Title
Logicist Computational Cognitive Modeling of Infinitary False Belief Tasks

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Bringsjord, Selmer
Govindarajulu, Naveen Sundar

Publication Date
2019

Peer reviewed

Logicist Computational Cognitive Modeling of *Infinitary* False Belief Tasks

Selmer Bringsjord (selmerbringsjord@gmail.com)
Rensselaer AI & Reasoning Lab, Dept of Cognitive Science, RPI; Troy, NY 12180 USA

Naveen Sundar Govindarajulu (naveensundarg@gmail.com)
Rensselaer AI & Reasoning Lab, Dept of Cognitive Science, RPI; Troy, NY 12180 USA

Abstract

We synoptically describe having achieved the unprecedented logicist cognitive computational simulation of quantified versions of any n-level (FBT_n, $\forall n \in \mathbb{N}$) false-belief task, and hence of what we call the *infinitary* false-belief task (FBT_ω); the achievement is enabled by the automated reasoner ShadowProver. Logicist cognitive computational simulation of the level-one (or, as it's currently known, "first-order") false-belief task (FBT_1) was achieved circa 2007 by Bringsjord et al. But subsequently cognitive science has seen the arrival such modeling and simulation successfully applied to the *second-order* false-belief task (FBT_2); see e.g. (Blackburn & Polyanskaya, forthcoming). (This is the level-two FBT in our hierarchy of tasks.) But now, courtesy of what we report, logicist cognitive computational simulation of any FBT_n is accomplished for the first time, and hence the infinitary false-belief task (FBT_ω) is reached as well.

Keywords: logic; cognitive modeling; false-belief task; sally-anne task; infinitary reasoning

The Level-1 and Level-2 False-Belief Tasks Many readers will be familiar with the standard false-belief task (FBT_1; a.k.a. the Sally-Anne task), first introduced by Wimmer and Perner (1983). But to ensure self-containedness we recapitulate: A subject (in an experiment carried out by e), agent a, perceives two agents a_1 and a_2 in front of two boxes b_1 and b_2. Agent a_1 puts an object o into b_1 in plain view of a_2. Agent a_2 then leaves, and in the absence of a_2, a_1 moves o from b_1 into b_2; this movement isn't perceived by a_2. Agent a_2 now returns, and a is asked by the experimenter e: "If a_2 desires to retrieve o, which box will a_2 look in?" If younger than four or five, a will reply "In b_2" (which of course fails the task); after this age subjects respond with the correct "In b_1." While some refer to this task as the "first-order" version of the false-belief task, we refer to it as the "level-one" version of the task.[1]

Table 1 lists some of the key epistemic propositions that hold of FBT_1^P after the switch happens, paired with their obvious symbolizations in our multi-operator quantified cognitive calculus used for handling false-belief tasks. We use the superscript 'P' to indicate that the task in question is passed; we reserve superscript 'F' to indicate that the task is failed.

[1]Use of the locution "n-order" is quite infelicitous, because this locution is long established in formal logic as a way to pick out the expressive power of extensional logics within a hierarchy of them. For instance, there is first-order logic, second-order logic, and so on. Since which of these logics is used to model and simulate a given false-belief task is a key parameter in the logicist modeling in question, we judge it to be wise to refer to such tasks at a given *level*, not an order, so as to avoid confusion that will otherwise obtain.

Table 1: Table for Level-1 (L1) FBT = FBT_1

Label	English Declarative Content	Formula
L1.1	a_1 believes a_2 believes o is in b_1.	$\mathbf{B}_{a_1}\mathbf{B}_{a_2}I(b_1)$
L1.2	a_1 believes o is in b_1.	$\mathbf{B}_{a_1}I(b_1)$
L1.3	a believes a_2 believes o is in b_1.	$\mathbf{B}_{a}\mathbf{B}_{a_2}I(b_1)$
L1.4	a bel. a_1 bel. a_2 bel. o is in b_1.	$\mathbf{B}_{a}\mathbf{B}_{a_1}\mathbf{B}_{a_2}I(b_1)$

The level-two (or "second-order") FBT is easily captured, as follows.[2] First, when agent a_2 leaves, he/she secretly perceives a_1 move o to box b_2. Formally, the key adjustment is an *addition* to (adjustments of) the lines seen in Table 1: e.g.

L2 a_2 believes a_1 believes a_2 believes o is in b_1.

Prior Relevant Achievements Circa 2007, cognition associated with the *false-belief task* (FBT_1, including both FBT_1^P and FBT_1^F) was modeled in formal logic expressive enough to handle quantification, and computationally simulated (Arkoudas & Bringsjord, 2008, 2009).[3] This type of research falls under what Bringsjord (2008) calls *logicist computational cognitive modeling* (LCCM). As far as we are aware, this work in 2007 marks the first robust logicist modeling and simulation of both passing and failing cognition in FBT.[4] Here is the crucial takeaway from study of prior work: No one, before now, has achieved logicist computational cognitive modeling of quantified false-belief tasks at level 3, 4, ..., even in the non-quantificational case; and no one has reached the infinitary case.

Level-k ($k \geq 3$) False-Belief Tasks In the level-three false-belief task, agent a_1 secretly views a_2's secretly viewing into the room from outside it. (All of this is easily visualized with help from iterated, hidden cameras that feed information to the agents. Because of space limitations we forego visual depictions.) For FBT_3, the characteristic formula is:

$$\mathbf{B}_{a_1}\mathbf{B}_{a_2}\mathbf{B}_{a_1}\mathbf{B}_{a_2}I(b_1) \tag{1}$$

[2]A nice place to start reviewing the literature on FBT_2 is (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999), which has complete references to the earliest introduction of FBT_1 and FBT_2 in the (empirical) literature. (In this regard, we certainly recommend that interested readers review (Perner & Wimmer, 1985).) There is no discussion in this literature of level-3-and-above FBTs, let alone of infinitary FBTs such as FBT_ω; and we haven't found any formal/mathematical literature on these more demanding FBTs either.

[3]While formal but certainly declarative, very impressive computational cognitive modeling of FBT_1 was achieved earlier by Wahl and Spada (2000). Stenning and van Lambalgen (2008) provide informal declarative notation for modeling false belief, but have no implementation/simulation.

[4]Bello, Bignoli, and Cassimatis (2007), as in the aforecited (Wahl & Spada, 2000), achieve computational cognitive cognitive modeling of FBT_1 that makes use of declarative representations, but not of any logics.

Quantified False-Belief Tasks The sub-formulae $I(b_n)(n \in \{1,2\}$ is expressible within the formal language of only the propositional calculus. If instead of a single object o being used, a given FBT involves a *group* G of, say, n objects, then the correlate to this sub-formula will require the machinery of at least the quantificational machinery of first-order logic. We are able to model and computationally simulate in this more demanding case, so that even if subjects have beliefs about a quantity m from G ($m \leq n$) being placed in the box, their cognition can't be captured.

FBT_ω: An Infinitary Quantified False-Belief Task Our inference system leverages a computable version of an infinitary inference rule to prove FBT_ω given that we can prove FBT_n $\forall n$ (N. Govindarajulu, Licato, & Bringsjord, 2013).[5]

Automation We use an automated reasoning system, ShadowProver, to model FBT_n and FBT_ω. ShadowProver is a quantified modal logic theorem prover that has been used to model, in LCCM fashion, intricate reasoning tasks, e.g. ethical reasoning in (N. Govindarajulu & Bringsjord, 2017; N. S. Govindarajulu, Bringsjord, Ghosh, & Peveler, Forthcoming in 2019) and self-consciousness in (Bringsjord, Licato, Govindarajulu, Ghosh, & Sen, 2015). Since characteristic statements for FBT_n and FBT_ω are structurally similar to common knowledge, we leverage ShadowProver's ability to use the operator (**C**) for such knowledge.[6]

Objections We mention here only that while it might be objected that humans have trouble with even third-order belief, many of our college-level subjects on the contrary have little trouble proving correct answers for any FBT_n.

Acknowledgments The "late-breaking" achievements described herein have been enabled by generous support from ONR and AFOSR, for which we are deeply grateful.

References

Arkoudas, K., & Bringsjord, S. (2008). Toward Formalizing Common-Sense Psychology: An Analysis of the False-Belief Task. In T.-B. Ho & Z.-H. Zhou (Eds.), *Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2008)* (pp. 17–29). Springer-Verlag. Retrieved from http://kryten.mm.rpi.edu/KA_SB_PRICAI08_AI_off.pdf

Arkoudas, K., & Bringsjord, S. (2009). Propositional Attitudes and Causation. *International Journal of Software and Informatics*, *3*(1), 47–65. Retrieved from [7]

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by Normally Developing Children and Children with Asperger Syndrome or High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *29*, 407–418.

Bello, P., Bignoli, P., & Cassimatis, N. (2007). Attention and Association Explain the Emergence of Reasoning About False Belief in Young Children. In *Proceedings of the 8th international conference on cognitive modeling* (pp. 169–174). Ann Arbor, MI: University of Michigan.

Blackburn, T. B. P., & Polyanskaya, I. (forthcoming). Being Deceived: Information Asymmetry in Second-Order False Belief Tasks. *topiCS*.

Bringsjord, S. (2008). Declarative/Logic-Based Cognitive Modeling. In R. Sun (Ed.), *The Handbook of Computational Psychology* (pp. 127–169). Cambridge, UK: Cambridge University Press. Retrieved from [8]

Bringsjord, S., Licato, J., Govindarajulu, N., Ghosh, R., & Sen, A. (2015). Real Robots that Pass Tests of Self-Consciousness. In *Proccedings of the 24th ieee international symposium on robot and human interactive communication (ro-man 2015)* (p. 498-504). New York, NY: IEEE.

Govindarajulu, N., & Bringsjord, S. (2017). On Automating the Doctrine of Double Effect. In C. Sierra (Ed.), *Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17)* (pp. 4722–4730). International Joint Conferences on Artificial Intelligence. Retrieved from https://doi.org/10.24963/ijcai.2017/658

Govindarajulu, N., Licato, J., & Bringsjord, S. (2013). Small Steps Toward Hypercomputation via Infinitary Machine Proof Verification and Proof Generation. In M. Giancarlo, A. Dennuzio, L. Manzoni, & A. Porreca (Eds.), *Unconventional computation and natural computation; lecture notes in computer science; volume 7956* (pp. 102–112). Berlin, Germany: Springer-Verlag.

Govindarajulu, N. S., Bringsjord, S., Ghosh, R., & Peveler, M. (Forthcoming in 2019). *Beyond The Doctrine Of Double Effect: A Formal Model of True Self-Sacrifice.* (Robots and Well Being: Proceedings of the International Conference on Robot Ethics and Safety Standards 2017)

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that ...": Attribution of Second-Order Beliefs by 5–10-Year-Old Children. *Journal of Experimental Child Psychology*, *39*, 437–471.

Stenning, K., & van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.

Wahl, S., & Spada, H. (2000). Children's Reasoning About Intentions, Beliefs and Behaviour. *Cognitive Science Quarterly*, *1*(1), 3–32.

Wimmer, H., & Perner, J. (1983). Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. *Cognition*, *13*, 103–128.

[5] Some infinitary logics allow various uncountably infinite elements, and certainly from a purely logico-mathematical perspective there's no reason why false-belief tasks involving such elements can't be specified, but so far we have only explored, and met with success on, *countably* infinite false-belief tasks.

[6] $\mathbf{C}\phi \equiv$ 'All agents $\mathbf{K}\phi$, all agents $\mathbf{K}$ that all agents $\mathbf{K}\phi$, ...

[7] http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf

[8] http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf