A Logic-based Computational Framework for Inferring Cognitive Affordances

Vasanth Sarathy, *Member, IEEE,* Matthias Scheutz, *Member, IEEE,*

***Abstract*—The concept of "affordance" refers to the relationship between human perceivers and aspects of their environment. Being able to infer affordances is central to commonsense reasoning, tool use and creative problem solving in artificial agents. Existing approaches to inferring affordances have focused on functional aspects, relying on either static ontologies or statistical formalisms to extract relationships between physical features of objects, actions, and the corresponding effects of their interaction. These approaches do not provide flexibility with which to reason about affordances in the open world, where affordances are influenced by changing context, social norms, historical precedence, and uncertainty. We develop a computational framework comprising a probabilistic rules-based logical representation coupled with a computational architecture (CALyX) to reason about affordances in a more general manner than described in the existing literature. Our computational architecture allows robotic agents to make deductive and abductive inferences about functional and social affordances, collectively and dynamically, thereby allowing the agent to adapt to changing conditions. We demonstrate our approach with experiments, and show that an agent can successfully reason through situations that involve a tight interplay between various social and functional norms.**

I. Introduction

NATURAL human activities involve using and manipulating objects around us and continuously reasoning about our environment. Consider the example of cooking activities in a restaurant kitchen: these activities require cutting vegetables, monitoring the stove and keeping tools and utensils clean, all while ensuring orders are prepared and served in a coordinated and timely manner. Not only are team members able to recognize various objects around the kitchen, but they know what to do with these objects, how to use them appropriately, how to help others use them (i.e., they can infer and act on complex object affordances). That is, the kitchen team is using these affordances to reason about the task at hand. Sometimes these types of activities involve standard reasoning tasks, like choosing a clean knife for cutting a tomato. Other times, these activities involve more creative reasoning tasks like solving puzzles and finding novel uses for objects, like using a dishcloth as an oven mitt.

Reasoning and using objects in this manner is a highly desirable skill for robotic agents as well. Helper robots will be critical in many application domains: helping our elderly and disabled in assisted living facilities, conducting search-and-rescue missions in unforgiving terrain to save human lives, assisting our astronauts on the space station, or even monitoring our surroundings to keep us safe from national security threats. In these critical sectors it is highly beneficial to endow robots with the ability to find creative ways to use and manipulate objects, especially when there is minimal and uncertain information. Unfortunately, although today's robots are proficient at recognizing object features, they are less skilled at recognizing what can be done with these objects.

In this work, we present a novel computational framework based on Dempster-Shafer (DS) theory [1] and "uncertain logic" for inferring object affordances. Our framework comprises a logic-based representational format and inference mechanism coupled with a nascent computational architecture, CALyX (Cognitive Affordances Logically eXpressed), to reason about not only functional and physical features of objects, but also social, historical, aesthetic and ethical aspects that we naturally consider when perceiving objects – generally, "cognitive affordances". For example, we know that dirty knives are typically not used for cutting vegetables, even though they can functionally accomplish the task. As such we will demonstrate, with examples, that with our proposed approach a robot will be able to reason about these kinds of complicated affordances in a unified, systematic, and effective manner.

II. Past Work

A. General Background

Gibson introduced the notion of "affordance" to account for human visual perception [2]. He considered affordances as latent properties of the environment that exists in the presence of animals and humans. His general description captured the deeply interconnected relationship between an animal and its environment in ecological terms. Since then, the idea of affordances has been adopted across a variety of disciplines including psychology, computer science, artificial intelligence, and human-computer design. However, despite this extensive adoption, the ontological and representational aspects of affordances have been the subject of rigorous debate.

Two contested questions focused on were what affordances are and where they are supposed to live? Do they belong to the environment, to the agent, or to the agent-environment system? Turvey proposed that affordances are dispositional properties of the environment and actualized by the actions of the agent [3]. Reed proposed a more radical theory stating that affordances, although disposed in the environment, are a scarce resource and actually play a role in regulating human

V. Sarathy and M. Scheutz are with the Department of Computer Science, Tufts University, Medford, MA, 02155 USA, e-mail: (see http://hrilab.tufts.edu/).

Manuscript received January 17, 2016.

adaptive behavior and natural selection [4]. Norman proposed two different kinds of affordances: a real affordance that is in the environment and a perceived affordance that is in the agent's mind [5].

Stoffregen argued that affordances do not belong to either the agent or the environment, but are instead emergent properties of an agent-environment system [6]. Several theories also explored explicit representational formats including describing affordances as relations connecting the abilities of the agent with environmental features [7], and further connecting the effects of agent behaviors on the course of events [8], or as relations connecting environmental attributes in overlapping conceptual spaces or regions [9], [10]. Scarantino claimed that affordances are not only relational, but also conditional [11]. Specifically, she argued that affordances are conditional upon various triggering conditions and related to the agent's set of potential abilities.

A number of these and other theories focused primarily on functional aspects of affordances [12], [13]. There has also been some limited work in introducing social considerations into an affordance framework. Schmidt argued to extend Scarantino's theory of conditional and relative nature of affordance to include the idea of social affordances [14]. Work by Kim in the particular space of cognitive robotics and object handover considered social etiquette and norms as well [15], [16].

Using and manipulating objects involves not only functional and physical aspects of objects, but other features including social conventions that govern the object's use, aesthetic considerations that limit what can and cannot be done with an object, ethical factors that guide moral action, and historical precedence that influences the designed purpose and intent for the objects. An affordance inference framework must allow for a broad definition of affordance, which we refer to as "cognitive affordance", one that accounts for functional as well as non-functional aspects and must be adaptable to allow for continuous changes to and evolution of these aspects over time. Some of the above-mentioned theories and representations are limited in their ability to reason about affordances more holistically and contextually.

In the next section, we will describe past approaches to reasoning with affordances as used in cognitive robotics. These approaches adopt some of the more conceptual theories mentioned above, e.g., those by [8], [10], [15], [16] and implement them in computational and robotic systems. Discussing these approaches allows us to more specifically place our own contribution in the context of past work.

B. Affordances in Cognitive Robotics

In cognitive robotics, there have primarily been two types of approaches to representing, inferring, and reasoning with affordances: (1) approaches based on statistical and machine learning formalisms, and (2) approaches based on ontological formalisms. These are very powerful approaches and have shown substantial benefits to robotic cognition. However, as we will discuss, these approaches are limited both representationally and architecturally. Specifically, they do not demonstrate flexible representational formats to account for social and other non-functional aspects of affordances, they do not allow for contextual reasoning, and they do not address uncertainty in perception and beliefs. These approaches are also limited architecturally because they mostly only involve bottom-up processing of sensory (mostly visual) information and thus do not allow for much top-down processing of sensory information, which is necessary for a more complete account of affordance perception.

Steedman used Linear Dynamic Event Calculus to formalize the relationship between objects and their affordances [17]. More recently, work by Abel and Tellex focused on using Markov Decision Processes to directly model affordances as mappings between a set of preconditions and goal states, to action possibilities [18]. Mastrogiovanni et al. have developed a framework, using Self-Organizing Neural Maps, for action selection and functional representation of everyday objects, places and actions in terms of affordances and capabilities, as regions in a proper metric space [9], [10].

The strength of these works lies in their joint modeling of affordances with the problem of planning and action sequencing. It allows for not just reasoning about actions, but also implementing action sequences that then allow for new affordances to emerge. However, while affordance perception involves action selection from a choice of action capabilities, its inference has broader applicability than just for planning. Affordance inference is important to other cognitive processes involved in commonsense reasoning, natural language explanations, and general environmental sense-making. We believe there is a benefit to representing and reasoning with affordances in a manner that disentangles it from planning, but still allowing for leveraging the extensive advances in the planning literature.

Montesano et al. have developed statistically-inspired causal models of affordance using Bayesian networks to formalize the relationship between object features, actions, and effects [19], [20]. Several others have modeled affordances as a relationship between action, object, and effect [21], [22], [23], [24]. A number of computational and robotic systems have also emerged to tackle various sub-problems relating to robotic affordances such as object grasping and handover [25], [26], [23].

The strengths of these works lies in their underlying model of affordances per Sahin's approach of relating objects, actions, and the effects [8], allowing for a close relationship with planning. But here too, inference of affordances is not separate from specific planning tasks and, therefore, is not applied more generally.

A few researchers have explored ontology-based approaches to represent functional affordances. For example, Varadarajan et al. have developed a detailed knowledge-ontology based on conceptual, functional and part properties of objects, and then used a combination of detection and query matching algorithms to pinpoint the affordances for objects [27], [28]. While being able to query an affordance knowledge-base is helpful from a deductive standpoint, this approach is limited in its flexibility for accounting for contextual shifts, and changing social norms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2016.2615326, IEEE Transactions on Cognitive and Developmental Systems

Moreover, the focus on much of the affordance work in cognitive robotics is on functional affordances, and so there is often no distinction provided between a hammer in a person's toolbox and a decorative hammer on display at the museum, both of which are functionally equivalent, but engender entirely different non-functional affordances. The social affordances associated with interacting with a museum object are vastly different from the social affordances of interacting with a hammer in a personal toolbox.

Shu et al. have recently presented a framework for reasoning about social affordances and provide a system that can act in social scenarios like handshaking, helping a person stand up, high-fiving, and handing over objects [29]. While Shu is reasoning about affordances in social interactions, the underlying affordance model is still largely devoid of contextual reasoning, and focused more on physical geometries of objects in these scenarios (in this case skeletal geometries). However, such physical aspects do not account for the contextual information that is not perceptual (e.g., high-fiving a friend versus a refraining from high-fiving an enemy) and is also subject to change.

Thus more generally, despite these past efforts, affordance representation faces many challenges that have not been overcome in the previous work. Specifically, past approaches fail to provide flexibility with which to reason about affordances in the open world, where they are influenced by changing context, social norms, historical precedence, and uncertainty. For example, none of the current approaches can systematically infer that coffee mugs afford grasping and drinking, while also simultaneously affording serving as a paperweight or cupholder, or depending on the context, as family heirloom not meant to be used at all. We argue that inferences of this sort are different from sole high-level reasoning or planning processes, for they require a continuous interplay between low-level sensory systems and high-level cognitive systems and between bottom-up (sensory mechanisms to higher-level cognition) and top-down processing (higher-level cognition to sensory mechanisms) of information in these systems. Critically, cognitive affordance representation and reasoning is a separate cognitive process in its own right and deserves its own architectural framework and inference machinery (separate from high-level reasoning and planning or low-level feature detection) that can then later be tied together with suitable perceptual and planning and reasoning frameworks. This is not to say that affordances are not influenced by perception, planning, and reasoning – they are – but affordance-based reasoning is fully explained by and thus not subsumed within these processes.

Next, we present an architecture for reasoning about affordances that has components distinct from perceptual processes (e.g., vision, haptics) and from action processes (e.g., planning and natural language interaction). Our framework enables reasoning about higher-level affordances that rely on cognition and contextual reasoning separately from perception and action.

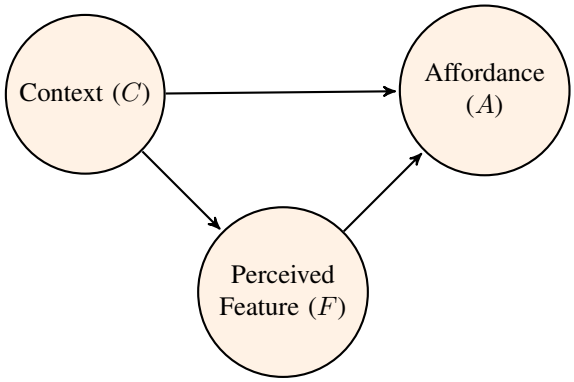

Fig. 1: Context-Sensitive Cognitive Affordance Model

III. The Computational Cognitive Affordance Framework

The proposed computational cognitive affordance framework consists of (1) a logic-based affordance representation and (2) a computational architecture (CALyX) that is context-sensitive and furthermore allows for top-down constraints on visual perception of the environment. Note that the proposed CALyX architecture is distinct from a (low-level) vision systems even though affordance reasoning can interface with it. Rather, the affordance representations used in CALyX are agnostic to the originating modality of the percepts (e.g., vision, haptics, natural language, etc.), allowing for reasoning at a higher than sensory-level (the sensory-level is sometimes referred to as detection in ecological psychology). Different from mere sensory processing of affordances, the higher-level representations and reasoning processes take into account perceptual, task-based, and other context as well as relevant mental states of the agent such as beliefs, intentions, goals, and desires.

A. Logic-Based Representation

We propose a novel representational format for cognitive affordances, illustrated in Fig. 1, in which an object's affordance (A) and its perceived features (F) depend on the context (C) [30]. We use Dempster-Shafer (DS) theory [1] – an uncertainty processing framework often interpreted as a generalization of the Bayesian framework – for inferring affordance (A) from object features (F) in contexts (C). More specifically, the proposed cognitive affordance model consists of four parts: (1) a set of perceivable object features (F), (2) a set of context states (C), (3) a set of object affordances (A), and (4) a set of "affordance rules" (R) connecting object features and context states to applicable affordances which take the overall form:

$$r :\equiv f \wedge c \implies_{[\alpha,\beta]} a$$

with $f \in F$, $c \in C$, $a \in A$, $r \in R$, $[\alpha, \beta] \subseteq [0, 1]$. Here, the confidence interval $[\alpha, \beta]$ is intended to capture the uncertainty associated with the affordance rule r such that if $\alpha = \beta = 1$ the rule is logically true, while $\alpha = 0$ and $\beta = 1$ assign maximum uncertainty to the rule. Rules can then be applied for a given feature percept f in given context c to obtain the implied affordance a under uncertainty about f, c, and the extent to which they imply the presence of a.

We have previously shown that these types of rules are very versatile and that we can employ "DS-theoretic modus ponens" to make uncertain deductive and abductive inferences [31]. Most critically, these rules allow us to address representational challenges with Bayesian models where $P(A|F,C)$ needs to be inferred by way of $P(F|A,C)$, $P(A|C)$, and $P(C)$ when we often have no practical way of obtaining the necessary probability distributions for all the affordances for an object. We will next provide an overview of our proposed computational architecture, which we will then use in combination with the above mathematical model to reason through two situations, each involving a tight interplay between social and functional affordances.

B. Computational Architecture (CALyX) - Overview

1) Introduction: We now present the computational *Cognitive Affordances Logically eXpressed* (CALyX) architecture (Fig. 2) for perceiving and reasoning about cognitive affordances in a unified manner. CALyX has two main components: (1) an *Affordance Reasoning Component* (ARC) for performing logic-based inferences of cognitive affordances, and (2) a *Perceptual Semantics and Attention Control Component* (PAC) for directing perception in a top-down manner and semantically analyzing perceptual information in a bottom-up manner. In addition, CALyX has two supporting memories: *Long-term Memory* (LTM) and *Working Memory* (WM), for storing and updating logical affordance rules and related uncertainties. These components work closely with sensory and perceptual systems (e.g., vision) and other components in a cognitive architecture to coordinate perceptual and action processing.

We will focus on the main components noted above and briefly touch upon other cognitive components as and when needed. It is important to note here that CALyX is only a part of a larger cognitive architecture and as such we do not expect it to cover other cognitive subsystems (e.g., those for planning or natural language processing) or provide an account for all manner of cognitive function. Instead, we focus on affordance perception and inference and note that CALyX serves as a intermediary subsystem linking lower level perceptual subsystems (e.g., vision, motor control, haptics) with higher-level belief, planning and goal management systems to facilitate top-down and bottom-up processing in the larger cognitive architecture (e.g., the DIARC architecture within which CALyX was developed [32]).

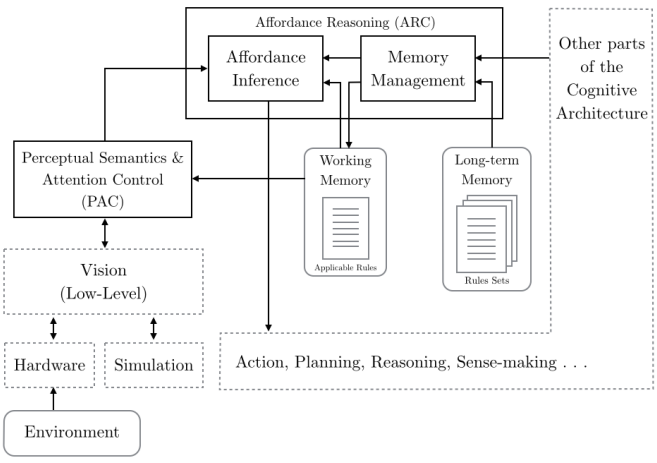

Fig. 2: Computational Architecture (CALyX). We depict our contribution in bold solid lines.

2) Cognitive Cycle: In each cognitive cycle, ARC selects applicable rules from LTM and populates WM. Once the rules are in Working Memory, both PAC as well as ARC use these rules as the basis for perception and inference. More specifically, PAC directs low-level perceptual systems like vision to perform visual searches in a focused manner only looking to determine beliefs for the specific perceptual features, F, relevant to the applicable rules in Working Memory. This is a top-down attentional strategy that helps the robot focus its senses on relevant parts of the environment given the rules in the WM while ignoring others. ARC performs DS-theoretic affordance inference on the rules in WM using beliefs about the relevant perceptual features from PAC and beliefs about contexts provided by other parts of the cognitive architecture. The outcome of the inference process is the generation of truth values of various affordances specified in WM and their associated uncertainty intervals, which are then used by the rest of the cognitive architecture for planning, reasoning and sense-making tasks.

3) Memory Management and Context-Sensitivity: In any given situation the robot might be subject to a set of overlapping contexts. For example, in a situation in which a robot is a kitchen helper, it might be subject to a context that refers to its role as a helper-robot. Simultaneously, the robot may also be in a more specific context that refers to a particular task that it must perform, for example, the task of handing over a knife to the human chef. This set of contextual aspects, C, collectively constitutes the agent's situation. As noted earlier, contexts may often not be perceivable, containing non-perceivable aspects of the task context and environment as well as the agent's own belief system. The fact that the agent is a kitchen helper, for example, is not necessarily perceivable by simply visually scanning the environment. Information about the robot's role, beliefs, desires and intentions may be provided by other high-level processing components in the robot's cognitive architecture. Thus, contextual aspects represent those descriptors of the situation which can be non-physical and even abstract.

This contextual information is passed into CALyX from other parts of the cognitive architecture and received by a Memory Management subcomponent of ARC. The Memory Management subcomponent searches through all available affordance rules of the form specified above in the agent's LTM and identifies rules that contain contexts that match those in the current situation. We use a matching threshold ζ to determine whether the current context "matches" the context presented in the rule. We can set the mass threshold to a value $0 \leq \zeta \leq 1$ and check if the mass of the current context exceeds this ζ threshold. Contextual aspects with masses satisfying the threshold condition will be considered by the Memory Management subcomponent. The Memory Management subcomponent aggregates the applicable rules

(i.e., rules applicable to contexts in the current situation) and populates WM. The WM stores rules of the form described earlier along with corresponding uncertainty intervals.

The Memory Management subcomponent passes the contextual aspects along with their mass assignments to the Affordance Inference subcomponent. The Affordance Inference subcomponent also has access to the WM of rules and accompanying uncertainties. In order to perform inference, the Affordance Inference subcomponent also needs uncertainty information about the set of perceptual aspects, F, identified in the applicable rules stored in WM. For this, it will turn to PAC.

4) Attention Control and Perceptual Semantics: PAC accesses the rules in WM and determines what perceptual aspects need to be evaluated. For example, if PAC needs to compute if there are grasp locations near the handle of a knife, it can resolve this perceptual relation query:

$$near(knife, G, holdPart(knife) = handle)?$$

PAC includes vision algorithms to resolve various sorts of relations including the spatial relation of $near()$. PAC directs low-level perception subcomponents (e.g., vision) to look for and identify uncertainties associated with relevant perceptual aspects. PAC returns to ARC the masses associated with perceptual aspects of the applicable rules.

5) Affordance Inference: The Affordance Inference subcomponent of ARC then performs DS-theoretic inference on the rules in WM using masses for the contextual aspects obtained from the Memory Management subcomponent and masses for the perceptual aspects obtained from PAC. ARC computes the uncertainties associated with affordances prescribed by the rules. In certain cases, the Memory Management subcomponent will selectively populate WM with rules that not only satisfy context, but also specify relevant affordance relations. This set of rules would be a subset of the applicable rules for the selected context.

Generally, affordance aspects and their associated uncertainty intervals and confidence measures are passed from ARC to other parts of the cognitive architecture including those subsystems responsible for planning, reasoning and sensemaking.

IV. Robot Kitchen Helper Experiment: Using and Handing Over Objects

For the experimental evaluation of the proposed computational cognitive affordance framework we will consider using and handing over objects in a kitchen as a running example to discuss the representation format, the uncertainty processing framework, and the inference algorithm in the implemented CALyX architecture. We will show how our framework assists the agent in reasoning about and deciding what action possibilities are available during each phase of the handover process, from grasping the object, to handing it over.

Note that handing over objects "properly" is an important skill for helper robotic agents. When performing a handover, the robot will need to reason about potential actions it can perform on objects (affordances), for example, selecting grasps or manipulating the object in certain ways. Existing approaches have focused on selecting one or two handover norms (e.g., orienting a handle towards the receiver), *a priori*, and then building object recognition and motion planning systems that are dependent on the preselected norms [25], [26]. These approaches fail to provide flexibility with which to reason about action choices in an open world, where norms and rules may change, norms may be added and removed, normative conflicts may arise, and other contextual factors may influence the propriety of a handover. In contrast, we intend to infer affordances based on (1) the semantic representation of certain visual percepts, (2) the agent's current context, and (3) the general domain and commonsense knowledge of the agent.

We will first provide a brief review of Dempster-Shafer theory and then use our framework to model the domain of cooking and assisting humans in the kitchen. Then we will walk through how an agent, staffed as a kitchen helper, reasons through the process of handing over knives.

A. Dempster-Shafer Theory Preliminaries

A set of elementary events of interest is called *Frame of Discernment* (FoD). The FoD is a finite set of mutually exclusive events $\Theta = \theta_1, ..., \theta_N$. The power set of Θ is denoted by $2^\Theta = \{A : A \subseteq \Theta\}$ [1].

Each set $A \subseteq \Theta$ has a certain weight, or *mass* associated with it. A *Basic Belief Assignment* (BBA) is a mapping $m_\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$ such that $\sum_{A \subseteq \Theta} m_\Theta(A) = 1$ and $m_\Theta(\emptyset) = 0$. The BBA measures the support assigned to the propositions $A \subseteq \Theta$ only. The subsets of A with non-zero mass are referred to as *focal elements* and comprise the set $\mathcal{F}_\Theta$. The triple $\mathcal{E} = \{\Theta, \mathcal{F}_\Theta, m_\Theta(\cdot)\}$ is called the *Body of Evidence* (BoE). For ease of reading, we sometimes omit $\mathcal{F}_\Theta$ when referencing the BoE.

Given a BoE $\{\Theta, \mathcal{F}_\Theta, m_\Theta(\cdot)\}$, the *belief* for a set of hypotheses A is $Bel(A) = \sum_{B \subseteq A} m_\Theta(B)$. This belief function captures the total support that can be committed to A without also committing it to the complement A^c of A. The *plausibility* of A is $Pl(A) = 1 - Bel(A^c)$. Thus, $Pl(A)$ corresponds to the total belief that does not contradict A. The $uncertainty$ interval of A is $[Bel(A), Pl(A)]$, which contains the true probability $P(A)$. In the limit case with no uncertainty, we get $Pl(A) = Bel(A) = P(A)$.

Dempster-Shafer theory can be considered a generalization of Bayesian theory. For example, a Bayesian would model Schrödinger's cat as a probability distribution over $\{dead, alive\}$, assigning a probability to each hypothesis. Dempster-Shafer would assign masses to each of $\{dead, alive, \{dead\ or\ alive\}\}$, without beliefs having to sum up, for example $Bel(dead)+Bel(alive) \neq Bel(dead \vee alive)$. One notable advantage of this uncertainty processing framework is that it allows for the allocation of probability masses to sets of hypotheses, and does not require an assumption about the probability distribution among members of that set.

Logical inference with uncertainty can be performed using DS-theoretic Logical inference with uncertainty can be performed using DS-theoretic "Modus Ponens" (denoted $\odot$) as discussed by Tang et al. [33]. We will use Tang's DS-theoretic

AND (denoted $\otimes$) to combine BoEs on different FoDs [33]. We choose to use Tang's models of Modus Ponens and AND over other proposed models because those models do not allow uncertainty to be multiplicatively combined.

We will use Yager's rule of combination (denoted $\bigcap$) to combine BoEs on the same FoD [34]. Yager's rule of combination aggregates evidences or confidence values from different sources, but within the same frame of discernment. Formally, when combining evidence from n different sources within the same frame, Θ, the combined multi-evidence BBA, according to Yager's rule is defined as follows:

$$m_\Theta(\emptyset) = 0$$

$$m_\Theta(A) = \sum_{\cap B_i = A} \prod_{i=1}^{n} m_{\Theta i}(B_i), \forall A \subseteq \Theta, A \neq \Theta, A \neq \emptyset$$

$$m_\Theta(\Theta) = \prod_{i=1}^{n} m_{\Theta i}(B_i) + \sum_{\cap B_i = \emptyset} \prod_{i=1}^{n} m_{\Theta i}(B_i)$$

Yager's rule of combination is chosen because it allows uncertainty to be pooled in the universal set, and due to the counter-intuitive results produced by Dempster's rule of combination, as discussed in [35].

For two logical formulae ϕ_1 (with $Bel(\phi_1) = \alpha_1$ and $Pl(\phi_1) = \beta_1$) and ϕ_2 (with $Bel(\phi_2) = \alpha_2$ and $Pl(\phi_2) = \beta_2$, applying logical AND yields $\phi_1 \otimes \phi_2 = \phi_3$ with $Bel(\phi_3) = \alpha_1 \cdot \alpha_2$ and $Pl(\phi_3) = \beta_1 \cdot \beta_2$.

For logical formulae ϕ_1 (with Bel$(\phi_1) = \alpha_1$ and $Pl(\phi_1) = \beta_1$) and $\phi_{\phi_1 \to \phi_2}$ (with $Bel(\phi_{\phi_1 \to \phi_2}) = \alpha_R$ and $Pl(\phi_{\phi_1 \to \phi_2}) = \beta_R$, the corresponding model of Modus Ponens is $\phi_1 \odot \phi_{\phi_1 \to \phi_2} = \phi_2$ with $Bel(\phi_2) = \alpha_1 \cdot \alpha_R$ and $Pl(\phi_2) = 1 - ((1 - Pl(\beta_1)) \cdot (1 - Pl(\beta_R)))$.

Moreover, we will use the "confidence measure" λ (defined in [36]) to be able to compare uncertainties associated with formulas ϕ and their respective uncertainty intervals $[\alpha, \beta]$:

$$\lambda(\alpha, \beta) = 1 + \frac{\beta}{\gamma} log_2 \frac{\beta}{\gamma} + \frac{1-\alpha}{\gamma} log_2 \frac{1-\alpha}{\gamma}$$

$$\text{where } \gamma = 1 + \beta - \alpha.$$

Here, ϕ is deemed more ambiguous as $\lambda(\alpha, \beta) \to 0$.

B. Semantic Representation of Visual Perception, F

The vision pipeline for an artificial agent involves various low-level components that are coupled together to process color and depth information and generate point clouds and 3D meshes. As noted earlier, PAC is configured to perform scene representation and semantic analysis to generate predicates that capture, qualitatively, certain aspects of the visual scene.

Let $F = \{\Theta_{F_1}, \Theta_{F_2}, \ldots, \Theta_{F_N}\}$ be the set of N different perceptual aspects such as color, shape, texture, relational information, and generally information obtained from the vision pipeline that an agent may interpret. Each aspect $\Theta_{F_i} = \{f_{i,1}, f_{i,2}, \ldots, f_{i,M}\}$ has a set of M mutually-exclusive candidate perceptual values (percepts), which come from the vision system as a BoE, $\mathcal{E}_{F_i} = \{\Theta_{F_i}, m_{\Theta_{F_i}}(\cdot)\}$. We will use $m_{f_{i,j}}$ to denote the candidate mass values of the percepts, where $i \in \{1 \ldots N\}$ and $j \in \{1 \ldots M\}$.

For the purposes of our example, we will represent the agent's visual perception of kitchen objects with nine binary visual aspects, each aspect with a percept and its negation. Thus, $\Theta_{F_i} = \{f_{i,j}, \neg f_{i,j}\}$, where $i \in \{1 \ldots 9\}$ and $j \in \{1\}$. The percepts and masses for each of the nine aspects, can be obtained from the low-level vision system, are shown below:

$holdPart(O)$ and $funcPart(O)$ are functions that return the name of the holding and functional parts of an object O. Thus, $funcPart(knife) = blade$ represents the knowledge that the blade is the functional part of the knife. Similarly, $holdPart(knife) = handle$ represents the knowledge that the handle is the holding part of the knife.

Aspect (Θ_{F_i})	Percept ($f_{i,j}$)	Mass ($m_{f_{i,j}}$)
Θ_{F_1}	$holdPart(O)$	$m_{f_{1,1}}$
Θ_{F_2}	$funcPart(O)$	$m_{f_{2,1}}$
Θ_{F_3}	$hasSharpEdge(O)$	$m_{f_{3,1}}$
Θ_{F_4}	$hasPointyTip(O)$	$m_{f_{4,1}}$
Θ_{F_5}	$hasOpening(O)$	$m_{f_{5,1}}$
Θ_{F_6}	$near(O, G,$part$)$	$m_{f_{6,1}}$
Θ_{F_7}	$grasped(O,$part$)$	$m_{f_{7,1}}$
Θ_{F_8}	$dirty(O)$	$m_{f_{8,1}}$
Θ_{F_9}	$inUse(O, H)$	$m_{f_{9,1}}$

$hasSharpEdge(O), hasPointyTip(O)$ and $hasOpening(O)$ represent the perception of various physical features on object O. In the case of knife we use algorithms developed by [37] to extract shape feature information from the object using object meshes. We then segment the objects (handle and blade) based on their relative sharpness.

$near(O, G,$part$)$ represents the location of a set of graspable points G on an object O in relation to a certain object part (holding or functional part). Thus, $near(knife, G, holdPart(knife) = handle)$ states that there are grasp points near the handle. The grasp points may be extracted from visual point clouds using algorithms developed by [38] that identify antipodal grasp information based on object geometries. We can then group these grasp points based on their location and proximity to the shape features noted above.

$dirty(O)$ represents a measure for whether a certain object is dirty or contains food particles. Thus, $dirty(knife)$ describes the knowledge that the knife is dirty. The value of this predicate is obtained from low-level vision components tasked with monitoring image characteristics of color and homogeneity.

$grasped(O,$part$)$ represents the agent's knowledge that it has grasped a certain part of the object. For example, $grasped(knife, holdPart(knife) = handle)$ represents the knowledge that the agent has grasped the handle.

$inUse(O, H)$ represents the agent's observation that an object O is currently in use by a person or agent H.

We selected these particular visual aspects because of their significance to the rules that we will discuss in more detail in the below sections. There is a potentially huge number of semantic aspects and relations in the environment and it would not be possible for the agent to keep track of them all. Our approach simplifies the task for PAC and the vision system to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2016.2615326, IEEE Transactions on Cognitive and Developmental Systems

only look for certain relevant perceptual features based on the agent's current context. We envision that our set of perceptual aspects, F, may change dynamically to include and exclude percepts as contexts and situations change over time.

C. Relevant Contextual Items, C

Knowledge of the agent's current context is provided to CALyX by certain high-level processing components such as the agent's belief, planning, and goal management systems. The context is representative of the agent's beliefs, goals, desires, and intentions, along with certain other abstract constructs in the agent's situation. Together, these contextual items, processed as predicates, represent qualitatively the agent's abstract context, i.e., knowledge not directly perceivable.

Let $C = \{\Theta_{C_1}, \Theta_{C_2}, \ldots, \Theta_{C_N}\}$ be the set of all contextual aspects an agent may need to interpret. Each contextual aspect $\Theta_{C_i} = \{c_{i,1}, c_{i,2}, \ldots, c_{i,M}\}$ has M mutually-exclusive candidate contextual states, which come from the high-level components as a BoE, $\mathcal{E}_{C_i} = \{\Theta_{C_i}, m_{\Theta_{C_i}}(\cdot)\}$. We will use $m_{c_{i,j}}$ to denote the candidate mass values of the contexts, where $i \in \{1 \ldots N\}$ and $j \in \{1 \ldots M\}$.

For the purposes of our example, similar to our representation of perceptual aspects, we will represent the agent's contextual knowledge with two binary contextual aspects. The first contextual aspect represents the agent's current domain or setting, L, and it includes a contextual value (context) of being a kitchen helper and its negation: $\Theta_{C_1} = \{c_{1,1}, \neg c_{1,1}\}$. The second contextual aspect represents the agent's tasks in the kitchen while playing two different social roles: (1) as a primary actor using objects, and (2) as a supporting assistant giving objects to others. This aspect includes two contextual values: $\Theta_{C_2} = \{c_{2,1}, c_{2,2}\}$. The contexts and masses for each of the two aspects, can be obtained from the agent's belief and planning systems:

Aspect (Θ_{C_i})	Context($c_{i,j}$)	Mass ($m_{c_{i,j}}$)
Θ_{C_1}	$domain(X, L)$	$m_{c_{1,1}}$
Θ_{C_2}	$task(X, use, O)$	$m_{c_{2,1}}$
	$task(X, give, O)$	$m_{c_{2,2}}$

$domain(X, L)$ represents the agent's, X, current domain, L. For example, $domain(self, kitchen)$ represents the knowledge that the agent is currently in the domain of working in the kitchen. The reason for the domain context is to help the agent constrain the set of possible affordances available on the object to the domain it is currently in. For example, the agent might not need to consider affordances of a knife as a camping tool or as a self-defense tool, while it is functioning as a kitchen helper. Thus, by choosing a domain, we can restrict what types of affordances the agent needs to reason about in its current task. This is not to say that the agent cannot think creatively or absorb affordance rules from other domains. But, as a simplification for this example, we choose contextual aspects that can help the agent effectively manage the computational complexity of affordance inference.

$task(X, use, O)$ represents the agent's, X, understanding of its current task-related context as being that of "using" object O. For example, $task(self, use, knife)$ means that the current task-context is that of the agent using the knife for its intended purpose of cutting.

$task(X, give, O)$ represents the agent's, X, understanding of it's current task-related context as being that of "giving" or "handing over" object O. For example, $task(self, give, knife)$ means that the current context is that of the agent handing over the knife to another.

We will discuss the rules themselves in more detail in the next sections.

D. Cognitive Affordances, A

The next part of the representational framework are the cognitive affordances A computed by CALyX based on applicable rules in WM. We use affordances here to represent action possibilities available to the agent at any given moment in time. The affordances are represented semantically with predicates for action possibilities.

Let $A = \{\Theta_{A_1}, \Theta_{A_2}, \ldots, \Theta_{A_N}\}$ be the set of N different cognitive affordance aspects. Each aspect $\Theta_{A_i} = \{a_{i,1}, a_{i,2}, \ldots, a_{i,M}\}$ has a set of M mutually-exclusive candidate affordance values (affordances), which come as a BoE, $\mathcal{E}_{A_i} = \{\Theta_{A_i}, m_{\Theta_{A_i}}(\cdot)\}$. We will use $m_{a_{i,j}}$ to denote the candidate mass values of the contexts, where $i \in \{1 \ldots N\}$ and $j \in \{1 \ldots M\}$.

For the purposes of our example, we will represent the agent's affordances with eight binary affordance aspects, each aspect with an affordance and its negation. Thus, $\Theta_{A_i} = \{a_{i,j}, \neg a_{i,j}\}$, where $i \in \{1 \ldots 8\}$ and $j \in \{1\}$. The percepts and masses for each of the eight aspects, can be obtained from our rules:

Aspect (Θ_{A_i})	Affordance ($a_{i,j}$)	Mass ($m_{a_{i,j}}$)
Θ_{A_1}	$cutWith(X, O)$	$m_{a_{1,1}}$
Θ_{A_2}	$pierceWith(X, O)$	$m_{a_{2,1}}$
Θ_{A_3}	$containWith(X, O)$	$m_{a_{3,1}}$
Θ_{A_4}	$graspable(X, O, \text{part})$	$m_{a_{4,1}}$
Θ_{A_5}	$sanitizeable(X, O)$	$m_{a_{5,1}}$
Θ_{A_6}	$useable(X, O)$	$m_{a_{6,1}}$
Θ_{A_7}	$giveable(X, O, H)$	$m_{a_{7,1}}$
Θ_{A_8}	$setOnTable(X, O, U)$	$m_{a_{8,1}}$

We will discuss each of these affordance aspects below:

1) Commonsense Physical Affordances: Various objects in the kitchen like knives, forks, spoons pots, pans, and appliances offer the agent with various physical affordances. Here we will consider three such affordances offered by a number of different objects: (1) $cutWith(X, O)$, (2) $pierceWith(X, O)$ and (3) $containWith(X, O)$, each representing an affordance of an object O available to an agent X in a kitchen scenario. Objects can have one or more of these affordances. For example, knife can have the affordance of cutting as well as piercing, depending on the shape of the knife.

2) Grasp Affordances: Many objects in the kitchen tend to have a use for which they are designed, and accordingly allow for holding and using the object in a particular way for this intended purpose. For example, knives are designed for cutting and thus can be grasped by the handle and used to cut with the blade. We account for grasp affordances with a $graspable(X, O, \text{part})$ predicate, which represents that the

object O is graspable by agent X at a certain part of the object. Thus, $graspable(self, knife, holdPart(knife) = handle)$ represents that the knife's handle has a grasp affordance in the current context.

3) Social Affordances: In the context of a kitchen, there are a number of social norms and rules that apply to ensure safety, etiquette, cleanliness and a generally friendly atmosphere. These rules present social affordances, i.e., action possibilities related to social interaction that can be made available to the agent. Here we consider $sanitizeable(X, O)$, which represents the possibility of washing and cleaning an object. As we will see with respect to the rules in the next section, social affordances can be represented both explicitly, as in $sanitizeable(X, O)$, and implicitly via socially-derived rules for conduct, e.g., presenting the handle first when giving objects to others.

4) Object Manipulation Affordances: Once the agent has begun interacting with the object, certain new affordances are made available to the agent: $useable(X, O)$ represents the agent's X ability to use object O for its intended purpose; $giveable(X, O, H)$ represents the agent's X ability to give object O to a human or another agent, H; and $setOnTable(X, O, U)$ represents the agent's X ability to place object O on surface U. These affordances allow the agent to consider its action possibilities once it is in the possession of the object.

Now, we recognize that these affordance are always available to the agent: the agent can cut, grasp, give, wash and place the knife at any time. Our affordance representation does not deny that latent affordances may exist in objects, but merely attaches uncertainties to their potential applicability. Certain dormant affordances will have low uncertainties unless certain contextual situations arise, and our rules seek to capture this type of reasoning with affordances.

It could also be argued that there are many more affordances for knives, and that we are limited in considering only a few. We agree with this argument and only present this exemplary set for demonstration and evaluation purposes. In reality, there are many more affordances, possibly unlimited, and our cognitive affordance inference framework can reason about all of them simultaneously. Although we will not address the issue of whether or not there are infinitely many affordances, we will contend that only a finite subset of them is relevant in any given set of contexts, applicable at a particular moment in time.

E. Cognitive Affordance Rules, R

The fourth part of our representational framework is the set of rules, R, that represent the cognitive affordance aspects, A, of the perceptual aspects, F, in a contextual aspects, C. We will present an exemplary set R of rules for the handover example below.

Let $R = \{\Theta_{R_1}, \Theta_{R_2}, \ldots, \Theta_{R_N}\}$ be the set of N different cognitive affordance rule aspects. Each rule aspect $\Theta_{R_i} = \{r_{i,1}, r_{i,2}, \ldots, r_{i,M}\}$ has a set of M mutually-exclusive candidate rule values (rules), which come as a BoE, $\mathcal{E}_{R_i} = \{\Theta_{R_i}, m_{\Theta_{R_i}}(\cdot)\}$. We will use $m_{r_{i,j}}$ to denote the candidate mass values of the contexts, where $i \in \{1 \ldots N\}$ and $j \in \{1 \ldots M\}$.

For the purposes of our example, we will represent the agent's affordances with 18 rule aspects (representing 18 rules), each aspect with a rule and its negation. Thus, $\Theta_{R_i} = \{r_{i,j}, \neg r_{i,j}\}$, where $i \in \{1 \ldots 18\}$ and $j \in \{1\}$. The percepts and masses for each of the 18 aspects, can be obtained from our rules:

Generally, the rules are of the form:

$$r^{i,j}_{m_{fc \to a}} := f \wedge c \implies a$$

The belief function, $Bel(R)$, captures the total support that can be committed to a rule, R, without also committing to the negation of the rule. The plausibility of R is, $Pl(R)$, corresponds to the total belief that does not contradict R. Together, the belief and plausibility represent the uncertainty interval, $[\alpha = Bel(R), \beta = Pl(R)]$. Thus, we write the rules in the form:

$$r^{i,j}_{[\alpha_{i,j}, \beta_{i,j}]} := f \wedge c \implies a$$

Below, we show each of the 18 rules for this example, presenting the uncertainty intervals for each of the rules. We have chosen uncertainty intervals in such a way that the more specific the rule, the more certainty and higher degree of belief the agent has about that particular rule. Thus, more specific the rule, narrower the uncertainty interval and higher the values for α and β. Also, for ease of reading, we have omitted the index $j = 1$.

Commonsense Physical Rules:

$r^1_{[0.8,1]} := hasSharpEdge(O) \wedge$
$domain(X, kitchen) \implies$
$cutWith(X, O)$
$r^2_{[0.8,1]} := hasPointyTip(O) \wedge$
$domain(X, kitchen) \implies$
$pierceWith(X, O)$
$r^3_{[0.8,1]} := hasOpening(O) \wedge$
$domain(X, kitchen) \implies$
$containWith(X, O)$

General Social Rules:

$r^4_{[0.95,0.95]} := dirty(O) \wedge$
$domain(X, kitchen) \implies$
$sanitizeable(X, O)$
$r^5_{[0.95,0.95]} := \neg inUse(O, H) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O))$
$r^6_{[0.95,0.95]} := \neg inUse(O, H) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O))$

General Object Grasp Rules:

$r^7_{[0.55,0.95]} := near(O, G, holdPart(O)) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O))$
$r^8_{[0.55,0.95]} := \neg near(O, G, holdPart(O)) \wedge$
$near(O, G, funcPart(O)) \wedge$
$domain(X, kitchen) \implies$

$graspable(X, O, funcPart(O))$

Task-based Social Rules:

$r^{9}_{[0.8,0.9]} := near(O, G, holdPart(O)) \wedge$
$task(X, use, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O))$

$r^{10}_{[0.8,0.9]} := near(O, G, funcPart(O)) \wedge$
$task(X, give, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O))$

$r^{11}_{[0.95,0.95]} := near(O, G, holdPart(O)) \wedge$
$\neg dirty(O)$
$task(X, use, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O))$

$r^{12}_{[0.95,0.95]} := near(O, G, funcPart(O)) \wedge$
$\neg dirty(O)$
$task(X, give, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O))$

Object Interaction Rules:

$r^{13}_{[0.8,0.9]} := grasped(O, holdPart(O)) \wedge$
$task(X, use, O) \wedge$
$domain(X, kitchen) \implies$
$useable(X, O)$

$r^{14}_{[0.8,0.9]} := grasped(O, funcPart(O)) \wedge$
$task(X, give, O) \wedge$
$domain(X, kitchen) \implies$
$giveable(X, O, H)$

$r^{15}_{[0.55,0.95]} := grasped(O, holdPart(O)) \wedge$
$domain(X, kitchen) \implies$
$setOnTable(X, O, T)$

$r^{15}_{[0.55,0.95]} := grasped(O, funcPart(O)) \wedge$
$domain(X, kitchen) \implies$
$setOnTable(X, O, U)$

$r^{17}_{[0.8,0.9]} := grasped(O, holdPart(O)) \wedge$
$dirty(O) \wedge$
$domain(X, kitchen) \implies$
$setOnTable(X, O, U)$

$r^{18}_{[0.8,0.9]} := grasped(O, funcPart(O)) \wedge$
$dirty(O) \wedge$
$domain(X, kitchen) \implies$
$setOnTable(X, O, U)$

Rules $r^1 - r^3$ relate to the agent's general commonsense understanding of the physical properties of objects. E.g., sharp edges provide a cutting affordance.

Rules $r^4 - r^6$ prescribe several social rules in a kitchen environment. For example, dirty objects need to be sanitized and only objects not currently used by someone else are available for grasping.

Rules r^7 and r^8 provide general rules on grasping objects. For example, if there are grasp points near the handle of a knife, then the handle has a grasp affordance, and if there are no grasp points near the handle, but there are some near the blade, then the blade has a grasp affordance.

Rules $r^9 - r^{12}$ relate to social etiquette and convention when using and giving objects. These rules provide a narrowing context depending on the agent's current task of using the object itself or giving the object to another. For example, when handing over an object, it is "proper" to hold the functional part of the object (e.g., knife) and present the handle towards the recipient.

Rules $r^{13} - r^{18}$ are kitchen rules that apply when the agent is manipulating and interacting with the object. For example, if the agent is holding the knife by its blade and is tasked with giving it to a human, then it is afforded the possibility of giving the knife.

F. Handover Inference

We will now turn to how a robot with our computational framework can use these rules to reason through the process of handing over a knife. Consider a robot helper receiving instructions from a human, Julia. Suppose Julia says to the robot: *"Bring me something clean I can use to cut this tomato."* The robot will need to parse this request and infer affordances of objects in its environment in context. Before we can describe how the robot can perform this inference for the knife-handover example, we will first describe our inference process and algorithm more generally.

G. Inferring Affordances with Uncertain Logic

The goal of defining cognitive affordance models is to infer object affordances based on (1) their perceivable features, (2) the known context, and (3) general domain and common sense knowledge. We propose to start with the first prototype inference algorithm shown in Algorithm 1 and refine it to tailor it specifically to a cognitive affordance model. The algorithm takes three parameters: (1) a BoE of candidate perceptions $\{\Theta_F, m_f\}$ is provided by the low-level vision system, (2) a BoE of relevant contextual items $\{\Theta_C, m_c\}$ provided by a knowledge base or some other part of the integrated system that can provide context information, and (3) a table of cognitive affordance rules R. Each rule $r_{f \wedge c \to a}$ in R is indexed by a feature perception f and a set of contextual items c, and dictates the mass assigned to $Bel(a)$ and $Pl(a)$ when the system believes the degree to which object features f were detected and that contextual items c are true. Here, a is a complex logical expression representing the affordance that can be derived from the perceived features f in context c.

The inference algorithm then examines each rule $r_{fc \to a} \in R$ (line 5), and m_{fc} is determined by performing $m_f \otimes m_c$ (line 6), where m_f specifies the degree to which object feature f is believed to be detected, and m_c specifies the degree to which each of the rule's associated contextual items is believed to be true. *Uncertain Modus Ponens* is then used to obtain m_a from $m_{fc \to a}$ and m_{fc} (line 6).

Note that since we allow multiple affordance rules to be considered, multiple affordances may be produced. Multiple rules may produce the same affordances for various reasons, possibly at different levels of belief or disbelief. However, we seek to return the set of *unique* affordances implied by a set of perceptions f.

After considering all applicable affordance rules, we group affordances that have the same content but different mass

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2016.2615326, IEEE Transactions on Cognitive and Developmental Systems

Algorithm 1 getAffordance($\{\Theta_F, m_f\}$,$\{\Theta_C, m_c\}$,R)

1: $\{\Theta_F, m_f\}$: BoE of candidate perceptual features
2: $\{\Theta_C, m_c\}$: BoE of relevant contextual items
3: R: Currently applicable rules
4: $S = \emptyset$
5: **for all** $r \in R$ **do**
6: $\quad S = S \cup \{(m_f \otimes m_c) \odot m_{r=fc \to a}\}$
7: **end for**
8: $G = group(S)$
9: $\psi = \emptyset$
10: **for all** group $g_a \in G$ **do**
11: $\quad \psi = \psi \cup \{\bigcap_{j=0}^{|g_a|} g_{a_j}\}$
12: **end for**
13: **return** ψ

assignments (line 8), and use Yager's rule of combination (line 11) to fuse each group of identical affordances, adding the resulting fused affordance to set ψ. This set then represents the set of affordance implied by the perceived features f.

Finally, we can use the confidence measure λ to determine whether an inferred affordance should be realized and acted upon. For example, we could check the confidence of each affordance $a \in \psi$ on its uncertainty interval $[\alpha_i, \beta_i]$: if $\lambda(\alpha_i, \beta_i) \leq \Lambda(c)$ (where $\Lambda(c)$ is a confidence threshold, possibly depending on context c), we do not have enough information to confidently accept the set of inferred affordances and can thus not confidently use the affordances to guide action. However, even in this case, it might be possible to pass on the most likely candidates to other cognitive systems. Conversely, if $\lambda(\alpha_i, \beta_i) > \Lambda(c)$, then we take the inferred affordance to be certain enough to use it for further processing.

H. DS-Theoretic Handover Inference

Returning to our knife-handover example, the robot, $X = self$, parses the request from Julia (i.e., for an object she can use to cut a tomato) and assigns its own task context and determines the types of affordances it is interested in exploiting in the kitchen environment. The agent is confident that it is in the kitchen context and that it is in the context of handing over an object in the kitchen, and assigns context masses as follows:

$domain(self, kitchen)$: $m_{c_{1,1}} = 1.0$
$task(self, give, O)$: $m_{c_{2,1}} = 0.95$
$task(self, use, O)$: $m_{c_{2,1}} = 0.05$

Specifically, the robot's cognitive architecture includes a natural language processing component, which processes Julia's instruction. The phrase "bring me something" is taken by the robot to indicate a "give" context as opposed to a "use" context. This contextual information is obtained outside of CALyX and passed into it as input. Similarly, the robot's cognitive architecture includes belief and goal management components, which process the robot's current role as a kitchen helper and compute the likelihood that is in the "kitchen" domain. This, too, is passed as input to our CALyX system.

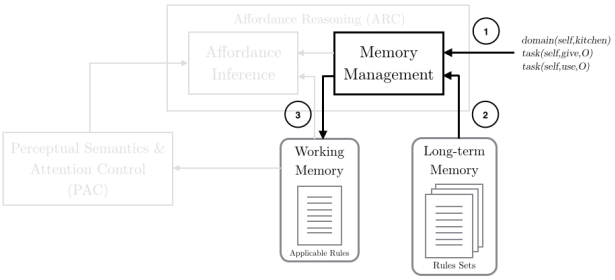

Fig. 3: Selecting Applicable Rules Based on Context

CALyX's Memory Management subcomponent receives this contextual information and selects applicable rules (Fig. 3, step 1). LTM potentially contains a large set of rules across various contexts that the robot has acquired over its lifetime. Given the specific domain and task contexts, the Memory Management subcomponent selects a subset of applicable rules from the LTM (step 2) and populates the WM with these rules (step 3).

Although the context has been established and the applicable rules have been identified, at this point the robot is not yet ready to do any affordance inference because it does not know whether the perceptual aspects specified by the rules are satisfied. Given the set of applicable rules in WM (Fig. 4, step 1), PAC can guide or direct sensory processing systems (step 2) to determine the uncertainties associated with each of the perceptual aspects specified in the set of applicable rules (step 3). For example, based on its set of rules, it knows to look specifically for certain visual percepts relevant to the rules, such as $near()$, $sharpEdge()$, and so on. Note, at this point, the robot is not aware of a specific object that it needs to find, but with PAC in combination with the low-level vision system, it can scan its environment and examine each object more closely to determine which perceptual aspects specified above are satisfied.

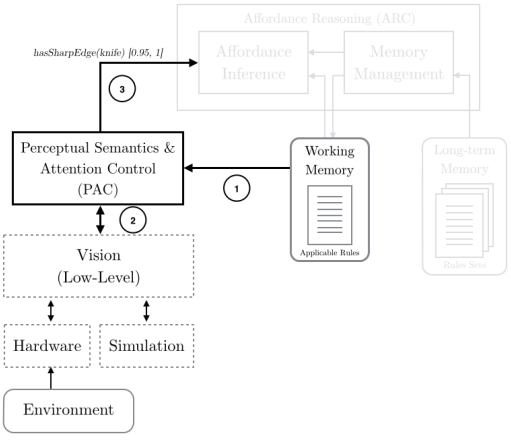

Fig. 4: Determining Beliefs of Perceptual Aspects

1) Spot the Knife - Directed Perception with PAC: Let us assume that the agent spots a knife on the counter. Upon examining the physical features of the knife, the agent determines masses for percepts in each of its perceptual aspects.

For example, upon spotting the knife, PAC assigns the percept $hasSharpEdge(knife)$ with a mass $m_{f_{3,1}} = 0.95$. However, because it is slightly unsure it also assigns the possibility that the knife either has or does not have a sharp edge, $\{hasSharpEdge(knife), \neg hasSharpEdge(knife)\}$, with a mass $= 0.05$. With these two masses, support for the percept $hasSharpEdge(knife)$ falls within the interval $[\alpha, \beta] = [0.95, 1]$. Similarly, we compute uncertainty intervals for all the relevant percepts, when the agent sees the knife:

$hasSharpEdge(knife)$ [0.95, 1]
$hasPointyTip(knife)$ [0.8, 0.9]
$hasOpening(knife)$ [0, 0]
$near(knife, G, holdPart(knife) = handle)$ [0.95, 0.95]
$near(knife, G, funcPart(knife) = blade)$ [0.95, 0.95]
$grasped(knife, holdPart(knife) = handle)$ [0, 0]
$grasped(knife, funcPart(knife) = blade)$ [0, 0]
$dirty(knife)$ [0.31, 0.81]
$inUse(knife, H)$ [0, 0]

To summarize, the agent has detected a knife that has a sharp edge and can be grasped, but it is not entirely sure if it is clean or dirty. Note, the agent has yet to pick up and grasp the object, so the $grasped()$ predicates evaluate to $[0, 0]$, which means *logically false* with maximum certainty.

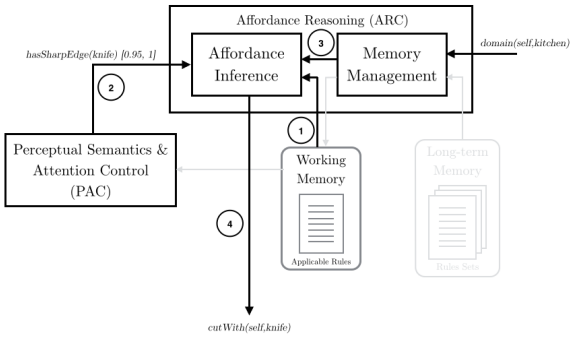

Fig. 5: Performing Inference of Cognitive Affordances

2) Affordance Inference with ARC: ARC examines each rule $r^i_{[\alpha_i,\beta_i]}$ (line 5 of Algorithm 1) in WM (Fig. 5, step 1), and m_{fc} is determined by performing $m_f \otimes m_c$ (line 6), where m_f specifies the degree to which the percept f is believed to be observed (Fig. 5, step 2), and m_c specifies the degree to which each of the rule's associated contextual value is believed to be true (Fig. 5, step 3). DS-based modus ponens is then used to obtain m_a from $m_{fc \to a}$ and m_{fc} (line 6) (Fig. 5, step 4).

For example, consider rule r^1:

$r^1_{[0.8,1]} := hasSharpEdge(O) \wedge domain(X, kitchen) \implies cutWith(X, O)$

The agent will apply perceptual and contextual information as follows, to determine the affordance implied by the rule:

$r^1_{[0.8,1]}(m_r = 0.8) :=$
$hasSharpEdge(knife)(m_f = 0.95) \wedge$
$domain(self, kitchen)(m_c = 1.0) \implies$

$cutWith(self, knife)(m_a = (m_f \otimes m_c) \odot m_r = 0.76)$

The uncertainty interval for the rule can then be computed as $[0.76, 1]$.The agent will subsequently perform this analysis for each of the other rules in the set to determine uncertainty intervals for the implied affordances.

To be able to generate a set of unique affordances, a, implied by feature, f, after considering all applicable affordance rules, we thus group affordances that have the same semantic content but different mass assignments (line 8) and use Yager's rule of combination (line 11) to fuse each group of identical intentions, adding the resulting fused intention to set ψ. Thus, affordances from rules r^1, r^3, r^6 and r^{17} will be fused as these rules all apply to the affordance of $graspable(self, knife, holdPart(knife) = handle)$. The agent will also fuse together rules r^2, r^4, r^7 and r^{18} as these rules all apply to the affordance of $graspable(self, knife, funcPart(knife) = blade)$. The agent will further fuse together rules r^{13}, r^{14}, r^{15} and r^{16} as these rules all apply to the same affordance of $setOnTable(self, knife, table)$.

Based on the application of each rule to the semantic visual percepts and contextual items, and fusing rules with similar implied affordances together we can generate a list of unique affordances available to the agent at the current moment in time, when it has seen the knife:

Available affordances (Upon seeing the knife), ψ
$cutWith(knife)[0.76, 1], \lambda = 0.29$
$pierceWith(knife)[0.64, 1], \lambda = 0.16$
$containWith(knife)[0, 1], \lambda = 0$
$sanitizeable(knife)[0, 0.95], \lambda = 0.004$
$graspable(self, knife,$ $holdPart(knife) = handle)[0.96, 0.99], \lambda = 0.78$
$graspable(self, knife,$ $funcPart(knife) = blade)[0.98, 0.99], \lambda = 0.88$

This information is passed from CALyX to other parts of the cognitive architecture like the robot's goal and action management system, which performs different operations based on these measured uncertainty intervals and associated λ confidence measures. The agent might decide that because there is a high degree of confidence that the object under consideration has a $cutWith$ affordance, it will choose to grasp it and then select to grasp it at the blade, given the context of a handover. Because it is unclear that the knife is dirty, the agent is less confident that the knife needs cleaning.

3) Grasp the Knife - Iterated Directed Perception with PAC: Affordance inference in this manner with CALyX is an iterative or cyclical process and affordances are computed and re-computed continuously to guide the robot's next actions. Thus, once the robot has grasped the knife, CALyX is once again tasked with inferring affordances to determine what the robot can and cannot do with the knife that it is holding. Here the contexts remain the same, so the rules in WM are likely unchanged. However, because the world state has changed (i.e., the knife is no longer on the table but in the hands of the robot), the perceptual aspects specified by the rules potentially have different truth values. PAC is once again called upon to update the uncertainty intervals associated with the visual perceptual aspects as follows:

$hasSharpEdge(knife)$ [1, 1]
$hasPointyTip(knife)$ [0.95, 0.95]
$hasOpening(knife)$ [0, 0]
$near(knife, G, holdPart(knife) = handle)$ [0.95, 0.95]
$near(knife, G, funcPart(knife) = blade)$ [0.05, 0.05]
$grasped(knife, holdPart(knife) = handle)$ [0, 0]
$grasped(knife, funcPart(knife) = blade)$ [1, 1]
$dirty(knife)$ [0.95, 1]
$inUse(knife, self)$ [1, 1]

It updates the $grasped()$ predicate information because it has now grasped the knife's blade. The robot also detects that the knife is dirtier than initially determined, and there are no longer any grasp points available near the blade, because it is already holding the blade. It also is more certain that the knife has a sharp edge and pointed tip. Finally, it also knows that because it is using the knife, the knife is "inUse".

4) Iterated Affordance Inference with ARC: Based on this update, ARC will re-calculate the uncertainty intervals and confidence measures associated with its affordances, as follows:

Available affordances (after grasping knife), ψ
$cutWith(knife)[0.8, 1], \lambda = 0.34$
$pierceWith(knife)[0.76, 1], \lambda = 0.29$
$containWith(knife)[0, 1], \lambda = 0$
$sanitizeable(knife)[0.9, 1], \lambda = 0.564$
$graspable(self, knife,$ $holdPart(knife) = handle)[0.5, 0.9], \lambda = 0.07$
$graspable(self, knife,$ $funcPart(knife) = blade)[0.03, 0.9], \lambda = 0.0006$
$setOnTable(self, knife, table)[0.84, 0.99], \lambda = 0.43$
$useable(self, knife)[0, 0.95], \lambda = 0.004$
$giveable(self, knife, Julia)[0.9, 1], \lambda = 0.56$

Having detected that the knife is dirtier than initially determined, the agent now has a higher confidence that the knife has a sanitizeable affordance. The agent also has additional affordances available. It has high confidence that the knife is giveable to Julia and that the knife can be set on the table. It knows that the knife is not currently useable to cut things by itself, mainly because it is holding the blade and the current context is a handover, and not use.

Once again, this information is passed to other parts of the agent's cognitive architecture like goal and action management systems, which perform different operations based on these measured uncertainties. The agent might decide to choose to realize one or more of the above observed affordances.

V. Experiment - Multi-Domain, Multi-Scenario Handover

A. Introduction

In our first example of a kitchen helper handing over a knife, we demonstrated the capability of our framework to reason about cognitive affordances of handing over objects. We limited the experiment to one domain (i.e., the kitchen helper) and we focused on a simple set of rules that govern social interactions in this domain. We demonstrated a flexible reasoning process that took into account social context. In this second experiment, we extended the object handover task and compared interactions across various domains and expanded our notion of object affordances to "social affordances" offered by humans in the scenario, as well. Object handovers are often complex interactions that involve more social intelligence than just reasoning about physical or social aspects of the objects alone. Often, in human-human interaction scenarios, the giver must tune into various social cues (e.g., eye-gaze) offered by the receiver indicating whether a handover must be initiated or not. Social context is highly relevant to object handovers, and we will demonstrate the flexibility of our framework in reasoning about situations when similar observations of the environment have very different meanings in different contexts.

For this experiment we built on some of the extensive previous work in determining parameters of a handover from a physical and temporal sense and in deciding its timing and trajectories [39], [40]. With regards to social cognition during handovers, Strabala et al. have extensively studied social cues that are crucial to coordinating a handover [16]. They provide four exemplary domains – elder care-giver, mechanic-helper, fire-brigade volunteer, and flyer-handout giver – that all feature a handover activity, but under very different contexts. In this experiment, we implemented an affordance-based handover reasoning mechanism for these four socially distinct domains. While Strabala et al. focused on discovering a unified handover structure that might apply in all these four domains, we retained the richness and distinctions of these four different domains, and instead reasoned about the affordance of "transferability" of an object prior to the handover. We will begin by describing the four domains in more detail.

B. Domains

We will consider four exemplary domains as originally introduced by Strabala et al., and expanded by us, as follows:.

1) **Care-giver** at an elder care facility: In this domain, a care-giver or assistant holding a glass of water is tasked with handing it over to the patient. The care-giver must be sure that the patient is ready to receive the water before beginning the transfer process. In many cases, it is further desirable that the patient make eye contact and orient her body towards the care-giver. Moreover, in this scenario, it is often not appropriate for the care-giver to handover the water when the patient is not attentive and looking away, even if the patient is reaching out or verbally requesting the water.
2) **Mechanic's Helper:** In this domain, a helper is tasked with handing over a wrench to a mechanic. The social cues in this domain, while similar in structure to the cues in the care-giver scenario, are vastly different in how they influence the interaction. Here, the mechanic may be under a car or focused on the task at hand, and, therefore, not attentive to the assistant. So the helper must be more attentive to other signals such as the mechanic reaching out with an outstretched arm and verbally requesting the wrench. Eye contact and body orientation may be less important in this

scenario. Indeed, sometimes the mechanic may be facing the helper but not be ready for the wrench, hence the handover must be confirmed through verbal signals or by reaching out.

3) **Fire-brigade:** In this domain, a helper or volunteer, who is one of many agents (humans and robots) in a line, is tasked with passing buckets of water from a source to the scene of a fire. This domain is different from the prior two domains in its ignorance of social context. Generally, there is a known procedure for swinging buckets and the urgent nature of the situation has eliminated the role of social etiquette. In this domain, it is often permissible to begin a handover procedure without many social cues like eye gaze, reaching, or verbalizing. Bodies are often not facing each other and the only real condition to begin the transfer is possession of the bucket of water. As long as the giver is holding a bucket of water, the transfer should begin.
4) **Flyer-handouts:** In this domain, a giver is tasked with handing out flyers on a busy university sidewalk. As Strabala et al. note, the giver has no prior relationship with the people on the sidewalk, and so established social norms apply [16]. In fact, this domain is the complement of the Fire-brigade domain because many social cues, including eye gaze, body orientation, reaching out actions and verbal confirmation, all apply. The passerby who is interested in receiving a flyer is likely to face the giver, make eye contact and request a handout while reaching out. The giver would be considered rude if she imposed a flyer on someone who was merely walking towards her or if the passerby provided verbal cues and hand motions that might appear to be a reaching action, when in fact they were signaling the opposite.

C. Representing Social Cues and Domain Rules

To represent these domains, we first selected well-established social cues offered by the receiver to the giver that are often considered relevant to handovers: eye gaze or eye contact, verbal confirmation, the action of reaching out and requesting the object, and body orientation [16]. We represented these social cues as predicates (with intuitive semantics) $eyeGaze(X)$, $verbalSignal(X)$, $reachingOut(X)$, and $bodyFacing(X)$, respectively, where X refers to the receiver. In addition to the social cues, we also represented the information that the robot is holding object O (i.e., the object to be handed over) with the predicate $holding(self, O)$. We represented the affordance of transferability with the predicate $transferable(O, X)$ to mean that object O is transferable to receiver X.

With these social cues, we assigned simple social rules or norms that are applicable generally (but to varying degrees) across all four domains, as follows:

$r^1 := eyeGaze(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^2 := verbalSignal(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^3 := reachingOut(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^4 := bodyFacing(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^5 := holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^6 := eyeGaze(X) \wedge bodyFacing(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^7 := verbalSignal(X) \wedge reachingOut(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$
$r^8 := eyeGaze(X) \wedge bodyFacing(X) \wedge verbalSignal(X) \wedge reachingOut(X) \wedge holding(self, O) \wedge goal(handover) \implies transferable(O, X)$

D. Representing the Domain Distinctions

We represented the different levels of importance of the rules in the four domains by assigning them different uncertainties depending on the domain. We selected uncertainties for the rules from three settings, as follows:

High = [0.95, 1]
Medium = [0.5, 0.6]
Low = [0.31, 0.81]

Here, "High" refers to the setting in which the rule is believed to be true with a low uncertainty (i.e., high degree of certainty). Similarly, the "Medium" and "Low" levels each correspond to a setting in which the rules are believed to be true, but with medium and high uncertainty, respectively.

We assigned an uncertainty level for each the rules in each of the domains, as shown in Table I. For example, since all social cues are important for the Flyer-handout domain, rule 8 featuring all of these cues is the most important one. In contrast, in the Fire-brigade domain, the only rule that is the most important is the commonsensical rule 5, which requires possession of the bucket before transfer, without any other social overhead. In the Elder care domain it is important that the patient face the giver and make eye contact, whereas in the mechanic domain, it is important that the mechanic reach out and ask for a wrench before initiating handover. Accordingly, rules corresponding to these social cues were given more importance and higher certainty.

Rule	Elder Care	Mechanic	Fire-Brigade	Flyer-Handout
r^1	Medium	Low	Low	Low
r^2	Low	Medium	Low	Low
r^3	Low	Medium	Low	Low
r^4	Medium	Low	Low	Low
r^5	Low	Low	High	Low
r^6	High	Low	Low	Medium
r^7	Low	High	Low	Medium
r^8	Low	Low	Low	High

TABLE I: Domain-Specific Rule Uncertainty Assignment

Scenarios	Eye Gaze	Verbal Signal	Reaching Out	Body Facing	Holding
S^1	T	F	F	T	T
S^2	F	T	T	F	T
S^3	T	T	T	T	T
S^4	F	T	F	F	T
$S^1(uncertain)$	sT	sF	sF	sT	sT
$S^2(uncertain)$	sF	sT	sT	sF	sT
$S^3(uncertain)$	sT	sT	sT	sT	sT
$S^4(uncertain)$	sF	sT	sF	sF	sT

TABLE II: Perceptual Uncertainty Assignment for 4 Scenarios

E. Experimental Scenarios

For our experiment, we considered four scenarios that are possible within any or all of these domains. These four scenarios represent the truth settings for the set of perceptual social cues - eye gaze, body orientation, verbal request or reaching out - in a given situation. For example, a mechanic under a car requesting a wrench may not make eye contact or orient their body towards the giver, so the perceptual cues like eye gaze and body orientation would be False in this scenario. But the mechanic is reaching out his arm and requesting a wrench verbally, so the perceptual cue for verbal request and reaching out would be True. We acknowledge the fact that with four social cues, there are 16 possible scenarios. In particular, each scenario would include a combination of true/false settings for the four perceptual cues: eye gaze, body orientation, verbal request or reaching out. However, we have limited our presentation in this paper to the four following exemplary scenarios that were the most informative, across all our domains:

Scenario 1 (S^1): When the receiver is making eye contact and facing the giver, but not providing any verbal cues or reaching out to the giver (e.g., an elderly patient signaling that they are ready to receive water).
Scenario 2: (S^2): When the receiver is verbally requesting the object from the giver and is extending her arm in anticipation of receiving the object. The receiver, however, is busy performing another task and is turned away and not looking at the giver.
Scenario 3: (S^3): Here the receiver is fully engaged in the handover and is providing all social cues.
Scenario 4: (S^4): Here, the receiver is only signaling verbally that he is ready to receive the object, but his attention maybe elsewhere as he is looking and turned away from the giver. He is also not extending his arm or reaching out to the giver.

We also generated four additional scenarios that were variations of the first four. These additional scenarios only differed from the original four in that we diminished the truth and false certainties using the "Somewhat True" and "Somewhat False" uncertainty setting:

True (T) = [0.95, 1]
False (F) = [0, 0.05]
Somewhat True (sT)= [0.62, 0.96]
Somewhat False (sF) = [0.04 0.38]

Here, the "Somewhat True" is an uncertainty setting for perceptual cues that are logically true, but with a greater amount of uncertainty than that of the "True" setting (i.e., wider uncertainty intervals). Similarly, the "Somewhat False" is an uncertainty setting for perceptual cues that are logically false, but with a greater amount of uncertainty than that of the "False" setting.

Overall, we represented the truth of the social cues in all scenarios per uncertainty levels noted above and shown in Table II. Note, these scenarios represent a truth setting for each of the perceptual cues in a situation. The scenarios themselves are independent of the domain. Thus, for example, in a domain, the giver could perceive a set of cues whose truth settings could fit any one of the scenarios.

F. Experimental Results and Discussion

We performed affordance inferences for a total of 32 situations involving our eight scenarios (four certain and four uncertain) across four domains. The results shown in Figure 6 depict the confidence measure across the various scenarios. Note, that higher λ values indicate a tighter uncertainty interval and consequently a more confident or clear outcome. For each scenario, we have depicted blue and yellow plots corresponding to whether or not the uncertainty setting was True/False or Somewhat True/Somewhat False, respectively. For example, in Scenario 1, the blue plots correspond to the case when the $eyeGaze()$ and $bodyFacing()$ cues were assigned a setting of True = [0.95, 1]. In this scenario, the yellow plots correspond to the case when the $eyeGaze()$ and $bodyFacing()$ cues were assigned an uncertainty setting of Somewhat True = [0.62, 0.96].

The results show that our computational framework conforms to our intuitions about these various scenarios. Specifically, we show some interesting distinctions between various domains. For example, in Scenario 2 (S^2), a transferable affordance is considered to exist when a mechanic extends his arm and requests a wrench even if he is not looking at or facing the robot. While, that same type of behavior from a patient in an elder-care facility may not offer the same kind of transferability affordance. We further note that the transferability affordance is always present across various scenarios in the Fire-brigade domain, while it is only present for the Flyer-handout domain when the receiver is fully engaged with the giver. As noted earlier the Fire-Brigade domain represents a domain without much social context and so it is expected that regardless of the social cues, buckets must be passed along. On

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2016.2615326, IEEE Transactions on Cognitive and Developmental Systems

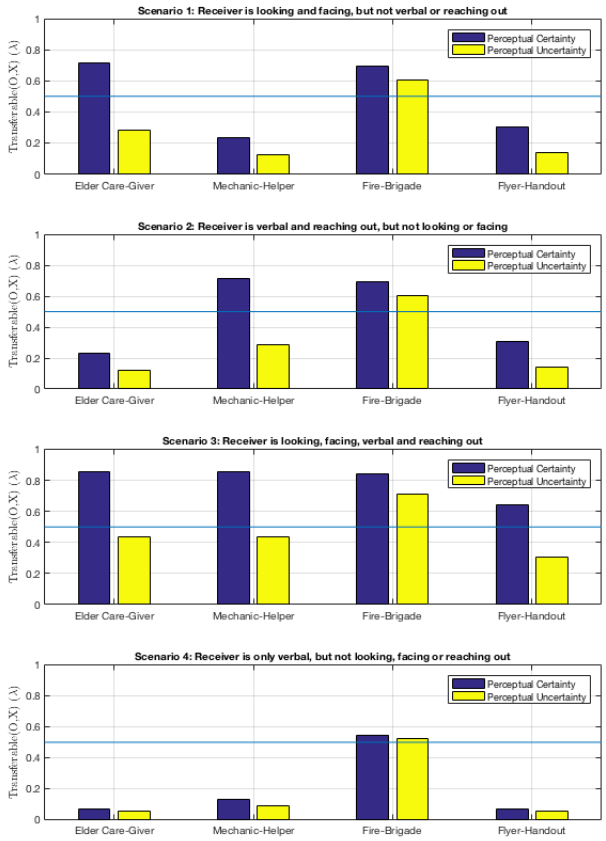

Fig. 6: Plots showing whether an object being held by the robot-giver has an affordance of being "transferable" to the receiver. Each plot corresponds to a particular perceptual scenario (truth values for the existence of various social cues) across four domains - Elder Care-Giver, Mechanic-Helper, Fire-Brigade Volunteer and an agent handing out Flyers. The perceptual cues in a scenario have assigned truth values and an uncertainty interval. The yellow plots correspond to the case when the cues have a higher level of uncertainty. The experiment was performed for four different scenarios and the results correspond to expected human-human interaction behavior in these scenarios and domains.

the other hand, the Flyer-Handout domain represents a domain with deep social context and so it is expected that a transfer should not be initiated unless all social cues are present.

We also show what happens when there is uncertainty introduced to the various perceptions (see yellow plots in Figure 6). It may be uncertain as to whether the receiver is facing the robot or has verbally requested the object. There may also be uncertainties associated with the amount of eye contact or whether an extended arm actually means that the receiver is "reaching out". Interestingly, even with uncertainty the overall structure of the results remained the same but the confidence of the conclusions dropped. That is, the robot is less sure of whether an object has a transferability affordance when it was less certain about whether or not it was perceiving the social cues correctly. In these situations, we can have the robot clarify the social cues and hold back and wait to handover the object. Note, however, the Fire-brigade domain exhibited the least change even in the face of uncertainty. This is because we emphasized the limited role played by social cues in these emergent and largely procedural handover domains.

VI. Discussion

A. CALyX

The novel CALyX architecture for inferring cognitive affordances has an affordance-based reasoning component (ARC) and an attention control and perceptual semantics component (PAC). In much of the past work, affordances have been treated as a subsystem of vision or planning. CALyX allows affordance processing to be a separate component, not subsumed in a vision system or a planning engine. The robot's vision system is then just one of several sensory systems that work with affordance processing. For example, certain perceptual aspects like an object's weight may need the robot's haptic and touch systems to resolve. Thus, an affordance rule involving weight might not involve the vision system at all, and consequently PAC will need to direct the robot's grippers and arms to determine weight information.

In CALyX, affordance rules inform and guide visual attention, that is PAC is guided by the perceptual aspects present in WM at any given moment in time. There have been findings in psychology to support this approach and these findings suggest that affordances influence visual attention by biasing and focusing the visual search on those objects that afford relevant actions. Particularly, researchers have shown support for top-down control in attention processing, task-based priming of visual search [41], [42], [43], and the influence of affordances on visual attention [44], [45], [46]. Developmentally, this is very advantageous as well because as the robot develops, our architecture allows the robot to take into consideration additional perceptual aspects and social cues as it learns them. These social cues will present themselves in the rules and allows the robot to reason about these cues in relation to others as we have shown here. Moreover, the framework allows the robot to revise its rules and beliefs without the need to undergo new rounds of training and learning, as is typically needed for other statistically-inclined affordance learning frameworks.

We have also shown ARC and PAC as separate components of CALyX. ARC and PAC perform different functions, as noted above. However, this separation is not merely a functional one. Each component is autonomous and does not rely on the other beyond an input-output relationship. PAC can potentially interface with not only the vision system but also a haptic system and auditory system to perform perceptual semantic analysis and attention on these other modalities when the affordance rules demand it. Similarly, ARC interfaces with various other higher-level cognitive systems (planning and reasoning) to perform affordance-based reasoning tasks.

CALyX is a flexible architecture and does not preclude maintaining an episodic memory of situations that involve acting on certain affordances and observing the effects. For example, although not described explicitly, CALyX does not exclude the robot from tracking and maintaining the effects

of grasping the knife and handing it over to the human. We noted earlier, that the truth values of the perceptual aspects can change from moment to moment and CALyX does not preclude tracking this information. In fact, observing the effects can influence the uncertainty of some rules in the current context. In the future, when the robot encounters similar contexts, it can remember these rules and reason accordingly.

We recognize that because our framework is rule-based, there is a possibility for rule conflicts. Rule conflicts can arise in many ways. One way is if a feature and its negation produce the same affordance. If there is a rule $(r_1 :\equiv handle \wedge context_1 \implies graspable)$, then there might be a conflict if there is then another rule that states that $(r_2 :\equiv \neg handle \wedge context_1 \implies graspable)$. Conflicts of the type between r_1 and r_2 are currently being resolved in our framework through our Yager fusion operator (which was designed for handling conflicting evidence of this sort) combining uncertainty evidence for $graspable$ from each of rules r_1 and r_2. Another way a conflict may arise is if the same feature in the same context can produce conflicting affordances. For example, if in addition to rule r_1, there is another rule implying a negation of an affordance, like $(r_3 :\equiv handle \wedge context_1 \implies \neg graspable)$. Currently, in the examples we have presented, we have not explicitly reasoned about "negative affordances" like $\neg graspable$. However, it is reasonable to expect that both $graspable$ and $\neg graspable$ could belong to the same frame of discernment, and consequently, fusing conflicting evidence from rules r_1 and r_3 can be handled in a similar manner.

Overall, a main advantage of the proposed architecture is that rational choices can still be made in the face of conflict because of the underlying character of the uncertain logic based inference algorithm. That is, the algorithm considers the rules (conflicting or otherwise) together through the fusion operator and collectively determines implications. Moreover, the architecture could be coupled with a higher-level predictive engine to test expectations against observations and adjust rule uncertainties.

Our architecture does not preclude, and in fact encourages the selection of an optimal choice of affordance, especially when there are many choices available. This is because at any given moment, the architecture presents a set of affordances along with uncertainty intervals. It makes no judgment about which one to select, because that is not the function of the affordance inference process. Selecting an optimal affordance to act on is the job of other cognitive functions like planning and goal management. CALyX provides helpful metrics such as the λ confidence measure, but it does not specify any further requirement in regards to selecting certain affordances.

One concern is that the number of rules increase, there is a potential for higher time and space complexity. The architecture does not explicitly address this beyond suggesting the use of a limited working memory to track applicable rules and only perform inference on this reduced rule-set.

CALyX is capable of withstanding changes in the environment and can adjust for these changes over its cognitive cycles. Short term changes in the environment can impact the agent's decision process. The architecture does not preclude adapting to current demands, even if that means pursuing short term changes temporarily. This is because, the architecture is more focused on moment-to-moment or cycle-to-cycle affordance inference and not more general planning and goal management issues.

Note that the proposed architecture does not preclude the agent from operating with a certain degree of autonomy. With the inclusion of an uncertainty interval and the confidence measure, not only does the architecture allow the agent to ask clarification questions when certain aspects are unclear, but we can also track and quantify the level of autonomy for the agent based on how frequently it encounters ambiguity. An agent with one too many uncertain rules or an agent with faulty sensors that result in uncertain percepts is less autonomous. This allows for the agent's general reasoning abilities to be quantifiable.

We should note that we have implemented our algorithm and the CALyX architecture in connection with a robotic vision system, and we have integrated it into the larger DIARC architecture [32], although these additional aspects are beyond the scope of this paper.

B. Learning the Rules

Thus far, we have not discussed, explicitly, the origin of the cognitive affordance rules and how an agent might generate or learn new rules, because this is not the focus of this paper. Our focus in this paper, instead, was to demonstrate our affordance representation format, inference algorithm, and architecture. Nevertheless, we expect these rules can be learned in a number of different ways from observation, demonstration and exploration, and using multiple different modalities including vision, natural language and haptic information. The agent could learn these types of rules from explicit natural language teaching and instruction as shown by Cantrell et al. [47]. The agent could also learn various rules from observation through reinforcement learning (RL) methods as shown by Bouralias [48] or through exploration from methods as shown by Forestier [49]. Alternatively, the agent could also acquire these rules through data mining and various association rule-mining techniques [50]. We expect rule-learning to be the subject of future work.

VII. Using Inferred Affordances - Future Work

The focus of the proposed architecture is affordance inference (what to do with the inferred affordances is beyond the scope of this paper). Generally, the architecture leaves open the possibility for how the affordance information can be utilized suitably. As discussed before, these affordance computations are useful in planning problems and in guiding a robot's next actions. In fact, the benefits of affordance computations extend further and as outlined below, affordance-based reasoning may be the basis for all manner of creative reasoning and sense-making.

A. Novel Tool Use

Consider the example of a robotic assistant helping a human with an assembly task in which the human has asked the

robot to tighten a loose screw. We would like for the robot to understand this task and the tools needed from an intuitive standpoint such that even in the absence of a screwdriver, it can reason through alternatives and find another substitute.

The robot may know of a number of rules related to its role as a helper. One rule may be: that if agent X is given a task to tighten a flat head screw S, and X sees an object O that has a flat-head edge, then the object O has a $tightenWith$ affordance. This rule can then be represented in DS-theoretic uncertain logic a follows:

$$r^0_{[\alpha_{R_0},\beta_{R_0}]} :\equiv \\ hasFlatEdge(O) \wedge task(X, tighten(S, flat)) \implies \\ tightenWith(S, O)$$

The robot (through attention control and perceptual semantics analysis in CALyX) can look around the room and determine (within a certain uncertainty interval) whether or not each of the various objects that it sees has a flat edge.

$$hasFlatEdge(Screwdriver)[0.95, 0.95] \\ hasFlatEdge(Knife)[0.9, 0.9] \\ hasFlatEdge(Coin)[0.75, 0.95] \\ hasFlatEdge(Pencil)[0, 0.95]$$

The robot can then apply DS-theoretic logical inference on rules, such as the one above, and infer uncertainties for the $tightenWith(S, O)$ affordance for each of the five objects. Based on this inference, the robot can deduce that knives and coins can be used to tighten screws in the absence of screwdrivers, but pencils cannot.

B. Creative Problem Solving

An affordance-based approach might shed light on insight and creative problem solving scenarios that require an ability to think about a problem from a different angle, [51], or in our case, a different context. Affordance-based creative reasoning approaches are not new and have been attempted by Olteteanu [52]. However, these approaches are limited for the same reasons as others in the affordance literature, in that they cannot account for complex affordances in different contextual circumstances. We believe that an affordance representation of the form presented in this work may assist in modeling both creative and commonsense reasoning processes more effectively. Moreover, when coupled with mental simulation engines, the agent need not physically actualize the affordances in order to see their effects, choosing instead to simulate mentally in a suitable physics engine.

C. Role of Affordances in Sense-making

Our perception of affordances in our environment enable us to not only know what we can do with objects around us, but they also serve to tell a story about our current situation. For example, chairs and tables in a restaurant allow people to sit and eat their food. However, a collection of chairs without any tables in the middle of the restaurant would strike us as a bit unusual. Our need to make sense of the situation drives us to dig deeper and learn more about the reasons why there are no tables. This same need is what allows us to discover problems when there is a mismatch between what we see and what we expect to see. Reasoning about cognitive affordances in a more general way, as outlined in this paper, has the potential to assist in such sense-making, which can be useful for artificial agents navigating in the open world.

VIII. Conclusion

As part of their interview process, many modern technology companies show prospective candidates an object they have never seen before and ask them to describe what they think is the object's function. The purpose of the question is to test the candidate and probe their intellect to identify candidates with strong mental representations of affordance. Clever answers are often rewarded and stand as an example of human creativity. The ultimate goal of our research is to endow robots with the ability to find creative ways to use and manipulate objects and their environment.

In this work, we took the first steps towards our goal and proposed a novel computational framework based on Dempster-Shafer (DS) theory for inferring cognitive affordances. We demonstrated how our framework can handle uncertainties and be extended to include the continuous and dynamic nature of real-world situations. We believe that this, much richer level of affordance representation is needed to allow artificial agents to be adaptable to novel open-world scenarios.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
[2] J. J. Gibson, *The ecological approach to visual perception*, 1979, vol. 39.
[3] M. T. Turvey, "Affordances and Prospective Control: An Outline of the Ontology," *Ecological Psychology*, vol. 4, no. 3, pp. 173–187, 1992.
[4] E. Reed, *Encountering the world: toward an ecological psychology*, 1996, vol. 34, no. 10.
[5] D. Norman, "The Psychology of Everyday Things," in *The Psychology of Everyday Things*, 1988, pp. 1–104.
[6] T. A. Stoffregen, "Affordances as Properties of the Animal-Environment System," *Ecological Psychology*, vol. 15, no. 2, pp. 115–134, 2003.
[7] A. Chemero, "An Outline of a Theory of Affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.
[8] E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk, "To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.
[9] F. Mastrogiovanni, A. Scalmato, A. Sgorbissa, and R. Zaccaria, "Problem Awareness for Skilled Humanoid Robots," *International Journal of Machine Consciousness*, vol. 3, no. 1, pp. 91–114, 2011.
[10] F. Mastrogiovanni and A. Sgorbissa, "A biologically plausible, neural-inspired planning approach which does not solve 'The gourd, the monkey, and the rice' puzzle," *Biologically Inspired Cognitive Architectures*, vol. 2, pp. 77–87, 2012.
[11] A. Scarantino, "Affordances Explained," *Philosophy of Science*, vol. 70, no. 5, pp. 949–961, 2003.
[12] L. W. Barsalou, S. A. Sloman, and S. E. Chaigneau, "The HIPE Theory of Function," in *Representing functional features for language and space: Insights from perception, categorization and development*, 2002, vol. 30322, no. 404, pp. 1–25.
[13] E. Bicici and R. S. Amant, "Reasoning About the Functionality of Tools and Physical Artifacts," *Department of Computer Science, North Carolina State University, Tech. Rep*, vol. 22, pp. 1–34, 2003.
[14] R. C. Schmidt, "Scaffolds for Social Meaning," *Ecological Psychology*, vol. 19, no. 2, pp. 137–151, 2007.
[15] J. Kim and J. Park, "Advanced Grasp Planning for Handover Operation Between Human and Robot: Three Handover Methods in Esteem Etiquettes Using Dual Arms and Hands of Home-Service Robot," *2nd International Conference on Autonomous Robots and Agents*, no. c, pp. 34–39, 2004.

[16] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, V. Micelli, and W. Garage, "Toward Seamless Human – Robot Handovers," *Journal of Human Robot Interaction*, vol. 2, no. 1, pp. 112–132, 2013.

[17] M. Steedman, "Formalizing Affordance," *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*, pp. 834–839, 2002.

[18] D. Abel, G. Barth-Maron, J. MacGlashan, and S. Tellex, "Affordance-Aware Planning," 2015.

[19] L. Montesano, M. Lop, A. Bernardino, and J. Santos-Victor, "Modeling affordances using Bayesian networks," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 4102–4107.

[20] L. Montesano and M. Lopes, "Learning grasping affordances from local visual descriptors," *2009 IEEE 8th International Conference on Development and Learning (ICDL 2009)*, 2009.

[21] E. Ugur, H. Celikkanat, E. Sahin, Y. Nagai, and E. Oztop, "Learning to grasp with parental scaffolding," *2011 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011)*, 2011.

[22] E. Ugur, E. Sahin, and E. Oztop, "Self-discovery of motor primitives and learning grasp affordances," *IEEE International Conference on Intelligent Robots and Systems*, pp. 3260–3267, 2012.

[23] E. Ugur, Y. Nagai, and E. Oztop, "Parental scaffolding as a bootstrapping mechanism for learning grasp affordances and imitation skills," *Proc. of the RAAD 2013 22nd InternationalWorkshop on Robotics in Alpe-Adria-Danube Region*, no. August 2014, pp. 1–19, 2013.

[24] B. Moldovan, P. Moreno, M. Van Otterlo, J. Santos-Victor, and L. De Raedt, "Learning relational affordance models for robots in multi-object manipulation tasks," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4373–4378, 2012.

[25] J. Aleotti, V. Micelli, and S. Caselli, "An Affordance Sensitive System for Robot to Human Object Handover," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 653–666, 2014.

[26] W. P. Chan, M. K. Pan, E. A. Croft, and M. Inaba, "Characterization of handover orientations used by humans for efficient robot to human handovers," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, sep 2015, pp. 1–6.

[27] K. M. Varadarajan and M. Vincze, "Knowledge representation and inference for grasp affordances," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6962 LNCS, pp. 173–182, 2011.

[28] K. Varadarajan, "Topological mapping for robot navigation using affordance features," *2015 6th International Conference on Automation, Robotics and Applications (ICARA). Proceedings*, 2015.

[29] T. Shu, M. S. Ryoo, and S.-C. Zhu, "Learning Social Affordance for Human-Robot Interaction," *International Joint Conference on Artificial Intelligence (IJCAI), 2016*, p. Accepted, apr 2016.

[30] V. Sarathy and M. Scheutz, "Semantic Representation of Objects and Function," in *Proceedings of the 2015 IROS Workshop on Learning Object Affordances*, 2015.

[31] T. Williams, G. Briggs, B. Oosterveld, and M. Scheutz, "Going Beyond Literal Command-Based Instructions : Extending Robotic Natural Language Interaction Capabilities," in *AAAI*, 2015, pp. 1387–1393.

[32] M. Scheutz, P. Schermerhor, J. Kramer, and D. Anderson, "First steps toward natural human-like HRI," *Autonomous Robots*, vol. 22, no. 4, pp. 411–423, 2007.

[33] Y. Tang, C. W. Hang, S. Parsons, and M. Singh, "Towards argumentation with symbolic dempster-shafer evidence," *Frontiers in Artificial Intelligence and Applications*, vol. 245, no. 1, pp. 462–469, 2012.

[34] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[35] L. A. Zadeh, *On the Validity of Dempster's Rule of Combination of Evidence*. Electronics Research Laboratory, University of California, 1979.

[36] R. C. Nunez, R. Dabarera, M. Scheutz, G. Briggs, O. Bueno, K. Premaratne, and M. N. Murthi, "DS-based uncertain implication rules for inference and fusion applications," *Information Fusion (FUSION), 2013 16th International Conference on*, pp. 1934–1941, 2013.

[37] L. Shapira, A. Shamir, and D. Cohen-Or, "Consistent mesh partitioning and skeletonisation using the shape diameter function," *Visual Computer*, vol. 24, no. 4, pp. 249–259, 2008.

[38] A. Ten Pas and R. Platt, "Localizing Handle-Like Grasp Affordances in 3-D Points Clouds Using Taubin Quadric Fitting," in *International Symposium on Experimental Robotics (ISER)*, 2014.

[39] S. Glasauer, M. Huber, P. Basili, A. Knoll, and T. Brandt, "Interacting in time and space: Investigating human-human and human-robot joint action," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2010, pp. 252–257.

[40] S. Shibata, K. Tanaka, and A. Shimizu, "Experimental analysis of handing over," *Robot and Human Communication, 1995. RO-MAN'95 TOKYO, Proceedings., 4th IEEE International Workshop on*, pp. 53–58, 1995.

[41] A. L. Yarbus, *Eye movements and vision*, 1967, vol. 6, no. 4.

[42] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005.

[43] E. Potapova, A. Richtsfeld, M. Zillich, and M. Vincze, "Incremental Attention-driven Object Segmentation," in *14th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2014, pp. 252—-258.

[44] J. Tünnermann and B. Mertsching, "Saliency and Affordance in Artificial Visual Attention," *Proceedings of First Workshop on Affordances: Affordances in Vision for Cognitive Robotics, Robotics Science and Systems*, 2014.

[45] K. L. Roberts and G. W. Humphreys, "Action-related objects influence the distribution of visuospatial attention." *Quarterly journal of experimental psychology (2006)*, vol. 64, no. January 2012, pp. 669–688, 2011.

[46] P. Garrido-Vásquez and A. Schubö, "Modulation of Visual Attention by Object Affordance," *Frontiers in Psychology*, vol. 5, no. February, pp. 1–11, 2014.

[47] R. Cantrell, K. Talamadupula, P. Schermerhorn, J. Benton, S. Kambhampati, and M. Scheutz, "Tell Me When and Why to Do It!: Runtime Planner Model Updates via Natural Language Instruction," in *Proceedings of the 2012 Human-Robot Interaction Conference*, 2012.

[48] A. Boularias, J. A. Bagnell, and A. Stentz, "Learning to Manipulate Unknown Objects in Clutter by Reinforcement," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Learning*, 2015, pp. 1336–1342.

[49] J. Fellrath and R. Ptak, "The role of visual saliency for the allocation of attention: Evidence from spatial neglect and hemianopia," *Neuropsychologia*, vol. 73, pp. 70–81, 2015.

[50] K. Wickramaratna, M. Kubat, K. Premaratne, and T. Wickramarathne, "Rule mining and missing value prediction in the presence of data ambiguities," in *The Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2009.

[51] L. Macchi and M. Bagassi, "The interpretative heuristic in insight problem solving," *Mind & Society*, vol. 13, no. 1, pp. 97–108, 2014.

[52] A.-M. Olteteanu and C. Freksa, "Towards affordance-based solving of object insight problems," in *Proceedings of First Workshop on Affordances: Affordances in Vision for Cognitive Robotics, Robotics Science and Systems*, 2014.

Vasanth Sarathy is a Ph.D. Student in Cognitive and Computer Science in the Department of Computer Science. He earned a B.S. in Electrical Engineering from the University of Arkansas in 2003, an S.M. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 2005, and a J.D. in Law from Boston University School of Law in 2010. His current research is in artificial intelligence and cognitive robotics with a focus on situated perception and computational creativity.

Matthias Scheutz is a Professor in Cognitive and Computer Science in the Department of Computer Science. He earned a Ph.D. in Philosophy from the University of Vienna in 1995 and a Joint Ph.D. in Cognitive Science and Computer Science from Indiana University Bloomington in 1999. He has more than 200 peer-reviewed publications in artificial intelligence, natural language processing, cognitive modeling, robotics, and human-robot interaction. His current research focuses on complex cognitive robots with natural language capabilities.