Piagetian Roboethics via Category Theory: Moving Beyond Mere Formal Operations to Engineer Robots Whose Decisions are Guaranteed to be Ethically Correct*

Selmer Bringsjord, Joshua Taylor
Bram van Heuveln, Micah Clark
Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

Ralph Wojtowicz
Metron Inc.
1818 Library Street, Suite 600
Reston VA 20190 USA

Konstantine Arkoudas
Telcordia Technologies, Inc.
Piscataway NJ 08854 USA

September 15, 2010

1 Introduction

This paper introduces an *approach* to, rather than the final results of, sustained research and development in the area of roboethics described herein. Encapsulated, the approach is to engineer ethically correct robots by giving them the capacity to reason *over*, rather than merely *in*, logical systems (where logical systems are used to formalize such things as ethical codes of conduct for warfighting robots). This is to be accomplished by taking seriously Piaget's position that sophisticated human thinking exceeds even abstract processes carried out *in* a logical system, and by

*The R&D described in this paper has been partially supported by IARPA's A-SpaceX program (and other IARPA/DTO/ARDA programs before this one, e.g., NIMD and IKRIS), and, on the category-theoretic side, by AFOSR funding to Wojtowicz at Metron Inc., and through Metron to Bringsjord. An NSF CPATH grant to explore "social robotics," on which Bringsjord is a Co-PI (N. Webb PI), has been helpful as well. Bringsjord is indebted to Jim Fahey for insights regarding roboethics (including, specifically, whether ethical reasoning can be mechanized), to Robert Campbell for information about lesser-known aspects of Piaget's work, and to Ron Arkin for lively, stimulating discussion about various approaches to roboethics. Joshua Taylor has been funded in the past in part by the Tetherless World Constellation at RPI.

exploiting category theory to render in rigorous form, suitable for mechanization, structure-preserving mappings that Bringsjord, an avowed Piagetian, sees to be central in rigorous and rational human ethical decision-making.

We assume our readers to be at least somewhat familiar with elementary classical logic, but we review basic category theory and categorical treatment of deductive systems. Introductory coverage of the former subject can be found in [1, 2]; deeper coverage of the latter, offered from a suitably computational perspective, is provided in [3]. Additional references are of course provided in the course of this paper.

2 Preliminaries

A category consists of a collection of objects and a collection of arrows, or morphisms. Associated with each arrow f are a domain (or source), denoted $\mathrm{dom}\, f$, and a codomain (or target), denoted $\mathrm{cod}\, f$. An arrow f with domain A and codomain B is denoted $f : A \to B$ or

$$A \xrightarrow{f} B.$$

Associated with a category is an associative composition operator $\circ$ which is total on compatible arrows. That is, for any arrow $f : A \to B$, $g : B \to C$, and $h : C \to D$, the category has an arrow $g \circ f : A \to C$, and that $(h \circ g) \circ f = h \circ (g \circ f)$. For each object A in a category, there is an identity arrow $\mathrm{id}_A : A \to A$ such that for any $f : A \to B$, it holds that $\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A$.

Many mathematical structures can be represented as categories. For instance, the natural numbers form a category with a single object, $\star$, and arrows named by the natural numbers $\{ n : \star \to \star \mid n \in \mathbb{N} \}$. Composition is defined as addition on the natural numbers such that $m \circ n = m + n$, and is readily seen to be associative. The identity arrow, $\mathrm{id}_\star$, is 0, as for any n, $n + 0 = 0 + n = n$.

In addition, many classes of mathematical structures can be represented as categories wherein individual mathematical structures are the objects of the category and arrows are morphisms between the objects. For instance, the category **Set** has sets as its objects and set functions as its arrows. Composition in **Set** is function composition (which is associative). The identity arrows of **Set** are the identity functions on sets.

A notable example of this type of category is **Cat**, whose objects are categories, and whose arrows are category morphims, or functors. A functor $\mathscr{F} : \mathscr{C} \to \mathscr{D}$ maps the objects and arrows of category $\mathscr{C}$ to the object and arrows of category $\mathscr{D}$ such that $\mathscr{F}(\mathrm{id}_A) = \mathrm{id}_{\mathscr{F}(A)}$ and $\mathscr{F}(f \circ g) = \mathscr{F}(f) \circ \mathscr{F}(g)$. Note that this requirement ensures that for any arrow $f : A \to B$ of $\mathscr{C}$, the domain and codomain of $\mathscr{F}(f)$ are $\mathscr{F}(A)$ and $\mathscr{F}(B)$, that is $\mathscr{F}(f) : \mathscr{F}(A) \to \mathscr{F}(B)$.

A logic combines a language, typically a set of formulae defined by a context-free grammar, and rules for constructing proofs, that is, derivations of certain formulae from others. Most logics can be represented as categories by taking their formulae as objects, and positing that there is an arrow $p : \phi \to \psi$ if and only if p is a proof of ψ from ϕ. Most logics, and all the logics with which we shall be concerned herein, are such that given proofs $p : \phi \to \psi$ and $q : \psi \to \rho$, we can construct a proof

$q \circ p : \phi \rightarrow \rho$, and also such that for any formula ϕ, there is a proof $\mathrm{id}_\phi : \phi \rightarrow \phi$. It is worth noting that either the arrows in such a category must either be taken as equivalence classes of proofs or that $\circ$ is a sort of normalizing proof composition (i.e., to satisfy the requirements that $p \circ \mathrm{id}_\phi = \mathrm{id}_\phi = \mathrm{id}_\phi \circ q$ and $(p \circ q) \circ r = p \circ (q \circ r)$).

In treating logics as categories, we shall define the arrows of a category through the use of arrow schemata. For instance, in the propositional calculus, given proofs of ψ and ρ from ϕ, there is a proof of $\psi \wedge \rho$ from ϕ. We indicate this with the following schema.

$$\frac{\phi \xrightarrow{p} \psi \quad \phi \xrightarrow{q} \rho}{\phi \xrightarrow{\wedge \mathrm{I}\ p,q} \psi \wedge \rho}\ \wedge \text{ intro}$$

As another example, given a proof of the disjunction $\psi \vee \rho$, and proofs of the conditionals $\psi \supset \sigma$ and $\rho \supset \sigma$ from ϕ, there is a proof of ρ from ϕ.

$$\frac{\phi \xrightarrow{p_0} \psi_1 \vee \ldots \vee \psi_n \quad \phi \xrightarrow{p_1} \psi_1 \supset \rho \quad \ldots \quad \phi \xrightarrow{p_n} \psi_n \supset \rho}{\phi \xrightarrow{\vee \mathrm{E}\ p_0,p_1,\ldots,p_n} \rho}\ \vee \text{ elim}$$

Functors between categories that represent logics map the formulae and proofs of one logic to the formulae and proofs of another. Such mappings, or translations, have been used in the history of formal logic to demonstrate many relationships between logics. Herein we shall be concerned with the use of functors between such categories as tools to shift between representations of reasoning tasks.

3 Piaget's View of Thinking

Many people, including many outside psychology and cognitive science, know that Piaget seminally—and by Bringsjord's lights, correctly—articulated and defended the view that mature human reasoning and decision-making consists in processes operating for the most part on formulas in the language of classical extensional logic (e.g., see [4]).[1] You may yourself have this knowledge. You may also know that Piaget posited a sequence of cognitive stages through which humans, to varying degrees, pass. How many stages are there, according to Piaget? The received answer is: four; in the fourth and final stage, *formal operations*, neurobiologically normal humans can reason accurately and quickly over formulas expressed in the logical system known as first-order logic, $\mathscr{L}_I$.[2]

Judging by the cognition taken by Piaget to be stage-three or stage-four (e.g., see Figure 1, which shows one of the many problems presented to subjects in [4]), the basic scheme is that an agent $\mathscr{A}$ receives a problem P (expressed as a visual scene accompanied by explanatory natural language), represents P in a formal language

[1] Many readers will know that Piaget's position long ago came under direct attack, by such thinkers as Wason and Johnson-Laird [5, 6]. In fact, unfortunately, for the most part academics believe that this attack succeeded. Bringsjord doesn't agree in the least, but this isn't the place to visit the debate in question. Interested readers can consult [7, 8]. Piaget himself retracted any claims of *universal* use of formal logic: [9].

[2] Various other symbols are used, e.g., the more informative $\mathscr{L}_{\omega\omega}$.

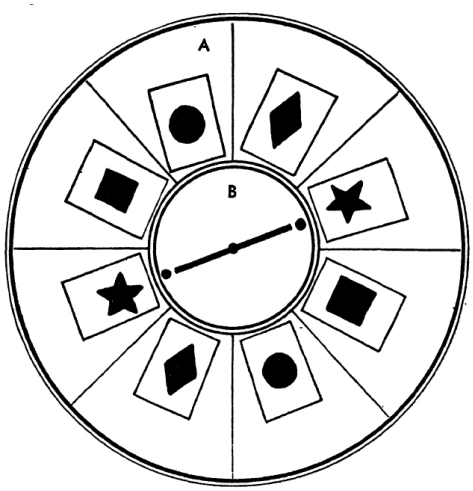

Figure 1: Piaget's famous "rigged" rotating board to test for the development of Stage-3-or-better reasoning in children. The board, A, is divided into sectors of different colors and equal surfaces; opposite sectors match in color. B is a rotating disk with a metal rod spanning its diameter—but the catch is that the star cards have magnets buried under them (inside wax), so the alignment after spinning is invariably as shown here, no matter how the shapes are repositioned in the sectors (with matching shapes directly across from each other). This phenomenon is what subjects struggle to explain. Details can be found in [4].

that is a superset of the language of $\mathscr{L}_I$, producing $[P]$, and then reasons over this representation (along with background knowledge Γ) using at least a combination of some of the proof theory of $\mathscr{L}_1$ and "psychological operators."[3] This reasoning allows the agent to obtain the solution $[S]$. To ease exposition, we shall ignore the heterodox operations that Piaget posits (see note 3) in favor of just standard proof theory, and we will moreover view $[P]$ as a triple (ϕ, C, Q), where ϕ is a (possibly complicated) formula in the language of $\mathscr{L}_I$, C is further information that provides context for the problem, and consists of a set of first-order formulas, and Q is a query asking for a proof of ϕ from $C \cup \Gamma$. So:

$$[P] = (\phi, C, Q = C \cup \Gamma \vdash \phi?)$$

For example, in the invisible magnetization problem shown in Figure 1, which requires stage-three reasoning in order to be solved, the idea is to explain how it is that $\phi^{\star\star}$, that is, that the rotation invariably stops with the two stars selected by the rod. Since Piaget is assuming the hypothetico-deductive method of explanation

[3] The psychological operators in question cannot always be found in standard proof theories. For example, Piaget held that the quartet I N R C of "transformations" were crucial to thought at the formal level. Each member of the quartet transforms formulas in certain ways. E.g., N is *inversion*, so that $\mathrm{N}(p \vee q) = \neg p \wedge \neg q$; this seems to correspond to DeMorgan's Law. But R is *reciprocity*, so $\mathrm{R}(p \vee q) = \neg p \vee \neg q$, and of course this isn't a valid inference in the proof theory for the propositional calculus or $\mathscr{L}_I$.

made famous by Popper [10], to provide an explanation is to rule out hypotheses until one arrives deductively at $\phi^{\star\star}$. In experiments involving child subjects, a number of incorrect (and sometimes silly) hypotheses are entertained—that the stars are heavier than the other shaped objects, that the colors of the sections make a difference, and so on. Piaget's analysis of those who discard mistaken hypotheses in favor of $\phi^{\star\star}$ is that they expect consequences of a given hypothesis to occur, note that these consequences fail to obtain, and then reason backwards by *modus tollens* to the falsity of the hypotheses. For example, it is key in the magnet experiments of Figure 1 that "for some spins of the disk, the rod will come to rest upon shapes other than the stars" is an expectation. When expectations fail, disjunctive syllogism allows $\phi^{\star\star}$ to be concluded. For our discussion of a sample functor over deductive systems as categories, it's important to note that while the hypotheses and context for the problem are naturally expressed using relation symbols, function symbols, and quantifiers from the language of $\mathscr{L}_I$, according to Piaget the final solution is produced by deduction in the propositional calculus.

4 From Piaget to Roboethics

What does all this have to do with roboethics? Well, for starters, notice that certain approaches to regulating the ethical decisions of lethal robots can be fairly viewed as aiming to engineer such robots by ensuring that they operate at Piaget's fourth stage. We believe this is true of both [11] and [12]. While in the first case an ethical code is to be expressed within some deontic/epistemic logic that subsumes classical logic,[4] and in the second there is no insistence upon using such more expressive logics, the bottom line is that in both cases there would seem to be a match with Piaget's fourth-stage: In both cases the basic idea is that robots work in a particular logical system, and their decisions are constrained by this work. In fact, it is probably not unfair to view an ethically relevant decision d by a robot to be correct if a formula in which d occurs can be proved from what is observed, and from background knowledge (which includes an ethical code or set of ethical rules, etc.)—so that a decision point becomes the solution of a problem with this now-familiar shape:

$$[P] = (\phi(d), C, Q = C \cup \Gamma \vdash \phi(d)?)$$

5 The Intolerable Danger of Fourth-Stage Robots

In a sentence, the danger is simply that if a lethal agent is unable to engage in at least something close to sophisticated human-level ethical reasoning and decision-making, and instead can only operate at Piaget's fourth stage (as that operation is formalized herein), it is evident that that agent will, sooner or later, go horribly awry. That is, it will perform actions that are morally wrong or fail to perform actions that

[4] A rapid but helpful overview of epistemic and deontic logic can be found in [13]. For more advanced work on computational epistemic logic see [14].

are morally obligatory, and the consequences will include extensive harm to human beings.

The reason such sad events will materialize is that a robot can flawlessly obey a "moral" code of conduct and still be catastrophically unethical. This is easy to prove: Imagine a code of conduct that recommends some action which, in the broader context, is positively immoral. For example, if human Jones carries a device which, if not eliminated, will (by his plan) see to the incineration of a metropolis (or perhaps a collection of metroplises), and a robot (e.g., an unmanned, autonomous UAV) is bound by a code of conduct not to destroy Jones because he happens to be a civilian, or be in a church, or at a cemetery... but has just one shot to save the day, and this is all the relevant information, it would presumably be immoral not to eliminate Jones. (This of course is just one of innumerable easily invented cases.)

Unfortunately, the approach referred to in the previous section is designed to bind robots by fixed codes of conduct (e.g., rules of engagement covering warfighters). This approach may well get us all killed—if in the real world a malicious agent like Jones arrives.

The approach that *won't* get us killed, and indeed perhaps the only viable path open to us if we want to survive, is to control robot behavior by operations over an ensemble of suitably stocked logical systems—operations from which suitable codes can be mechanically *derived* by robots on the fly. Once the code has been derived, it can be applied in a given set of circumstances.

6 But Then Why Piaget's Paradigm?

But if Piaget posits four stages, and deficient approaches to ethically correct robots already assume that such robots must operate at the fourth and final stage, what does the Piagetian paradigm have to offer those in search of ways to engineer ethically correct robots? The key fact is that Piaget actually posited stages *beyond* the fourth one—stages in which agents are able to operate over logical systems. For example, we know that logicians routinely create new logical systems (and often new components thereof that are of independent interest); this was something Piaget was aware of, and impressed by. But most people, even scholars with psychology of reasoning in academia, are not aware of the fact that Piaget's scheme made room for cognition beyond the fourth stage.

In fact, the truth of the matter is that Piaget made room for an arbitrary number of ever more sophisticated stages beyond the fourth. Piaget scholar and Clemson psychologist Robert Campbell writes:

> For [Piaget] there was no fixed limit to human development, and, wisely, he did not attempt to forecast future creative activity. Piaget did suggest that beyond formal operations, there are postformal operations, or "operations to the *n*th power." Inevitably these would be of a highly specialized nature, and might be found in the thinking of professional mathematicians or experts in some other field. (Lecture presented at the Institute of Objectivist Studies Summer Seminar, Charlottesville, VA, July 7 and 8, 1997. Available on the web at http://hubcap.clemson.edu/∼campber/piaget.html.)

What Piaget had in mind for post-formal stages would seem to coincide quite natural with our formal framework, in which post-formal reasoning involves the meta-processing of logics and formal theories expressed in those logics. Unfortunately, most of Piaget's later work has yet to be translated into English. For example, Campbell notes in the same lecture cited immediately above that a straightforward example of "operations to the *n*th power," according to Piaget, is the construction of axiomatic systems in geometry, which requires a level of thinking beyond Stage IV (= beyond formal operations). But the textual confirmation (e.g., Piaget: "one could say that axiomatic schemas are to formal schemes what the latter are to concrete operations") comes from work not yet translated from the French, viz., [15], p. 226.

7 Category Theory For Fifth-Stage Robots

Category theory is a remarkably useful formalism, as can be easily verified by turning to the list of spheres to which it has been productively applied—a list that ranges from attempts to supplant orthodox set theory-based foundations of mathematics with category theory [16, 17] to viewing functional programming languages as categories [3]. However, for the most part—and this is in itself remarkable—category theory has not energized AI or computational cognitive science, even when the kind of AI and computational cognitive science in question is logic-based.[5] We say this because there is a tradition of viewing logics or logical systems from a category-theoretic perspective. For example, Barwise [20] treats logics, from a model-theoretic viewpoint, as categories; and as some readers will recall, Lambek [21] treats proof calculi (or as he and others often refer to them, *deductive systems*) as categories. Piaget's approach certainly seems proof-theoretic/syntactic; accordingly, we provide now an example of stage-five category-theoretic reasoning from the standpoint of proof theory. (While Piaget, as we have noted, allows for the possibility of any number of stages beyond IV, we simplify the situation and refer to post-formal processing as 'Stage V.')

The example is based on two logical systems known to be directly used by Piaget, the propositional calculus $\mathscr{L}_{PC}$ and full first-order logic $\mathscr{L}_I$. We will work with the categories corresponding to these logics, **PC** and **FOL**, respectively. The review of basic category theory given in §2 should make the structure of **PC** and **FOL** relatively clear, but discussion of several points is in order.

Given that there are many formalizations of the propositional calculus (e.g., axiomatic methods with but one inference rule, natural-deduction-style systems, etc.), there are actually many categories that we might accept as **PC**. However, the consequence relations for propositional calculus and first-order logic are fixed, and we do require that **PC** represent a sound and complete proof calculus for the propositional calculus, that is, that there are arrows from ϕ to ψ if and only if $\phi \models_{PC} \psi$. For most proof systems, there will be infinitely many such proofs, and so infinitely many arrows for each consequence. We also maintain that in **PC** there is

[5] Bringsjord is as guilty as anyone, in light of the fact that even some very recent, comprehensive treatments of logicist AI and computational cognitive science are devoid of category-theoretic treatments. E.g., see [18, 19].

an object $\top$, to be read as "true" and that **PC** contains for every object ϕ an arrow $\top_\phi : \phi \to \top$. In addition, this construction provides the appropriate identity arrow, $\mathrm{id}_\top = \top_\top$. We impose the same restrictions on **FOL**; we require there be a "true" object $\top$, that the proof calculus respects the consequence relation, and so on. We also require that the arrows of **FOL** are generated by a superset of the schemata that generate the arrows of **PC**. Particularly, the schemata for **PC** define sentential proofs, while the extra arrows for **FOL** define proofs involving quantification and equality.

In this treatment we have followed the traditional scheme [21], but we must leave open paths for unsound and incomplete proof calculi because, clearly, in Piaget's work, proof calculi for humans would not necessarily include the full machinery of standard ones for the propositional and predicate calculi; and moreover, humans, according to Piaget, make use of idiosyncratic transformations that we would want to count as deductions (see note 3). While even the traditional scheme may seem to require some forcing of proof calculi into a categorical framework (e.g., by an equivalence relation imposed on proofs or by non-trivial proof composition), there are proof calculi which match this paradigm well. For instance, Arkoudas' NDL [22] explicitly calls out deductions, which can be composed.

We now provide a cognitively plausible functor-based mechanism for performing limited types of reasoning in **FOL** using **PC**. The standard truth functional form [1, Chapter 10] of a first order formula ϕ is a propositional formula that preserves the truth functional connectives present in ϕ, but maps all other formulae, viz., atomic formulae and quantifications, to propositional variables. That is, given an injection ι that maps atomic formulae and quantifications to propositional variables, the truth functional form of a formula ϕ, denoted $\tau(\phi)$ is defined as follows.

$$
\begin{aligned}
\tau(\top) &= \top \\
\tau(\neg\phi) &= \neg\tau(\phi) \\
\tau(\phi \wedge \psi) &= \tau(\phi) \wedge \tau(\psi) \\
\tau(\phi \vee \psi) &= \tau(\phi) \vee \tau(\psi) \\
\tau(\phi \supset \psi) &= \tau(\phi) \supset \tau(\psi) \\
\tau(\phi) &= \iota(\phi) \qquad\qquad \phi \text{ atomic or a quantification}
\end{aligned}
$$

We now define the category **PC**′ whose objects are the formulae in the image of ι along with $\top$ and compound formulae built up therefrom using the sentential connectives; this is exactly the image of τ. The arrows of **PC**′ are those defined by the same schemata used to define the arrows of **PC**. It is trivial to confirm that every object and arrow of **PC**′ is also an object or arrow of **PC**.

We can now construct a functor $\star$: **PC**′ $\to$ **FOL**. Since τ is an injection of first-order formulae, and a surjection to the objects of **PC**′, it is a bijection between the objects of **FOL** and **PC**′. The arrows of **PC**, and hence of **PC**′ are defined by a subset of the schemata used to define the arrows of **FOL**. The functor $\star$: **PC**′ $\to$ **FOL** simply maps each object ϕ of **PC**′ to its corresponding first-order formula $\tau^{-1}(\phi)$, and each arrow $p : \phi \to \psi$ to the arrow $p : \tau^{-1}(\phi) \to \tau^{-1}(\psi)$.

The function τ and functor $\star$ form a cognitively plausible mechanism for repre-

senting what we noted to be happening in connection with the magnet mechanism above, viz., subjects are representing phenomena associated with the apparatus using relations and quantifiers (as objects of **FOL**), but then encoding this information (via τ) in the propositional calculus (as objects of $\mathbf{PC}'$).

It seems to us plausible that in the case of the magnet challenge, humans who successfully meet it essentially do a proof by cases, in which they rule out as unacceptable certain hypotheses for why the rod always stops at the stars. Assuming this is basically correct, it seems undeniable that though humans perceive all the relations that are in play (colors, shapes, and so on), and in some sense reason over them, something like the function τ and functor $\star$ are applied to more detailed reasoning of **FOL** to distill down to the core reasoning, expressible in $\mathbf{PC}'$, and hence drop explicit reference to relations. The situation as we see it is summed up in Figure 2.

8 Demonstrations and Future Research

As we said at the outset of the present paper, our goal here has been to introduce an approach to roboethics. Nonetheless, we have made some concrete progress. For example, demonstration of an actual magnet puzzle-solving robot operating on the basis of the approach described above was engineered by Taylor and Evan Gilbert, and given by Bringsjord at the Roboethics Workshop at ICRA 2009, in Kobe, Japan. This demonstration used PERI, shown in Figure 4; and a snapshot derived from the video of the demonstration in question is shown in Figure 5.

Of course, we seek robots able to succeed on many of Piaget's challenges, not only on the magnet problem of Figure 1, and we we are developing Piagetian challenges of our own design that catalyze post-stage-four reasoning and decision-making. We are also working on microcosmic versions of the ethically charged situations that robots will see when deployed in warfare and counter-terrorism, where post-stage-four reasoning and decision-making is necessary for successfully handling these situations. These coming demonstrations are connected to NSF-sponsored efforts on our part to extend CMU's Tekkotsu [23, 24] framework so that it includes operators that are central to our logicist approach to robotics, and specifically to roboethics—for example, operators for belief (**B**), knowledge (**K**), and obligation ($\bigcirc$ of standard deontic logic). The idea is that these operators would link to their counterparts in bona fide calculi for automated and semi-automated machine reasoning. One such calculus has already been designed and implemented: the *cognitive event calculus*; see [25]. This calculus includes the full event calculus, a staple in AI; for example, see [26]. Given that our initial experiments will make use of simple hand-eye robots recently acquired by the RAIR Lab from the Tekkotsu group at CMU, Figure 3, which shows one of these robots, sums up the situation (in connection with the magnet challenge).

Finally, while our title contains 'Piagetian Roboethics,' the approach described in this short paper can of course generalize to robotics *simpliciter*. This generalization will be pursued in the future. In fact, the direction described herein is the kernel of an approach to logicist AI and computational cognitive science, whether or not the

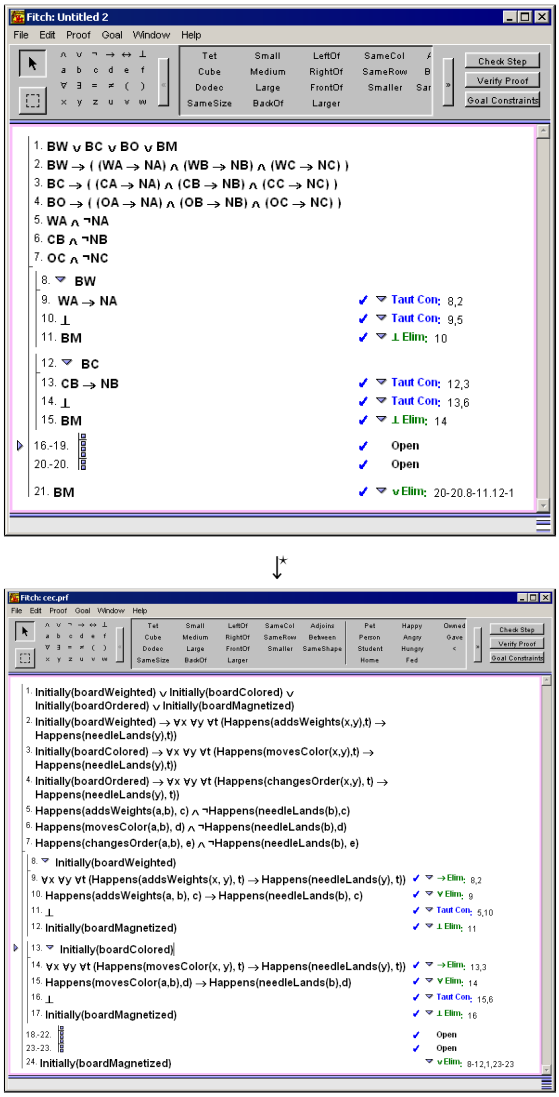

Figure 2: This figure shows two proofs, one expressed in **PC′**, the other in **FOL**. The first-order proof produces the conclusion that what causes the metal rod to invariably stop at the stars is that there are hidden magnets. The basic structure is proof by cases. Of the four disjuncts entertained as the possible source of the rod-star regularity, the right one is deduced when the others are eliminated. The functor $\star$ is shown here to indicate that the basic structure can be produced as a proof couched exclusively in the propositional calculus.

agents involved are physical or non-physical. Therefore, in the future, the general concept of agents whose intelligence derive from reasoning and decision-making

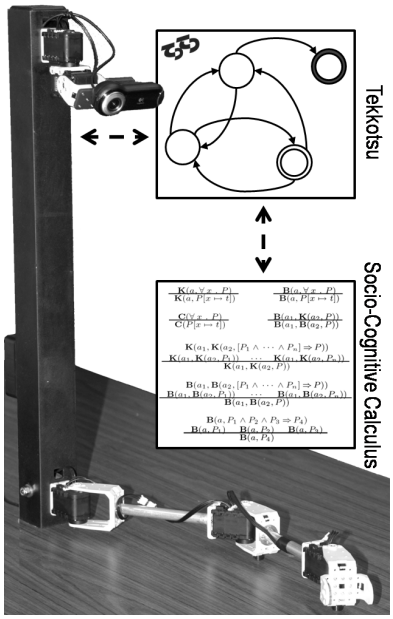

Figure 3: The basic configuration for our initial experiments involving Tekkotsu.

over logical systems (and their components) as categories will be pursued as well. It is not implausible to hold in Piagetian fashion that sophisticated human cognition, whether or not it is directed at ethics, exploits coordinated functors over many, many logical systems encoded as categories. These systems range from the propositional calculus, through description logics, to first-order logic, to temporal, epistemic, and deontic logics, and so on.

References

[1] J. Barwise and J. Etchemendy, *Language, Proof, and Logic*. New York, NY: Seven Bridges, 1999.

[2] H. D. Ebbinghaus, J. Flum, and W. Thomas, *Mathematical Logic (second edition)*. New York, NY: Springer-Verlag, 1994.

[3] M. Barr and C. Wells, *Category Theory for Computing Science*. Montréal, Canada: Les Publications CRM, 1999.

[4] B. Inhelder and J. Piaget, *The Growth of Logical Thinking from Childhood to Adolescence*. New York, NY: Basic Books, 1958.

[5] P. Wason, "Reasoning," in *New Horizons in Psychology*. Hammondsworth, UK: Penguin, 1966.

Figure 4: The RAIR Lab's PERI

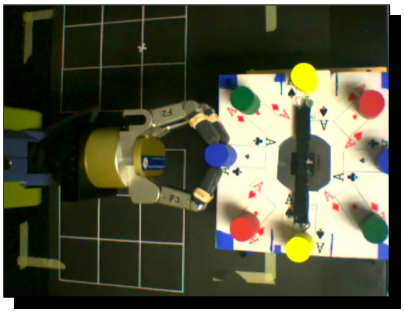

Figure 5: The RAIR Lab's PERI in the Process of Solving the Magnet Puzzle. Note that to make robot manipulation possible, playing cards and wooden cylinders have been used—but the problem here is isomorphic to Piaget's original version. Credit is given to Evan Gilbert and Trevor Housten for construction of the apparatus and programming of PERI.

[6] P. Wason and P. Johnson-Laird, *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press, 1972.

[7] S. Bringsjord, E. Bringsjord, and R. Noel, "In defense of logical minds," in *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum, 1998, pp. 173–178.

[8] K. Rinella, S. Bringsjord, and Y. Yang, "Efficacious logic instruction: People are not irremediably poor deductive reasoners," in *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, J. D. Moore and K. Stenning, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2001, pp. 851–856.

[9] J. Piaget, "Intellectual evolution from adolescence to adulthood," *Human Development*, vol. 15, pp. 1–12, 1972.

[10] K. Popper, *The Logic of Scientific Discovery*. London, UK: Hutchinson, 1959.

[11] S. Bringsjord, K. Arkoudas, and P. Bello, "Toward a general logicist methodology for engineering ethically correct robots," *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 38–44, 2006. [Online]. Available: http://kryten.mm.rpi.edu/bringsjord_inference_robot_ethics_preprint.pdf

[12] R. C. Arkin, "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture – Part iii: Representational and architectural considerations," in *Proceedings of Technology in Wartime Conference*, Palo Alto, CA, January 2008, this and many other papers on the topic are available at the url here given. [Online]. Available: http://www.cc.gatech.edu/ai/robot-lab/publications.html

[13] L. Goble, Ed., *The Blackwell Guide to Philosophical Logic*. Oxford, UK: Blackwell Publishing, 2001.

[14] K. Arkoudas and S. Bringsjord, "Metareasoning for multi-agent epistemic logics," in *Fifth International Conference on Computational Logic In Multi-Agent Systems (CLIMA 2004)*, ser. Lecture Notes in Artificial Intelligence (LNAI). New York: Springer-Verlag, 2005, vol. 3487, pp. 111–125. [Online]. Available: http://kryten.mm.rpi.edu/arkoudas.bringsjord.clima.crc.pdf

[15] J. Piaget, *Introduction a l' Épistémologie Génétique. La Pensée Mathématique*. Paris, France: Presses Universitaires de France, 1973.

[16] J.-P. Marquis, "Category theory and the foundations of mathematics," *Synthese*, vol. 103, pp. 421–447, 1995.

[17] F. W. Lawvere, "An elementary theory of the category of sets," *Proceedings of the National Academy of Science of the USA*, vol. 52, pp. 1506–1511, 2000.

[18] S. Bringsjord, "Declarative/logic-based cognitive modeling," in *The Handbook of Computational Psychology*, R. Sun, Ed. Cambridge, UK: Cambridge University Press, 2008, pp. 127–169. [Online]. Available: http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf

[19] ——, "The logicist manifesto: At long last let logic-based AI become a field unto itself," *Journal of Applied Logic*, vol. 6, no. 4, pp. 502–525, 2008. [Online]. Available: http://kryten.mm.rpi.edu/SB_LAI_Manifesto_091808.pdf

[20] J. Barwise, "Axioms for abstract model theory," *Annals of Mathematical Logic*, vol. 7, pp. 221–265, 1974.

[21] J. Lambek, "Deductive systems and categories i. Syntactic calculus and residuated categories," *Mathematical Systems Theory*, vol. 2, pp. 287–318, 1968.

[22] K. Arkoudas, "Simplifying Proofs in Fitch-Style Natural Deduction Systems," *Journal of Automated Reasoning*, vol. 34, no. 3, pp. 239–294, Apr. 2005.

[23] D. Touretzky, N. Halelamien, E. Tira-Thompson, J. Wales, and K. Usui, "Dual-coding representations for robot vision in Tekkotsu," *Autonomous Robots*, vol. 22, no. 4, pp. 425–435, 2007.

[24] D. S. Touretzky and E. J. Tira-Thompson, "Tekkotsu: A framework for AIBO cognitive robotics," in *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*. Menlo Park, CA: AAAI Press, 2005.

[25] K. Arkoudas and S. Bringsjord, "Propositional attitudes and causation," *International Journal of Software and Informatics*, vol. 3, no. 1, pp. 47–65, 2009. [Online]. Available: http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf

[26] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall, 2002.