) and Epistemic Logic (Hintikka, 1962; Rendsvig and Symons, 2019). The model captures standard responses to disagreement known from the epistemological literature, i.e., Steadfastness (Kelly, 2005; Titelbaum, 2015) and (strong) Conciliationism (Elga, 2007; Matheson, 2009), and shows surplus by mirroring the drastic nature of deep disagreement qua World View Switching.

These merits notwithstanding it's fair to ask whether our framework of default logic does a better job modeling deep disagreement than its most obvious rivals. Consider first *Classical Propositional Logic.* In this framework one could represent fundamental epistemic principles as material conditionals. Default logic, as defined above, is merely an extension of classical propositional logic, so in case the classical framework does equally well modeling deep disagreement, Ockham's Razor would force us to adopt the classical alternative.

This move would come with a serious drawback for our present purposes, however. Notice that in our model from §4 we represent fundamental epistemic principles as default rules—i.e., as metalinguistic items beyond the scope of explicit evaluation. In contrast, treating such principles as material conditionals would make them part of an object-language, and thus eligible to explicit evaluation (as objects of belief). This is clearly an undesirable feature of the classical framework when it comes to modeling deep disagreement. Fundamental epistemic principles are supposed to be the kind of principles we normally take for granted in disagreements; not just ordinary targets of evaluation (Wittgenstein, 1969; Fogelin, 2005; Lynch, 2010, 2016; Kappel, 2012). Hence, our modeling of fundamental epistemic principles as metalinguistic default rules seems superior to at least one rival.[15]

But how about *Subjective Bayesianism*? In a Bayesian framework one could represent fundamental epistemic principles as update functions.[16] This would enable us to model a common sense response to (deep) disagreement, which is

[15] As an anonymous reviewer points out it's quite plausible to think that *dynamic epistemic logic*-style ('DEL-style') systems could actually do a better job than at least the classical propositional logic here. Some dynamic epistemic logics (2010, Chapter 4)—which can represent inference rules in the object language—could perhaps even do better than our own default logic framework in a number of ways. However, this is only with the important cost of simplicity in the object-language. For further discussion of the link between these DEL-style logics and default reasoning, see (Velazquez-Quesada, 2010; Solaki, 2021).

[16] See (Talbott, 2016; Schupbach, 2022) for more on (Subjective) Bayesianism.

neglected by our default logic-model, viz., adjusting one's confidence levels appropriately in the target-proposition under dispute (and in one's background assumptions and epistemic standards).

Even so, for the purposes of modeling *deep* disagreement in particular there seems to be a serious downside to Bayesianism. On the assumption that there is only one rational update function, the Bayesian will be unable to model *rational irresolvability*. For a disagreement to count as rationally irresolvable—by Bayesian lights—the parties involved would need to endorse non-equivalent update functions. Otherwise even agents with radically different priors would eventually converge on a rational credence (see for instance (Earman, 1992, Chapter 6) for the formal details). This seems to count in favor of our default logic-model because its agent-relative ranking of defaults allows two completely rational agents to disagree with each other.[17]

Summa: we have constructed an elementary model of deep disagreement using the technical tools of default logic, and compared the result with some obvious competitors. We have seen that our simple model fares quite well in comparison to both Classical Propositional Logic and Subjective Bayesianism. Of course we haven't made a decisive argument for default logic vis-à-vis modeling deep disagreement, but as stated, our proposed model is merely meant as a provisional one to be further discussed and refined, as is indeed the very core of the model-building perspective.

Acknowledgments

I would like to thank the audience at an Arché seminar meeting in December 2022, University of St Andrews, for constructive feedback on an earlier version of this paper. From memory the audience included Andrea Oliani, Franz Berto, Matteo Nizzardo, Sophie Nagler, Viviane Fairbank, and Greg Restall. Further, I would like to thank the audiences at the conference Disagreements: From Theory to Practice, March 2020, University of Tartu; and 1st SEECRs Workshop, July 2022, University of Edinburgh, for helpful Q&A-sessions. Finally, I want to express special gratitude to Rasmus K. Rendsvig and Makan Nojoumian, for written comments on my drafts.

[17]This is at least true on some accounts that understand rationality as some form of internal coherence.

Funding

My research is supported by the Niels Bohr Foundation (Royal Danish Academy of Sciences) and the Royal Institute of Philosophy (London).

References

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(2):510–530.

Antonelli, A. (2005). *Grounded Consequence for Defeasible Logic*. Cambridge University Press.

Antoniou, G. (1999). A Tutorial on Default Logics. *ACM Computing Surveys (CSUR)*, 31(4):337–359.

Antoniou, G., O'Neill, T., and Thurbon, J. (1996). Studying properties of classes of default logics: Preliminary report. pages 558–569. Springer.

Baader, F. and Hollunder, B. (1993). How to Prefer More Specific Defaults in Terminological Default Logic. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1993)*, pages 669–674.

Baader, F. and Hollunder, B. (1995). Priorities on Defaults with Prerequisites, and their Application in Treating Specificity in Terminological Default Logic. *Journal of Automated Reasoning*, 15(1):41–68.

Bonevac, D. (2018). Defaulting on reasons. *Noûs*, 52(2):229–259.

Brewka, G. (1989). Preferred Subtheories: An Extended Logical Framework for Default Reasoning. In Sridharan, N., editor, *Proceedings of the eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1043–1048. Morgan Kaufmann Publishers.

Brewka, G. (1994a). Adding Priorities and Specificity to Default Logic. In *Logics in Artificial Intelligence: European Workshop JELIA '94*, pages 247–260. Springer.

Brewka, G. (1994b). Reasoning about Priorities in Default Logic. In *AAAI*, volume 1994, pages 940–945.

Brewka, G. and Eiter, T. (2000). Prioritizing Default Logic. In Hölldobler, S., editor, *Intellectics and Computational Logic*, volume 19 of *Applied Logic Series*, pages 27–45.

Christensen, D. (2007). Epistemology of disagreement: the good news. *Philosophical Review*, 116:187–217.

Earman, J. (1992). *Bayes or Bust?* MIT Press.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41:478–502.

Feldman, R. (2005). Deep disagreement, rational resolutions, and critical thinking. *Informal Logic*, 25(1).

Fogal, D. and Worsnip, A. (2021). Which reasons? which rationality? *Ergo*, 8.

Fogelin, R. (2005). The logic of deep disagreements. *Informal logic*, 25(1).

Friemann, R. (2005). Emotional backing and the feeling of deep disagreement. *Informal Logic*, 25(1).

Hansson, S. O. (2000). Formalization in philosophy. *Bulletin of Symbolic Logic*, 6(2):162–175.

Hansson, S. O. (2017). Logic of belief revision. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition.

Hazlett, A. (2014). Entitlement and mutually recognized reasonable disagreement. *Episteme*, 11(1):1–25.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca, N.Y.

Horty, J. F. (2007a). Defaults with Priorities. *Journal of Philosophical Logic*, 36(4):367–413.

Horty, J. F. (2007b). Reasons as Defaults. *Philosophers' Imprint*, 7:1–28.

Horty, J. F. (2012). *Reasons as defaults*. Oxford: Oxford University Press.

Horty, J. F. (2016). Reasoning with Precedents as Constrained Natural Reasoning. In Lord, E. and Maguire, B., editors, *Weighing Reasons*. Oxford University Press.

Jønch-Clausen, K. and Kappel, K. (2015). Social epistemic liberalism and the problem of deep epistemic disagreements. *Ethical Theory and Moral Practice*, 18(2):371–384.

Kappel, K. (2012). The problem of deep disagreement. *Discipline Filosofiche*, 22(2):7–25.

Kappel, K. (2021). Higher order evidence and deep disagreement. *Topoi*, 40(5):1039–1050.

Kappel, K. and Andersen, F. J. (2019). Moral disagreement and higher-order evidence. *Ethical Theory and Moral Practice*, 22(5):1103–1120.

Kelly, T. (2005). The epistemic significance of disagreement. In Hawthorne, J. and Gendler, T., editors, *Oxford Studies in Epistemology, 1*, pages 167–96. Oxford: Oxford University Press.

Knoks, A. (2021a). Conciliatory reasoning, self-defeat, and abstract argumentation. *The Review of Symbolic Logic*, pages 1–48.

Knoks, A. (2021b). Misleading higher-order evidence, conflicting ideals, and defeasible logic. *Ergo, An Open Access Journal of Philosophy*, 8(6):141–174.

Knoks, A. (2022). Conciliatory views, higher-order disagreements, and defeasible logic. *Synthese*, 200(2):1–23.

Lagewaard, T. (2021). Epistemic injustice and deepened disagreement. *Philosophical Studies*, 178(5):1571–1592.

Lynch, M. (2010). Epistemic circularity and epistemic disagreement. In Haddock A, Millar A, P. D., editor, *Social epistemology*. Oxford University Press.

Lynch, M. P. (2016). After the spade turns: disagreement, first principles and epistemic contractarianism. *International Journal for the Study of Skepticism*, 6(2-3):248–259.

Makinson, D. (1994). General Patterns in Non-Monotonic Reasoning. In Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming: Nonmonotonic Reasoning and Uncertain Reasoning (vol. 3)*, pages 35–110. Oxford University Press.

Martin, B. (2021). Searching for deep disagreement in logic: the case of dialetheism. *Topoi*, 40(5):1127–1138.

Matheson, J. (2009). Conciliatory views of disagreement and higher-order evidence. *Episteme*, 6(3):269–279.

Mccarthy, J. (1986). Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28(1):89–116.

McDermot, D. and Doyle, J. (1980). Non-monotonic logic 1. *Artificial Intelligence*, 13:41–72.

Poole, D. (1988). A Logical Framework for Default Reasoning. *Artificial Intelligence*, 36:27–47.

Pritchard, D. (2010). Epistemic relativism, epistemic incommensurability and wittgensteinian epistemology. *S. Hales, Blackwell Companion to Relativism*, pages 266–285.

Ranalli, C. (2020). Deep disagreement and hinge epistemology. *Synthese*, 197(11):4975–5007.

Ranalli, C. (2021). What is deep disagreement? *Topoi*, pages 983–998.

Ranalli, C. and Lagewaard, T. (2022a). Deep disagreement (part 1): Theories of deep disagreement. *Philosophy Compass*, 17(12):e12886.

Ranalli, C. and Lagewaard, T. (2022b). Deep disagreement (part 2): Epistemology of deep disagreement. *Philosophy Compass*, 17(12):e12887.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13:81–132.

Reiter, R. and Criscuolo, G. (1981). On Interacting Defaults. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, pages 270–276.

Rendsvig, R. and Symons, J. (2019). Epistemic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Rintanen, J. (1995). On Specificity in Default Logic. In Mellish, C., editor, *IJCAI'95*, pages 1474–1479. AAAI Press.

Rintanen, J. (1998). Lexicographic Priorities in Default Logic. *Artificial Intelligence*, 106(2):221–265.

Schupbach, J. N. (2022). *Bayesianism and Scientific Reasoning*. Cambridge University Press.

Solaki, A. (2021). *Logical models for bounded reasoners*. PhD thesis, University of Amsterdam.

Talbott, W. (2016). Bayesian Epistemology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition.

Thomason, R. (2018). Logic and artificial intelligence. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Titelbaum, M. (2015). Rationality fixed point (or: In defense of right reason). In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, pages 253–294. Oxford University Press.

Velazquez-Quesada, F. R. (2010). *Small steps in dynamics of information*. University of Amsterdam.

Williamson, T. (2007). *The philosophy of philosophy*. Wiley Blackwell.

Williamson, T. (2013). Gettier cases in epistemic logic. *Inquiry*, 56(1):1–14.

Williamson, T. (2017). Model-building in philosophy. In Blackford, R. and Broderick, D., editors, *Philosophy's Future*, pages 159–71. Wiley Blackwell.

Williamson, T. (2019). Evidence of evidence in epistemic logic. In Skipper, M. and Steglich-Petersen, A., editors, *Higher-order evidence. New essays*, pages 265–297. Oxford University Press.

Wittgenstein, L. (1969). *On certainty*. Blackwell Oxford.