Novel Intensional Defeasible Reasoning for AI: Is it Cognitively Adequate?

Michael Giancola[1], Selmer Bringsjord[1] and Naveen Sundar Govindarajulu[1]

[1]*Rensselaer AI & Reasoning (RAIR) Lab, Department of Computer Science, Department of Cognitive Science, Rensselaer Polytechnic Institute (RPI), Troy NY 12180, United States*

1. Extended Abstract

The importance of defeasible (or nonmonotonic) reasoning has long been recognized in AI, and proposed ways of formally modeling and computationally simulating such non-deductive reasoning via logics and automated reasoning go back to early, seminal work in the field. But from that time to now, logic-based AI has not produced a logic, with associated automation, that handles defeasible reasoning suffused with arbitrarily iterated *intensional* operators like *believes*, *knows*, etc. We present a novel logic-based approach for solving defeasible reasoning problems that demand intensional operators and reasoning. We exploit two central problems. The first is the "Nixon Diamond," (ND) [1] a simple but illuminating specimen in defeasible-reasoning research in AI. We show how the contradiction inherent in ND can be resolved by constructing two arguments — corresponding to the two branches of the Diamond — one of which "defeats" the other. The solution is found by enabling reasoning about the agent's beliefs regarding the *context* of the Diamond's assertions. Such reasoning about beliefs inherently requires an intensional logic. Our second problem is a variant of a much-studied and deeper one from cognitive science: Byrne's "Suppression Task" (ST) [2]. We present a challenging new version of ST that is explicitly and unavoidably intensional — and then show that our new AI approach can meet this challenge. We thus claim that our approach is "AI adequate" — but hold that it is not *cognitively* adequate until empirical experiments in cognitive science, run with relevant classes of subjects, align with what our AI approach yields. The rest of this extended abstract will present a high-level overview of both the mechanisms we use to solve the two problems — namely, *cognitive likelihood calculi* — and the solutions themselves.

1.1. Cognitive Likelihood Calculi

While a full discussion of the technical specifications of cognitive calculi, let alone cognitive *likelihood* calculi, is impossible due to space constraints, we summarize their key attributes. A

Cognitive Aspects of Knowledge Representation: An IJCAI 22 Workshop, July 23–29, 2022, Vienna, Austria
mike.j.giancola@gmail.com (M. Giancola); selmer.bringsjord@gmail.com (S. Bringsjord); naveen.sundar.g@gmail.com (N. S. Govindarajulu)
https://mjgiancola.github.io/ (M. Giancola); https://www.rpi.edu/~brings (S. Bringsjord); http://www.naveensundarg.com/ (N. S. Govindarajulu)

cognitive calculus is a multi-operator quantified intensional logic with modal operators that capture cognitive attitudes of human cognition (e.g. $\mathbf{K}$ for "knows", $\mathbf{B}$ for "believes"). For the purposes of this paper, a cognitive calculus consists essentially of two components: (1) multi-sorted first-order logic with intensional/modal operators for modeling cognitive attitudes and (2) inference schemata that — in the tradition of proof-theoretic semantics [3] — fully express the semantics of the modal operators.

A cognitive *likelihood* calculus additionally includes an uncertainty sub-system in order to ascribe likelihoods to formulae. That is, such a calculus must contain syntactic forms and inference schemata which dictate the ways in which likelihoods can be associated with formulae and how they can be used and propagated in proofs. Note that since formulae can in this approach have a relative "strength" σ of certainty, cognitive likelihood calculi necessarily cannot be purely deductive, but are instead *inductive*.

Syntax

$$\phi ::= \{\neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \phi \rightarrow \psi \mid \forall x : \phi(x) \mid \exists x : \phi(x) \mid \mathbf{K}(a, \phi) \mid \mathbf{B}^{\sigma}(a, \phi)$$
$$\text{where } \sigma \in [-6, -5, \ldots, 5, 6]$$

Inference Schemata

$$\frac{\mathbf{K}(a, \phi)}{\phi}\ [I_{\mathbf{K}}] \qquad \frac{\mathbf{B}^{\sigma_1}(a, \phi_1), \ldots, \mathbf{B}^{\sigma_m}(a, \phi_m), \{\phi_1, \ldots, \phi_m\} \vdash \phi, \{\phi_1, \ldots, \phi_m\} \nvdash \bot}{\mathbf{B}^{min(\sigma_1, \ldots, \sigma_m)}(a, \phi)}\ [I_{\mathbf{B}}]$$

The syntax of the calculus used herein subsumes first-order logic; it additionally contains modal operators for knowledge $\mathbf{K}$ and uncertain belief $\mathbf{B}^{\sigma}$. The first schema, $[I_{\mathbf{K}}]$, says that if an agent a knows a formula ϕ, then ϕ must hold. The second schema, $[I_{\mathbf{B}}]$, says that if an agent a holds an arbitrary number of beliefs about formulae ϕ_1 to ϕ_m with corresponding strengths σ_1 to σ_m, then a can infer a belief in anything provable from those beliefs, with two restrictions: First, the beliefs cannot prove a contradiction. The second restriction is that the strength of the inferred belief must be at the level of the weakest belief used to infer it.[1]

Finally, in this calculus, beliefs can take on 13 possible likelihood values. The following are the descriptors of the non-negative[2] likelihood values: CERTAIN (6), EVIDENT (5), OVERWHELMINGLY LIKELY (4), VERY LIKELY (3), LIKELY (2), MORE LIKELY THAN NOT (1), and COUNTERBALANCED (0).

1.2. The Nixon Diamond

The "Nixon Diamond" (ND) is a famous specimen in the AI literature on nonmonotonic/defeasible logic; see Figure 1. ND contains two arguments which seem to directly contradict. First, Nixon was a Quaker, and Quakers are pacifists. Second, Nixon was a Republican, and Republicans are not pacifists. Hence, Nixon is a pacifist and a non-pacifist!

To us, the important question to ask when considering how to solve the diamond is this: What would human reasoners familiar with the concepts involved (e.g. pacifism) and background knowledge (e.g. about Quakerism's core tenets) actually conclude about Nixon? We shall return

[1] This schema is essentially a formalization of *The Weakest Link Principle*.
[2] The negative values are simply the negation of the corresponding positive value.

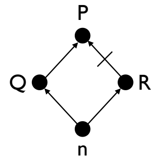

Figure 1: The Nixon Diamond

α_1	α_2
$\mathbf{B}_c^4(\forall x\, Qx \rightarrow Px)$	$\mathbf{B}_c^3(\forall x\, Rx \rightarrow \neg Px)$
$\mathbf{B}_c^4(Qn \rightarrow Pn)$ [∀E]	$\mathbf{B}_c^3(Rn \rightarrow \neg Pn)$ [∀E]
$\mathbf{B}_c^1 Qn$	$\mathbf{B}_c^5 Rn$
$\therefore \mathbf{B}_c^1 Pn$ [MP]	$\therefore \mathbf{B}_c^3 \neg Pn$ [MP]

Table 1: Competing Arguments α_1 and α_2

to this question below; the important point at the moment is that this is a driving question behind the new family of intensional defeasible logics herein introduced.

Our analysis posits a single cognizer, c. Invoking the symbolization introduced in Figure 1, set $\Gamma := \{\forall x(Qx \rightarrow Px), \forall x(Rx \rightarrow \neg Px), Qn, Rn\}$,. Clearly, $\Gamma \vdash Pn \wedge \neg Pn$.

Now, our cognizer's beliefs about the members of Γ are simply determined by (1) inferring new beliefs via the inference schemata of our cognitive likelihood calculus[3] and (2) adjudicating clashes in favor of higher likelihood. Therefore, our cognizer derives the pair of arguments shown in Table 1.

The point here isn't the particular upshot obtained via the likelihood values employed in Table 1, which is that our cognizer ought to believe that Nixon is not a pacifist. It's true that this table is intended to be a "real-world" instantiation, based as it is on background reasoning about the concepts involved.[4] But the point is that, from our AI perspective (uninformed by empirical experiments), what is to be ultimately believed by a first-rate cognizer is based on bringing to bear the key attributes of a cognitive likelihood calculus, in conjunction with relevant information. In particular, with respect to fine-grained arguments, what's *really* going on in the minds of first-rate human reasoners is presumably that, in turn, there are other background fine-grained arguments in play in support of such propositions as that Republicans are non-pacifists. Thus, ultimately, what emerges as the rational belief for a cognizer at a particular time will depend upon the processing of *many* interrelated arguments, and their internal structure. Of course, again, we say this as AI researchers in the *hope* of cognitive adequacy.

1.3. The <u>Intensional</u> Suppression Task

We turn now to the promised variant of the original ST that is more demanding from a logicist-AI perspective, but, at least to us, not much more demanding from a human-reasoning perspective. We cannot review the original Suppression Task here due to space constraints; the interested reader is referred to [2].
In our intrinsically intensional version of the suppression task[5], the three premises are these:

(p1_{int}) If Mary has an essay to finish, then Mary will study late in the library.

[3]Note that, as our cognitive calculus subsumes first-order logic, the inference schemata contain those of first-order logic as well, i.e. the standard introduction & elimination schemata.

[4]Nixon, after his father converted from Methodism to his mother's Quakerism, had two parents who were Quakers, but at most that makes it *more likely than not* or *likely* that *he* was a Quaker. Nixon formally registered as a Republican, making it *evident* that he is in fact one. Furthermore, Quakerism is doctrinally distinguished by pacifism; in contrast, relatively few Republicans identify as pacifists.

[5]Herein we only present an intensional adaptation of one of Byrne's 12 experiments. In the full paper we adapted three of them, but our calculus is fully capable of modeling all 12.

($p2_{int}$) Mary's mother knows that Mary's father knows that Mary has an essay to finish.

($p3_{int}$) If the library stays open, then Mary will study late in the library.

We next have the following three options:

($o1_{int}$) Mary will study late in the library.

($o2_{int}$) Mary will not study late in the library.

($o3_{int}$) Mary may or may not study late in the library.

Now imagine posing this question to a rational human-level agent: Which of these three options logically follow from the three premises? The correct answer is ($o1_{int}$), only. However, Byrne's original experiment found that the addition of ($p3_{int}$) led most human cognizers to suppress the valid inference. It can be expected that most people, as in in Byrne's original experiment, would fail to correctly select ($o1_{int}$) for generally the same reasons they failed to select (o1) in the original ST.[6]

Assume we have a fully rational agent a who is capable of reasoning via our cognitive likelihood calculus. Proving the formal equivalent of ($o1_{int}$) is fairly straightforward, given the established inference schemata:[7]

$$\mathbf{B}^6(\mathfrak{a}, \mathbf{K}(m, \mathbf{K}(f, \text{ToFinish}(mary, essay)))) \quad \text{[Given]}$$
$$\mathbf{B}^6(\mathfrak{a}, \mathbf{K}(f, \text{ToFinish}(mary, essay))) \quad [[I_\mathbf{B}],\ \text{using } [I_\mathbf{K}]]$$
$$\mathbf{B}^6(\mathfrak{a}, \text{ToFinish}(mary, essay))) \quad [[I_\mathbf{B}],\ \text{using } [I_\mathbf{K}]]$$
$$\mathbf{B}^6(\mathfrak{a}, \text{ToFinish}(mary, essay) \rightarrow \text{StudyLate}(mary)) \quad \text{[Given]}$$
$$\mathbf{B}^6(\mathfrak{a}, \text{StudyLate}(mary)) \quad [[I_\mathbf{B}],\ \text{using modus ponens}]$$

1.4. Automated Solutions

Finally, we briefly note that we have an automated reasoner able to automatically generate and adjudicate the arguments presented herein. That automated reasoner, called ShadowAdjudicator, is under active development and is open-source.[8]

Acknowledgments

This research is partially enabled by support from ONR and AFOSR (Award # FA9550-17-1-0191).

References

[1] R. Reiter, A Logic for Default Reasoning, Artificial Intelligence 13 (1980) 81–132.
[2] R. Byrne, Suppressing Valid Inferences with Conditionals, Journal of Memory and Language 31 (1989) 61–83.
[3] N. Francez, Proof-theoretic Semantics, College Publications, London, UK, 2015.

[6]We note that while we intuitively hypothesize that the changes we made to the experiment would not impact the results, our position is that establishing cognitive adequacy would require a new experiment in order to confirm our hypothesis. However, conducting such an experiment is in the expertise of cognitive scientists, not AI researchers. Of course, whatever the result empirically, our logico-mathematics can model and simulate it.

[7]Our cognitive likelihood calculus can also model the *invalid* reasoning undertaken by humans who suppress the valid inference; due to space constraints, it is left to the full paper.

[8]https://github.com/RAIRLab/ShadowAdjudicator