Strength Factors: An Uncertainty System for a Quantified Modal Logic

Naveen Sundar Govindarajulu, Selmer Bringsjord

Strength Factors: An Uncertainty System for a Quantified Modal Logic

Naveen Sundar Govindarajulu and **Selmer Bringsjord**
Rensselaer Polytechnic Institute, Troy, NY
{naveensundarg,selmer.bringsjord}@gmail.com

Abstract

We present a new system $\mathcal{S}$ for handling uncertainty in a quantified modal logic (first-order modal logic). The system is based on both probability theory and proof theory and is derived from Chisholm's epistemology. We concretize Chisholm's system by grounding his undefined and primitive (i.e. foundational) concept of **reasonableness** in probability and proof theory. We discuss applications of the system. The system described below is a work in progress; hence we end by presenting a list of future challenges.

1 Introduction

We introduce a new system $\mathcal{S}$ for talking about uncertainty of iterated beliefs in a quantified modal logic with belief operators. The quantified modal logic we use is based on the **deontic cognitive event calculus** ($\mathcal{DCEC}$), which belongs to the family of **cognitive calculi** that have been used in modeling complex cognition. Here, we use a subset of $\mathcal{DCEC}$ that we term **micro cognitive calculus** ($\mu\mathcal{C}$). Specifically, we add a system of uncertainty derived from Chisholm's epistemology [Chisholm, 1987].[1] The system $\mathcal{S}$ is a work in progress and hence the presentation here will be abstract in nature.

One of our primary motivations is to design a system of uncertainty that is easy to use in end-user facing systems. There have been many studies that show that laypeople have difficulty understanding raw probability values (e.g. see [Kaye and Koehler, 1991]); and we believe that our approach borrowed from philosophy can pave the way for systems that can present uncertain statements in a more understandable format to lay users.

$\mathcal{S}$ can be useful in systems that have to interact with humans and provide justifications for their uncertainty. As a demonstration of the system, we apply the system to provide a solution to the lottery paradox. Another advantage of the system is that it can be used to provide uncertainty values for counterfactual statements. Counterfactuals are statements that an agent knows for sure are false. Among other cases, counterfactuals are useful when systems have to explain their actions to users (*If I had not done* α*, then* ϕ *would have happened*). Uncertainties for counterfactuals fall out naturally from our system. Before we discuss the calculus and present $\mathcal{S}$, we go through relevant prior work.

2 Prior Work

Members in the family of cognitive calculi have been used to formalize and automate highly intensional reasoning processes.[2] More recently, using $\mathcal{DCEC}$ we have presented an automation of **the doctrine of double effect** in [Govindarajulu and Bringsjord, 2017].[3] We quickly give an overview of the doctrine to illustrate the scope and expressivity of cognitive calculi such as $\mathcal{DCEC}$. The doctrine of double effect is an ethical principle that has been shown to be used by both untrained laypeople and experts when faced with moral dilemmas; and it plays a central role in many legal systems. Moral dilemmas are situations in which all available options have both good and bad consequences. The doctrine states that an action α in such a situation is permissible *iff* — (1) it is morally neutral; (2) the net good consequences outweigh the bad consequences by some large amount γ; and (3) at least one or more of the good consequences are *intended*, and none of the bad consequences are intended. The conditions require both intensional operators and a calculus (e.g. the event calculus) for modeling commonsense reasoning and the physical world. Other tasks automated by cognitive calculi include the false-belief task [Arkoudas and Bringsjord, 2008] and *akrasia* (succumbing to temptation to violate moral principles) [Bringsjord *et al.*, 2014].[4] Each cognitive calculus is a sorted (i.e. typed) quantified modal logic (also known as sorted first-order modal logic). Each calculus has a well-defined syntax and proof calculus. The proof calculus is based on natural deduction [Gentzen, 1935], and includes all the introduction

[1] See the SEP entry on Chisholm for a quick overview of Chisholm's epistemology: https://plato.stanford.edu/entries/chisholm/#EpiIEpiTerPriFou.

[2] By "intensional processes", we roughly mean processes that take into account knowledge, beliefs, desires, intentions, etc. of agents. Compare with extensional systems such as first-order logic that do not take into account states of minds of other agents. This is not be confused with "intentional" systems which would be modeled with intensional systems. See [Zalta, 1988] for a detailed treatment of intensionality.

[3] This work will be presented at IJCAI 2017.

[4] Arkoudas and Bringsjord [2008] introduced the general family of **cognitive event calculi** to which $\mathcal{DCEC}$ belongs.

and elimination rules for first-order logic, as well as inference schemata for the modal operators and related structures.

On the uncertainty and probability front, there have been many logics of probability, see [Demey *et al.*, 2016] for an overview. Since our system builds upon probabilities, our approach could use a variety of such systems. There has been very little work in uncertainty systems for first-order modal logics. Among first-order systems, the seminal work in [Halpern, 1990] presents a first-order logic with modified semantics to handle probabilistic statements. We can use such a system as the foundation for our work, and use it to define the base probability function **Pr** used below. (Note that we leave **Pr** unspecified for now.)

3 The Formal System

The formal system $\mu\mathcal{C}$ is a modal extension of the the event calculus. The event calculus is a multi-sorted first-order logic with a family of axiom sets. The exact axiom set is not important. The primary sorts in the system are shown below.

Sort	Description
Agent	Human and non-human actors.
Moment or Time	Time points and intervals. E.g. simple, such as t_i, or complex, such as $birthday(son(jack))$.
Event	Used for events in the domain.
ActionType	Action types are abstract actions. They are instantiated at particular times by actors. E.g.: "eating" vs. "jack eats."
Action	A subtype of Event for events that occur as actions by agents.
Fluent	Used for representing states of the world in the event calculus.

Full $\mathcal{DCEC}$ has a suite of modal operators and inference schemata. Here we focus on just two: an operator for belief **B** and an operator for perception **P**. The syntax of and inference schemata of the system are shown below. S is the set of all sorts, f are the core function symbols, t shows the set of terms, and ϕ is the syntax for the formulae.

Syntax

$$S ::= \begin{cases} \mathsf{Object} \mid \mathsf{Agent} \mid \mathsf{Self} \sqsubset \mathsf{Agent} \mid \mathsf{ActionType} \mid \mathsf{Action} \sqsubseteq \mathsf{Event} \mid \\ \mathsf{Moment} \mid \mathsf{Formula} \mid \mathsf{Fluent} \mid \mathsf{Numeric} \end{cases}$$

$$f ::= \begin{cases} action : \mathsf{Agent} \times \mathsf{ActionType} \to \mathsf{Action} \\ initially : \mathsf{Fluent} \to \mathsf{Formula} \\ holds : \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula} \\ happens : \mathsf{Event} \times \mathsf{Moment} \to \mathsf{Formula} \\ clipped : \mathsf{Moment} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula} \\ initiates : \mathsf{Event} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula} \\ terminates : \mathsf{Event} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula} \\ prior : \mathsf{Moment} \times \mathsf{Moment} \to \mathsf{Formula} \end{cases}$$

$$t ::= x : S \mid c : S \mid f(t_1, \ldots, t_n)$$

$$\phi ::= \begin{cases} t : \mathsf{Formula} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \phi \to \psi \mid \phi \leftrightarrow \psi \\ \mathbf{P}(a,t,\phi) \mid \mathbf{B}(a,t,\phi) \end{cases}$$

The above calculus lets us formalize statements of the form *"John believes now that Mary perceived that it was raining."* One formalization could be:

$$\exists t < now : \mathbf{B}\Big(john, now, \mathbf{P}\big(mary, t, holds(raining, t)\big)\Big)$$

The figure below shows the inference schemata for $\mu\mathcal{C}$. $R_{\mathbf{P}}$ captures that perceptions get turned into beliefs. $R_{\mathbf{B}}$ is an inference schema that lets us model idealized agents that have their beliefs closed under the $\mu\mathcal{C}$ proof theory. While normal humans are not deductively closed, this lets us model more closely how deliberate agents such as organizations and more strategic actors reason. Assume that there is a background set of axioms Γ we are working with.

Inference Schemata

$$\frac{\mathbf{P}(a,t_1,\phi_1),\ \ \Gamma \vdash t_1 < t_2}{\mathbf{B}(a,t_2,\phi)}\ [R_{\mathbf{P}}]$$

$$\frac{\mathbf{B}(a,t_1,\phi_1),\ldots,\mathbf{B}(a,t_m,\phi_m),\ \ \{\phi_1,\ldots,\phi_m\} \vdash \phi,\ \ \Gamma \vdash t_i < t}{\mathbf{B}(a,t,\phi)}\ [R_{\mathbf{B}}]$$

4 The Uncertainty System $\mathcal{S}$

In the uncertainty system, we augment the belief modal operators with a discrete set of uncertainty factors termed as *strength factors*. The factors are not arbitrary and are based on how derivable a proposition is for a given agent.

Chisholm's epistemology has a primitive undefined binary relation that he terms **reasonableness** with which he defines a scale of strengths for beliefs one might have in a proposition. Note that Chisholm's system is agent free while ours is agent-based. Let $\phi \succ_t^a \psi$ denote that ϕ is more reasonable than ψ to an agent a at time t. We require that $\succ_t^a$ be *asymmetric*: i.e., *irreflexive* and *anti-symmetric*. That is, for all ϕ, $\phi \not\succ_t^a \phi$; and for all ϕ and ψ,

$$(\phi \succ_t^a \psi) \Rightarrow (\psi \not\succ_t^a \phi)$$

We also require that $\succ_t^a$ be transitive. In addition to these conditions, we have the following five requirements governing how $\succ_t^a$ interacts with the logical connectives $\wedge, \neg$ and **B** (the first three conditions can be derived from the defintion of $\succ$ sketched out later):

$[\mathbf{C}_{\wedge_1}]$ $\big(\psi_1 \succ_t^a \phi_1\big)$ and $\big(\psi_2 \succ_t^a \phi_2\big) \Rightarrow \Big(\psi_1 \succ_t^a \phi_1 \wedge \phi_2\Big)$

$[\mathbf{C}_{\wedge_2}]$ $\big(\psi_1 \wedge \psi_2 \succ_t^a \phi\big) \Rightarrow \Big[\big(\psi_1 \succ_t^a \phi\big)$and $\big(\psi_2 \succ_t^a \phi\big)\Big]$

$[\mathbf{C}_{\neg}]$ There is no ϕ such that $(\bot \succ_t^a \phi)$; and for all $\phi (\phi \succ_t^a \bot)$

$[\mathbf{C}_{\mathbf{B}_1}]$ $\Big(\mathbf{B}(a,t,\phi) \succ_t^a \mathbf{B}(a,t,\neg\phi)\Big) \Rightarrow \Big(\mathbf{B}(a,t,\phi) \succ_t^a \neg\mathbf{B}(a,t,\phi)\Big)$

$[\mathbf{C}_{\mathbf{B}_2}]$ For all ϕ, $\begin{bmatrix} \big(\mathbf{B}(a,t,\phi) \succ_t^a \mathbf{B}(a,t,\neg\phi)\big) \text{ or} \\ \big(\mathbf{B}(a,t,\neg\phi) \succ_t^a \mathbf{B}(a,t,\phi)\big) \end{bmatrix}$

We also add a **belief consistency condition** which requires that:

$$\Big(\Gamma \vdash \mathbf{B}^p(a,t,\phi)\Big) \Leftrightarrow \Big(\Gamma \nvdash \mathbf{B}^p(a,t,\neg\phi)\Big)$$

For convenience, we define a new operator, the withholding operator **W** (this is simply syntactic sugar):

$$\mathbf{W}(a,t,\phi) \equiv \neg\mathbf{B}(a,t,\phi) \wedge \neg\mathbf{B}(a,t,\neg\phi)$$

We now reproduce Chisholm's system below. Note the formula used in the definitions below are meta-formula and not strictly in $\mu\mathcal{C}$.

Strength Factor Definitions

Acceptable An agent a at time t finds ϕ acceptable *iff* withholding ϕ is not more reasonable than believing in ϕ.

$$\mathbf{B}^1(a,t,\phi) \Leftrightarrow \begin{cases} \mathbf{W}(a,t,\phi) \not\succ_t^a \ \mathbf{B}(a,t,\phi); \text{ or} \\ \Big(\neg\mathbf{B}(a,t,\phi) \wedge \neg\mathbf{B}(a,t,\neg\phi)\Big) \not\succ_t^a \ \mathbf{B}(a,t,\phi) \end{cases}$$

Some Presumption in Favor An agent a at time t has some presumption in favor of ϕ *iff* believing ϕ at t is more reasonable than believing $\neg\phi$ at time t:

$$\mathbf{B}^2(a,t,\phi) \Leftrightarrow \Big(\mathbf{B}(a,t,\phi) \ \succ_t^a \ \mathbf{B}(a,t,\neg\phi)\Big)$$

Beyond Reasonable Doubt An agent a at time t has beyond reasonable doubt in ϕ *iff* believing ϕ at t is more reasonable than withholding ϕ at time t:

$$\mathbf{B}^3(a,t,\phi) \Leftrightarrow \begin{cases} \mathbf{B}(a,t,\phi) \ \succ_t^a \ \mathbf{W}(a,t,\phi); \text{ or} \\ \Big(\mathbf{B}(a,t,\phi) \succ_t^a \Big(\neg\mathbf{B}(a,t,\phi) \wedge \neg\mathbf{B}(a,t,\neg\phi)\Big)\Big) \end{cases}$$

Evident A formula ϕ is evident to an agent a at time t *iff* ϕ is beyond reasonable doubt and if there is a ψ such that believing ψ is more reasonable for a at time t than believing ϕ, then a is certain about ψ at time t.

$$\mathbf{B}^4(a,t,\phi) \Leftrightarrow \begin{cases} \mathbf{B}^3(a,t,\phi) \wedge \\ \exists\psi : \begin{bmatrix} \mathbf{B}(a,t,\psi) \succ_t^a \mathbf{B}(a,t,\phi) \\ \Rightarrow \mathbf{B}^5(a,t,\psi) \end{bmatrix} \end{cases}$$

Certain An agent a at time t is certain about ϕ *iff* ϕ is beyond reasonable doubt and there is no ψ such that believing ψ is more reasonable for a at time t than believing ϕ.

$$\mathbf{B}^5(a,t,\phi) \Leftrightarrow \begin{cases} \mathbf{B}^3(a,t,\phi) \wedge \\ \neg\exists\psi : \mathbf{B}(a,t,\psi) \succ_t^a \mathbf{B}(a,t,\phi) \end{cases}$$

The above definitions are from Chisholm but more rigorously formalized in $\mu\mathcal{C}$. The definitions and the conditions $\{[\mathbf{C}_{\wedge_1}], [\mathbf{C}_{\wedge_2}, [\mathbf{C}_{\neg}], [\mathbf{C}_{\mathbf{B}_1}], [\mathbf{C}_{\mathbf{B}_1}]\}$ give us the following theorem.

Theorem: Higher Strength subsumes Lower Strength

For any p and q, if $p > q$, we have: $\mathbf{B}^p(a,t,\phi) \Rightarrow \mathbf{B}^q(a,t,\phi)$

Proof: $\mathbf{B}^5 \Rightarrow \mathbf{B}^3$ and $\mathbf{B}^4 \Rightarrow \mathbf{B}^3$ by definition. $\mathbf{B}^5 \Rightarrow \mathbf{B}^4$ by the second clause in the definitions of $\mathbf{B}^4$ and $\mathbf{B}^5$. $\mathbf{B}^3 \Rightarrow \mathbf{B}^1$ by the asymmetry property of $\succ_t^a$.

For $\mathbf{B}^2 \Rightarrow \mathbf{B}^1$, we have a proof by contradiction. Assume that (in shorthand):

$$(\mathbf{B}\phi \succ \mathbf{B}\neg\phi) \text{ but } (\neg\mathbf{B}\phi \wedge \neg\mathbf{B}\neg\phi) \succ \mathbf{B}\phi$$

Using $[\mathbf{C}_{\mathbf{B}_1}]$ on the former and $[\mathbf{C}_{\wedge_2}]$ on the latter, we get

$$\mathbf{B}\phi \succ \neg\mathbf{B}\phi \text{ and } \neg\mathbf{B}\phi \succ \mathbf{B}\phi$$

Using transitivity, we get $\mathbf{B}\phi \succ \mathbf{B}\phi$. This violates irreflexivity, therefore $\mathbf{B}^2 \Rightarrow \mathbf{B}^1$.

For $\mathbf{B}^3 \Rightarrow \mathbf{B}^2$, if the condition for $\mathbf{B}^2$ does not hold, by $\mathbf{C}_{\mathbf{B}_2}$ we have:

$$\mathbf{B}\neg\phi \succ \mathbf{B}\phi$$

Using the condition for $\mathbf{B}^3$ and transitivity, we get

$$\mathbf{B}\neg\phi \succ \neg\mathbf{B}\phi \wedge \neg\mathbf{B}\neg\phi$$

giving us $\mathbf{B}^3\neg\phi$, and we started with $\mathbf{B}^3\phi$. This violates the belief consistency condition. ■

The definitions almost give us $\mathcal{S}$ except for the fact that $\succ_t^a$ is undefined. While Chisholm gives a careful and informal analysis of the relation, he does not provide a more precise definition. Such a definition is needed for automation. We provide a three clause defintion that is based on both probabilities and proof theory.

There are many probability logics that allow us to define probabilities over formulae. They are well studied and understood for propositional and first-order logics. Let $\mathcal{L}$ be the set of all formulae in $\mu\mathcal{C}$. Let $\mathcal{L}_p$ be a pure first-order subset of $\mathcal{L}$. Assume that we have the following partial probability function defined over $\mathcal{L}_p$[5]:

$$\mathbf{Pr} : \mathsf{Agent} \times \mathsf{Moment} \times \mathsf{Formula} \mapsto \mathbb{R}$$

Then we have the first clause of our definition for $\succ_t^a$.

Clause I. Defining $\succ$

If $\mathbf{Pr}(a,t,\phi)$ and $\mathbf{Pr}(a,t,\psi)$ are defined then:

$$\Big(\phi \ \succ_t^a \ \psi\Big) \Leftrightarrow \Big(\mathbf{Pr}(a,t,\phi) > \mathbf{Pr}(a,t,\psi)\Big)$$

We might not always have meaningful probabilities for all propositions. For example, consider propositions of the form *"I believe that Jack believes that* ϕ*."* It is hard to get precise numbers for such statements. In such situations, we might look at the ease of derivation of such statements given a knowledge base Γ. [6] Given two competing statements ϕ and ψ, we can say one is more reasonable than the other if we can easily derive one more than the other from Γ. This assumes that we can derive ϕ and ψ from Γ. We assume we have a cost function $\rho : \mathsf{Proof} \mapsto \mathbb{R}^+$ that lets us compute costs of proofs. There are many ways of specifying such functions. Possible candidates are length of the proof, time for computing the proof, depth vs breadth of the proof, unique symbols used in the proof etc. We leave this choice unspecified but any such function could work here. Let $\vdash_{a,t}$ denote provability w.r.t. to agent a at time t.

[5] Something similar to the system in [Halpern, 1990] that accounts for probabilities as statistical information or degrees of belief can work.

[6] Another possible mechanism can leverage Dempster-Shafer models of uncertainty for first-order logic [Nunez *et al.*, 2013].

Clause II. Defining $\succ$

If one of $\mathbf{Pr}(a,t,\phi)$ and $\mathbf{Pr}(a,t,\psi)$ is not defined, but if $\Gamma \vdash_{a,t} \phi$ and $\Gamma \vdash_{a,t} \psi$:

$$\Big(\phi \succ_t^a \psi\Big) \Leftrightarrow \Big(\rho\big(\Gamma \vdash_{a,t} \phi\big) < \rho\big(\Gamma \vdash_{a,t} \psi\big)\Big)$$

Clauses I and II might not always be applicable as the premises in the definitions might not always hold. A more common case could be when we cannot derive the propositions of interest from our background set of axioms Γ. For example, if we are interested in the uncertainty values for statements that we know are false, then it should be the case that they be not derivable from our background set of axioms. In this situation, we look at Γ and see what minimal changes we can make to it to let us derive the proposition of interest. Trivially, if we cannot derive ϕ from Γ, we can add it to Γ to derive it, as $\Gamma + \phi \vdash \phi$. This is not desirable for two reasons.

First, simply adding to Γ might result in a contradiction. In such cases we would be looking at removing a minimal set of statements Λ from Γ. Second, we might prefer to add a more simpler set of propositions Θ to Γ rather than ϕ itself to derive ϕ. Recapping, we go from (1) to (2) below:

$$\Gamma \nvdash \phi \quad (1)$$
$$\Gamma \cup \Theta - \Lambda \vdash \phi \quad (2)$$

When we go from (1) to (2) we would like to modify the background axioms as minimally as possible. Assume that we have a similarity function π for sets of formulae. We then choose Θ and Λ as given below ($Con[S]$ denotes that S is consistent):

$$\langle \Theta, \Lambda \rangle = \underset{\langle \Theta, \Lambda \rangle}{\arg\min}\, \pi(\Gamma, \Gamma \cup \Theta - \Lambda); \text{ such that } \begin{cases} \Gamma \cup \Theta - \Lambda \vdash \phi; \text{ and} \\ Con\big[\Gamma \cup \Theta - \Lambda\big] \end{cases}$$

Consider a statement such as *"It rained last week"* when it did not actually rain last week, and another statement such as *"The moon is made of cheese."* Both statements denote things that did not happen, but intuitively it seems that former should be more easier to accept from what we know than the latter. There are many similarity measures which can help convey this. Analogical reasoning is one such possible measure of similarity. If the new formulae are structurally similar to existing formulae, then we might be more justified in accepting such formulae. For example, one such measure could be the analogical measure used by us in [Licato *et al.*, 2013].

Now we have the formal mechanism in place for defining the final clause in our definition for our reasonableness. Let $\delta_t^a(\Gamma, \phi)$ be the distance between Γ and closest consistent set under π that lets us prove ϕ:

$$\delta_t^a(\Gamma, \phi) \equiv \min_{\langle \Theta, \Lambda \rangle} \left\{ \pi\big(\Gamma, \Gamma \cup \Theta - \Lambda\big) \,\middle|\, \begin{array}{l} \big(\Gamma \cup \Theta - \Lambda\big) \vdash_t^a \psi; \text{ and} \\ Con\big[\Gamma \cup \Theta - \Lambda\big] \end{array} \right\}$$

Clause III. Defining $\succ$

If one of $\mathbf{Pr}(a,t,\phi)$ and $\mathbf{Pr}(a,t,\psi)$ is not defined, and one of $\Gamma \vdash_{a,t} \phi$ and $\Gamma \vdash_{a,t} \psi$ does not hold, then

$$\Big(\phi \succ_t^a \psi\Big) \Leftrightarrow \Big[\delta_t^a(\Gamma, \phi) < \delta_t^a(\Gamma, \psi)\Big]$$

The final piece of $\mathcal{S}$ is inference rules for belief propagation with uncertainty values. This is quite straightforward. Inferences propagate uncertainty values from the premises with the lowest strength factor; and inferences happen only with beliefs that are close in their uncertainty values, with maximum difference being parametrized by u, with default $u = 2$.

Inference Schemata for $\mathcal{S}$

$$\frac{\mathbf{P}(a,t_1,\phi_1),\ \Gamma \vdash t_1 < t_2}{\mathbf{B}^5(a,t_2,\phi)}\ [R_{\mathbf{P}}^s]$$

$$\frac{\mathbf{B}^{s_1}(a,t_1,\phi_1),\ldots,\mathbf{B}^{s_m}(a,t_m,\phi_m),\{\phi_1,\ldots,\phi_m\} \vdash \phi, \Gamma \vdash t_i < t}{\mathbf{B}^{min(s_1,\ldots,s_m)}(a,t,\phi)}\ [R_{\mathbf{B}}^s]$$
$$\text{with} \quad max(\{s_1,\ldots,s_m\}) - min(\{s_1,\ldots,s_m\}) \leq u$$

5 Usage

In this section, we illustrate $\mathcal{S}$ by applying it solve problems of foundational interest such as the lottery paradox [Kyburg Jr, 1961, p. 197] and a toy version of a more real life example, a murder mystery example (following in the traditions of logic pedagogy). Finally, we very briefly sketch abstract scenarios in which $\mathcal{S}$ can be used to generate uncertainty values for counterfactual statements and to generate explanations for actions.

5.1 Paradoxes: Lottery Paradox

In the lottery paradox, we have a situation in which an agent a comes to believe ϕ and $\neg\phi$ from a seemingly consistent set of premises Γ_L describing a lottery. Our solution to the paradox is that the agent simply has different strengths of beliefs in the proposition and its negation. We first go over the paradox formalized in $\mu\mathcal{C}$ and then present the solution.

Let Γ_L be a meticulous and perfectly accurate description of a 1,000,000,000,000-ticket lottery, of which rational agent a is fully apprised. Assume that from Γ_L it can be proved that either ticket 1 will win or ticket 2 will win or ... or ticket 1,000,000,000,000 will win. Lets write this (exclusive) disjunction as follows (here $\oplus$ is an exclusive disjunction):

$$\Gamma_L \vdash win(t_1) \oplus win(t_2) \oplus \ldots \oplus win(t_{1,000,000,000,000})$$

The paradox has two strands of reasoning. The first strand yields $\mathbf{B}(a, now, \phi)$ and the second strand yields $\mathbf{B}(a, now, \neg\phi)$ with $\phi \equiv \exists t : win(t)$.

Strand 1: Since a believes all propositions in Γ_L, a can then deduce from this the belief that there is at least one ticket that will win, a proposition represented as:

$$\boxed{\mathrm{S}_1} \quad \mathbf{B}\Big(a, now, \exists t : win(t)\Big)$$

Strand 2: From Γ_L it can be proved that the probability of a particular ticket t_i winning is 10^{-12}.

$$\Big[\mathbf{Pr}\Big(a, now, win\big(t_1\big)\Big) = 10^{-12}\Big] \wedge \Big[\mathbf{Pr}\Big(a, now, win\big(t_2\big)\Big) = 10^{-12}\Big] \wedge \ldots \wedge \Big[\mathbf{Pr}\Big(a, now, win\big(t_{1T}\big)\Big) = 10^{-12}\Big]$$

For the next step, note that the probability of ticket t_1 winning is lower than, say, the probability that if you walk outside a minute from now, you will be flattened on the spot by a door from a 747 that falls from a jet of that type cruising at 35,000 feet. Since you, the reader, have the rational belief that death won't ensue if you go outside (and have this belief precisely because you believe that the odds of your sudden demise in this manner are vanishingly small), the inference to the rational belief on the part of a that t_1 won't win sails through — and this of course works for each ticket. Hence we have as a valid belief (though not derivable in $\mu\mathcal{C}$ from Γ_L):

$$\mathbf{B}\Big(a, now, \neg win\big(t_1\big)\Big) \wedge \mathbf{B}\Big(a, now, \neg win\big(t_2\big)\Big) \wedge \ldots \wedge \mathbf{B}\Big(a, now, \neg win\big(t_{1T}\big)\Big)$$

From $R_{\mathbf{B}}$ and above, we get:

$$\mathbf{B}\Big(a, now, \neg win\big(t_1\big) \wedge \neg win\big(t_2\big) \wedge \ldots \wedge \neg win\big(t_{1T}\big)\Big)$$

Applying $R_{\mathbf{B}}$ to the above and Γ_L, we get:

$$\boxed{\mathrm{S}_2} \quad \mathbf{B}\Big(a, now, \neg\exists t : win(t)\Big)$$

The two strands are complete, and we have derived contradictory beliefs labeled S_1 and S_2. Our solution consists of two new uncertainty infused strands that result in beliefs of sufficiently varying strengths that block inferences that could combine them.

Strand 1 and **Strand 2** demonstrate the standard informal reasoning which leads to the paradox. We replicate the reasoning in $\mu\mathcal{C}$ and show that the paradox is not derivable.

Strand 3: Assume that a is certain of all propositions in Γ_L, then using $R^s_{\mathbf{B}}$, we have:

$$\boxed{\mathrm{S}_3} \quad \mathbf{B}^5\big(a, now, \exists t : win(t)\big)$$

Strand 4: Since $\mathbf{Pr}(a, now, win(t_i)) < \mathbf{Pr}(a, now, \neg win(t_i))$, using Clause I and the strength factor definitions, we have now that for all t_i

$$\mathbf{B}^2\big(a, now, \neg win(t_i)\big)$$

Using the reasoning similar to that in Strand 2, we get:

$$\boxed{\mathrm{S}_4} \quad \mathbf{B}^2\Big(a, now, \neg\exists t : win(t)\Big)$$

Strands 3 and 4 resolve the paradox. Note that $R^s_{\mathbf{B}}$ cannot be applied to S_3 and S_4 and churn out arbitrary propositions, as the default value of the u parameter in $R^s_{\mathbf{B}}$ requires beliefs to be no more than 2 levels apart. ■

5.2 Application: Solving a Murder

We look at a toy example in which an agent s has to solve a murder that happened at time t_3. s believes that either Alice or Bob is the murderer. The agent knows that there is a gun gun involved in the murder and that the owner of the gun at t_3 committed the murder. s also knows that Alice is the owner of the gun initially at time t_0.

Presumption in Favor of Alice Being the Murderer

From just these facts, the agent has some presumption for believing that Alice is the murderer.

Proof Sketch: All the above statements can be taken as certain beliefs $\mathbf{B}^5$ of s. For convenience, we consider the formulae directly without the belief operators.

In order to prove the above, we need to prove that it is easier for the agent to derive that Alice is the murderer than to derive that Alice is not the murderer. First, to prove the former, the agent just has to assume that Alice's ownership of the gun did not change from t_0 to t_3. Second, in order for the agent to believe that Alice did not commit the murder but Bob committed it, the agent must be willing to admit that something happened to change Alice's ownership of the gun from time t_0 to t_3 that results in Bob owning the gun. One possibility is that Alice simply sold the gun to Bob. Both the scenarios are shown as proofs in the Slate theorem proving workspace [Bringsjord *et al.*, 2008] in the Appendix. Figure 1 shows a proof modulo belief operators of $\mathbf{B}(s, now, Murderer(\mathsf{Alice}))$ from $\Gamma \cup \Theta_1$ and Figure 2 shows a proof of $\mathbf{B}(s, now, \neg Murderer(\mathsf{Alice}))$ from $\Gamma \cup \Theta_2$.

If we assume that Θ_1 and Θ_2 exhaust the space of allowed additions, then it easy to see how syntactic measures of complexity will yield that $\delta^a_t\big(\Gamma, \Gamma \cup \Theta_1\big) < \delta^a_t\big(\Gamma, \Gamma \cup \Theta_2\big)$ as Θ_2 is more complex than Θ_1. This lets us derive that s has some presumption in favor of $Murderer(\mathsf{Alice})$. ■

What happens if the agent knows or has a belief with certainty that Alice's ownership of the gun did not change from t_0 to t_3?

Beyond Reasonable Doubt that Alice is the Murderer

If the agent is certain that Alice's ownership of the gun did not change from t_0 till t_3, the agent has beyond reasonable doubt that she is the murderer.

Proof Sketch: In this case we directly have that:

$$\Gamma \vdash \mathbf{B}(s, now, Murderer(\mathsf{Alice}))$$
$$\Gamma \nvdash \neg\mathbf{B}(s, now, Murderer(\mathsf{Alice}))$$
$$\Gamma \nvdash \neg\mathbf{B}(s, now, \neg Murderer(\mathsf{Alice}))$$

In order to flip the last two statements above, we need to modify Γ, but we can derive that Alice is the murderer without any modifications, and since $\delta^a_t(\Gamma, \Gamma) = 0$, it easier to believe Alice is the murderer than to withhold that Alice is the murderer. ■

5.3 Counterfactuals

At time t, assume that an agent a believes in a set of propositions Γ and is interested in propositions $holds(f, t')$ and $holds(g, t')$ with $t' < t$ and:

$$\Gamma \vdash \neg holds(f, t') \wedge \neg holds(g, t')$$

We may need non-trivial uncertainty values, but in this case, **Pr** will assign a trivial value of 0 to both the propositions. We can then look at closest consistent sets to Γ under δ:

$$\Gamma_1 \vdash holds(f,t')$$
$$\Gamma_2 \vdash holds(g,t')$$

Clause III from the definition for reasonableness gives us:

$$\mathbf{B}\Big(a,t,holds(f,t')\Big) \succ_t^a \mathbf{B}\Big(a,t,holds(g,t')\Big)$$
$$\Leftrightarrow$$
$$\delta_t^a(\Gamma,\Gamma_1) < \delta_t^a(\Gamma,\Gamma_2)$$

5.4 Explanations

The definitions of the strength factors and reasonableness above can be used to generate high-level schemas for explanations. These schemas can be used instead of simply presenting raw probability values to end-users. While we have not fleshed out such explanation schemas, we illustrate one possible schema. Say an agent performs an action α on the basis of ϕ. In this case, the agent could generate an explanation that at the highest level simply says that it is more reasonable for the agent to believe ϕ than for the agent to believe in $\neg\phi$. The agent could then further explain why it was reasonable for it by using one of the three clauses in the reasonableness definition.

6 Inference Algorithm Sketch

Describing the inference algorithm in detail is beyond the scope of this paper, but we provide a high-level sketch here.[7] Our proof calculus is simply an extension of standard first-order proof calculus under different modal contexts. For example, if a believes that b believes in a set of propositions Γ and $\Gamma \vdash_{FOL} \psi$, then a believes that b believes ψ. We convert $\mathbf{B}(a,t_a,\mathbf{B}(b,t_b,Q))$ into the pure first-order formula $Q(context(a,t_a,b,t_b))$ and use a first-order prover. The conversion process is a bit more nuanced as we have to handle negations, properly handle substitutions of equalities, uncertainties and transform compound formulae within iterated beliefs.

7 Conclusion and Future Work

We have presented initial steps in building a system of uncertainty that is both probability and proof theory based that could lend itself to (1) solving foundational problems; (2) being useful in applications; (3) generating uncertainty values for counterfactuals; and (4) building understandable explanations.

Shortcomings of $\mathcal{S}$ can be cast as challenges, and many challenges exist, some relatively easy and some quite hard. Among the easy challenges are defining and experimenting with different candidates for **Pr**, ρ, π and δ. On the more difficult side, we have to come up with tractable computational mechanisms for computing the $\min_{\langle\Theta,\Lambda\rangle}$ in the definition for δ. Also on the difficult side, is the challenge of coming up efficient reasoning schemes. While we have an exact inference algorithm, we believe that an approximate algorithm that selectively discards beliefs in a large knowledge base during reasoning will be more useful. Future work also includes comparison with other uncertainty systems and exploration of conditions under which uncertainty values of $\mathcal{S}$ are similar/dissimilar with other systems (thresholded appropriately).

[7] More details can be found here: https://goo.gl/2Vz2nJ

Acknowledgements

We are grateful to the Office of Naval Research for their funding of projects titled *"Advanced Logicist Machine Learning"* and *"Making Morally Competent Robots"* and to the Air Force Office of Scientific Research for funding the project titled *"Great Computation Intelligence: Mature and Further Applied"* that enabled the research presented in this paper. We are also thankful for the insightful reviews provided by the three anonymous referees.

A Appendix: Slate Proofs

The figures below are vector graphics and can be zoomed to more easily read the contents.

Figure 1: Alice is the murder: $\mathbf{B}\Big(s,t,Murderer(\text{Alice})\Big)$

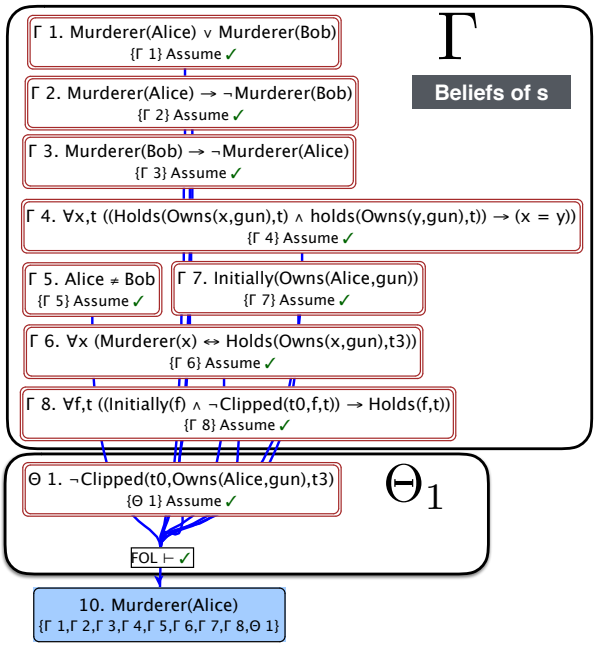
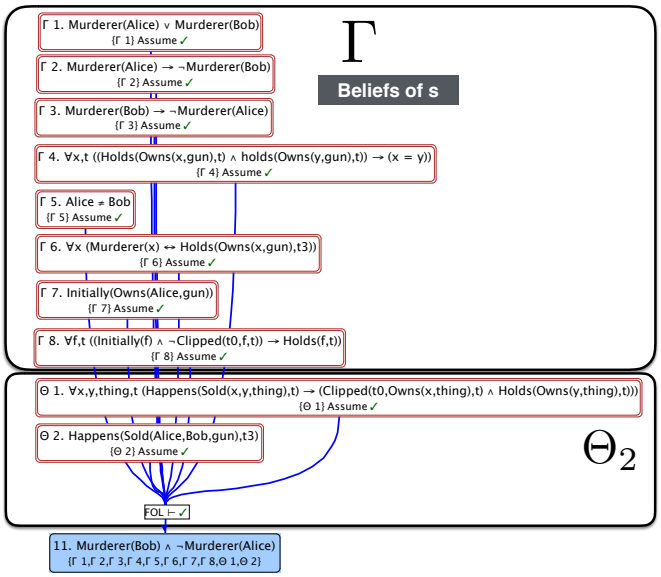


Figure 2: Alice is not the murder: $\mathbf{B}\Big(s, t, \neg \mathit{Murderer}(\mathrm{Alice})\Big)$



References

[Arkoudas and Bringsjord, 2008] Konstantine Arkoudas and Selmer Bringsjord. Toward Formalizing Common-Sense Psychology: An Analysis of the False-Belief Task. In T.-B. Ho and Z.-H. Zhou, editors, *Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2008)*, number 5351 in Lecture Notes in Artificial Intelligence (LNAI), pages 17–29. Springer-Verlag, 2008.

[Bringsjord *et al.*, 2008] Selmer Bringsjord, Joshua Taylor, Andrew Shilliday, Micah Clark, and Konstantine Arkoudas. Slate: An Argument-Centered Intelligent Assistant to Human Reasoners. In Floriana Grasso, Nancy Green, Rodger Kibble, and Chris Reed, editors, *Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)*, pages 1–10, Patras, Greece, July 21 2008. University of Patras.

[Bringsjord *et al.*, 2014] Selmer Bringsjord, Naveen Sundar Govindarajulu, Daniel Thero, and Mei Si. Akratic Robots and the Computational Logic Thereof. In *Proceedings of ETHICS • 2014 (2014 IEEE Symposium on Ethics in Engineering, Science, and Technology)*, pages 22–29, Chicago, IL, 2014. IEEE Catalog Number: CFP14ETI-POD.

[Chisholm, 1987] Roderick Chisholm. *Theory of Knowledge 3rd ed*. Prentice-Hall, Englewood Cliffs, NJ, 1987.

[Demey *et al.*, 2016] Lorenz Demey, Barteld Kooi, and Joshua Sack. Logic and Probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[Gentzen, 1935] Gerhard Gentzen. Investigations into Logical Deduction. In M. E. Szabo, editor, *The Collected Papers of Gerhard Gentzen*, pages 68–131. North-Holland, Amsterdam, The Netherlands, 1935. This is an English version of the well-known 1935 German version.

[Govindarajulu and Bringsjord, 2017] Naveen Sundar Govindarajulu and Selmer Bringsjord. On Automating the Doctrine of Double Effect. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 2017. Preprint available at this url: https://arxiv.org/abs/1703.08922.

[Halpern, 1990] Joseph Y Halpern. An Analysis of First-order Logics of Probability. *Artificial intelligence*, 46(3):311–350, 1990.

[Kaye and Koehler, 1991] D. H. Kaye and Jonathan J. Koehler. Can Jurors Understand Probabilistic Evidence? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154(1):75–81, 1991.

[Kyburg Jr, 1961] Henry E Kyburg Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT, 1961.

[Licato *et al.*, 2013] John Licato, Naveen Sundar Govindarajulu, Selmer Bringsjord, Michael Pomeranz, and Logan Gittelson. Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox. In Francesca Rossi, editor, *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI–13)*, pages 1004–1009, Beijing, China, 2013. Morgan Kaufmann.

[Nunez *et al.*, 2013] Rafael C. Nunez, Matthias Scheutz, Kamal Premaratne, and Manohar N. Murthi. Modeling Uncertainty in First-Order Logic: A Dempster-Shafer Theoretic Approach. In *8th International Symposium on Imprecise Probability: Theories and Applications*, 2013.

[Zalta, 1988] Edward N Zalta. *Intensional Logic and the Metaphysics of Intentionality*. MIT Press, Cambridge, MA, 1988.