Beyond Grasping - Perceiving Affordances Across Various Stages of Cognitive Development

Vasanth Sarathy and Matthias Scheutz
Human-Robot Interaction Laboratory, Tufts University
Medford, MA 02155, USA
Email: vasanth.sarathy@tufts.edu, matthias.scheutz@tufts.edu

***Abstract*—The concept of "affordance" has typically represented the relationship between human perceivers and their environment. Affordance perception, representation, and inference are central to commonsense reasoning, tool-use and creative problem-solving in artificial agents. Existing approaches to representing affordances have focused on its physical 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 fail to provide flexibility with which to reason about affordances through various developmental stages, where they are more influenced by changing context, social norms, historical precedence, and uncertainty. We develop a formal rules-based logical representational format coupled with an uncertainty-processing framework to reason about cognitive affordances in a more general manner than shown in the existing literature. Our framework, which is retained through cognitive development, allows agents to make deductive and abductive inferences about functional and social affordances. We demonstrate our approach with an example, and show that an agent can successfully reason through situations that involve a tight interplay between various social and functional norms.**

I. Introduction

Reasoning about the environment and deciding what to do with objects are highly desirable skills that require not only recognizing objects but also perceiving what action possibilities are available (i.e., perceiving affordances). Through exploration, infants begin to acquire affordance perception skills gradually. At first, the infant learns about objects by touching and sucking them. Within a few months, they can account for the object's distance and attempt simple grasping. Affordance learning continues throughout childhood with the toddler learning about various physical affordances offered by toys and objects around them [1], [2]. As they get older, children begin to develop a stronger sense for not only physical affordances, but also more functional aspects of objects, understanding for example, that mugs can not only be grasped, but can also be used for drinking. Through social interaction and adult instruction, children also begin to learn various social rules associated with these objects. For example, they begin to learn that sometimes mugs can be used for drinking, and other times, when appropriate, mugs can be used as pen holders. Deciding how to use objects in this manner, in situations that involve a tight interplay between various functional and social norms, is a highly desirable skill for robots, as well. Unfortunately, although robots are getting more and more proficient at perceiving physical affordances like grasping, they are less skilled at resolving higher-order affordances that involve social factors, changing context and uncertainty.

In this paper, we present a novel approach based on Dempster-Shafer (DS) theory [3] and "uncertain logic" for inferring object affordances through various stages of cognitive development, from infancy through adulthood. As part of this effort, we have developed a language for reasoning about affordances that can represent complicated activities and can account for the dynamic and continuous nature of real-world scenarios. We will demonstrate our approach with an example of a simple task of reasoning about household containers.

II. Background

James Gibson [4] introduced the concept of "affordance" to represent the relationship between the agent and its environment. Since then, many competing theories have emerged aimed at unpacking affordance and assembling useful representational formats to reason about actions that the agent can or must take in its environment [5], [6], [7], [8], [9], [10]. These general theories, were largely philosophical and exhibited limited, if any, formalism or mechanisms for reasoning in robotic systems. A number of these and other theories focused primarily on functional aspects of affordances [11], while a few introduced social considerations into an affordance framework [12], [13].

Work in cognitive and developmental robotics as well as in AI originated from these general theories and diverged in two directions: statistical approaches and ontological approaches. The statistical approaches modified and implemented these general theories in specific domains using statistical formalisms to represent and compute affordances [14], [15], [16], [17], [18]. The ontological approaches focused on developing 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 [19].

Despite these efforts, affordance representation faces many challenges that have not been overcome in the previous work. These 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, these current approaches cannot reason that coffee mugs afford grasping and drinking, while also affording serving as a paperweight or cupholder, or depending on the context, as family heirloom not meant to be used at all. From a developmental standpoint, these approaches also fail to provide a robust representation capable of handling different types of affordances learned through various stages of development.

III. Proposed Affordance Model

One reason it is difficult to represent affordances in this flexible way - accounting for contextual and social aspects - is the underlying complexity associated with actualizing (i.e., perceiving and utilizing) relevant affordances from an extremely large number of potential affordances available in the environment. An object affords a large number of physical actions such as grasping and pushing, functional actions related to its use and other social actions based on its meaning to an agent. Moreover, each of these actions can, in turn, be influenced by contextual factors including goals and intentions, prior knowledge and interpretations, ensemble scene information, mental state, experience and developmental state, and social and moral norms, among others. Existing machine learning and ontological approaches suffer performance degradation when modeling this broader class of affordances due to the need for complicated data structures and inherent modeling limitations.

We propose a novel affordance model and formal representational framework that enables accounting for these contextual aspects explicitly: serving as filters and imposing constraints on the process of perception as well as on the selection of an action. The intuition behind our approach is that not all affordances are relevant in every context, a fact we can then use to reduce the complexity of the search space. The proposed model is rules-based, and represents the relationship between perceptual information in the environment and the possibilities for action in the presence of contextual constraints.

More specifically, the rules (R) are premise-conclusion pairs in which the premises are variables representing perceptual invariants (F) in the environment and certain contextual information (C), and the conclusions are variables representing affordances (A) actualizable by the agent in the situation. Each of the rules and the variables in the premise are assigned confidence intervals, grounded in Dempster-Shafer theory, representing uncertainties in our beliefs about them. Using the DS-theoretic framework coupled with Uncertain logic, we can compute the confidence intervals associated with the affordances in the rule's conclusion.

The proposed representation allows for constraints to be propagated through the inference process, thereby limiting the search space and simplifying the combinatorial complexity of affordance perception. In particular, contextual information act as constraints and filter what rules the agent must consider in a given situation. The rules, themselves, further limit what affordances should be perceived in the current context and what actions the agent should take. Interestingly, the proposed framework also allows the agent to reason abductively and decide where it should focus its attention (i.e., what perceptual invariants to detect in the environment).

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

$\{\Theta_F, m_f\}$: Candidate perceptual features
$\{\Theta_C, m_c\}$: Relevant contextual items
R: Currently applicable rules
$S = \emptyset$
for all $r \in R$ **do**
  $S = S \cup \{(m_f \otimes m_c) \odot m_{r=fc \to a}\}$
end for
$G = group(S)$
$\psi = \emptyset$
for all group $g_a \in G$ **do**
  Fusion: $\psi = \psi \cup \{\bigcap_{j=0}^{|g_a|} g_{a_j}\}$
end for
return ψ

Mathematically, the set of affordance rules (R) take the overall form:

$$r_{[\alpha,\beta]} :\equiv f \wedge c \implies a$$

with $f \in F$, $c \in C$, $a \in A$, $r \in R$, $[\alpha, \beta] \subset [0, 1]$. 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. Similarly, each of the variables f and c also have confidence values associated with them, and are used for inferring affordances as described in more detail below. 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. It has been previously shown that these types of rules are very versatile and that one can employ DS-theoretic modus ponens to make uncertain deductive and abductive inferences [20]. Most critically, these rules allow us to address representational challenges with mere Bayesian models such as inferring $P(A|F, C)$ by way of $P(F|A, C)$, $P(A|C)$, and $P(C)$ when we often have no practical way to obtain the necessary probability distributions for all the potential affordances for an object.

IV. Inferring Affordances

The purpose of 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. As a preliminary matter, in DS-theory, $\Theta = \{\theta_1, ..., \theta_N\}$ represents the sample space of finite mutually exclusive and exhaustive outcomes of an experiment, where each θ_i are elementary events. Each elementary event and combinations thereof have a certain weight or mass m associated with it. Thus, in the proposed model for example, the variable representing the set of perceptual features F includes several sample spaces or perceptual aspects Θ_F (e.g., shape, color, etc.) with elementary events f representing the detection of this aspect, having a mass m_f. Similarly, we can define the sample space, elementary events and masses for the contextual information C as $\{\Theta_C, m_c\}$, and rules R as $\{\Theta_R, m_{fc \to a}\}$. The confidence interval $[\alpha, \beta]$ for each of these variables are computed from the masses of the elementary events using Dempster-Shafer theoretic notions of belief and plausibility[1].

The algorithm takes three parameters: (1) The set of candidate perceptions $\{\Theta_F, m_f\}$ is provided by the low-level vision system, (2) a set of relevant contextual items $\{\Theta_C, m_c\}$ provided by the cognitive system's knowledge base or some other part of the integrated system (e.g., belief inference components) that can provide context information, and (3) a table of cognitive affordance rules R. Here, 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.

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 the confidence interval $[\alpha(a),\ \beta(a)]$ when the system believes that 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 $r_{f \wedge c \to a} \in R$ (line 5), performs a DS-theoretic AND operation and an *Uncertain Modus Ponens* to obtain m_a (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 assignments (line 8), and use Yager's rule of combination (line 11) defined in [22] 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 "ambiguity measure" λ defined in [23] to determine whether an inferred affordance should be realized and acted upon. For example, we could check the ambiguity 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 an ambiguity 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.

V. Example: How to use a Coffee Mug

We will now present an evaluation of our representation and our algorithm, and demonstrate the capabilities facilitated by our approach. We will use an example of reasoning about containers to demonstrate the behavior of our algorithm.

Containers of all types are highly useful objects in our daily lives. We reason about containers constantly, from deciding where to put our toothbrush in the morning, to selecting a mug for coffee and organizing our desk and placing items in drawers and boxes. The idea of a container is a subject of recent work in commonsense reasoning [24] and important in AI and robotics research. Existing approaches have focused on geometric and physical aspects of the properties of a container, which while relevant to infant affordance perception, do not translate to later stages of development when functional, social and contextual factors play an increasingly important role. Specifically, these approaches do not provide flexibility with which to reason about action choices, where the social norms and rules may change and other contextual factors may influence the use of a certain object as a container.

Consider this example: If asked to stow away a pen in a container, adults would consider a dry mug on our office table as a candidate. Now, if the same dry mug were in a kitchen along with other mugs, adults might not consider putting the pen into a coffee mug in the kitchen. The same object, i.e., a coffee mug has different affordances depending on context, and these affordances sometimes have more to do with social tendencies than physical constraints. While an infant may not be able to perceive these subtleties, an adult can and is expected to be able to do so. For example, the infant may not be able to distinguish between contexts in this manner or it may not have learned about the social etiquette associated with kitchens and offices, or it may not even have learned the containment affordance. We will demonstrate that the representation and algorithm described above allows us to reason about exactly these types of subtleties.

VI. Containment Formalized

A. 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. These components also generate point clouds and 3D meshes, and 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. We will use $m_{f_{i,j}}$ to denote the candidate mass values of the percepts, where $i \in \{1 \ldots N\}$ and $j \in \{1 \ldots M\}$.

[1]The lower bound captures the total support that can be committed to a event without also committing to the negation of the event. The upper bound captures the total belief that does not contradict the event. More details about the confidence intervals and other aspects of DS-theory are presented in [21].

For the purposes of our example, we will represent the agent's visual perception of our domain objects with five binary visual aspects, each aspect with a percept and its negation.

Aspect (Θ_{F_i})	Percept ($f_{i,j}$)	Mass ($m_{f_{i,j}}$)
Θ_{F_1}	$isCylinder(O_1)$	$m_{f_{1,1}}$
Θ_{F_2}	$hasOpening(O_1)$	$m_{f_{2,1}}$
Θ_{F_3}	$hasHandle(O_1)$	$m_{f_{3,1}}$
Θ_{F_4}	$isLiquid(O_2)$	$m_{f_{4,1}}$
Θ_{F_5}	$isSmaller(O_2, O_1)$	$m_{f_{5,1}}$

$isCylinder(O_1)$ represents the knowledge that the object O_1 is cylindrical in shape. $hasOpening(O_1)$ represents the knowledge that the object O_1 has an opening on a surface that allows for the placement of other objects. $hasHandle(O_1)$ represents the knowledge that the object O_1 has a handle with which to grasp it. $isLiquid(O_2)$ represents the knowledge that the object O_2 is a liquid, and $isSmaller(O_2, O_1)$ represents the knowledge that the object O_2 is smaller in volume than object O_1, thus allowing O_2 to fit inside O_1.

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 are an infinite 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 the vision system to 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.

B. Relevant Contextual Items, C

Knowledge of the agent's current context is provided by certain high-level processing components such as the agent's belief, planning and goal management system. The context is representative of the agent's beliefs, goals, desires, and intentions, along with certain other abstract constructs in the agent's narrative 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. 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 one binary contextual aspects, which includes a contextual value (context) and its negation:

Aspect (Θ_{C_i})	Context($c_{i,j}$)	Mass ($m_{c_{i,j}}$)
Θ_{C_1}	$domain(L)$	$m_{c_{1,1}}$

$domain(L)$ represents the agent's current domain, L. For example, $domain(kitchen)$ represents the knowledge that the agent is currently in the domain of working in the kitchen.

C. Cognitive Affordances, A

The next part of our representational framework are the cognitive affordances A computed by several rules. 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). We will use $m_{a_{i,j}}$ to denote the candidate mass values of the affordances, 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 one affordance aspect, with an affordance and its negation.

Aspect (Θ_{A_i})	Affordance ($a_{i,j}$)	Mass ($m_{a_{i,j}}$)
Θ_{A_1}	$containWith(X, O_2, O_1)$	$m_{a_{1,1}}$

The $containWith(X, O_2, O_1)$ represents the property of object O_1 which allows for an agent to contain object O_2 within O_1 by performing an action X.

Now, we recognize that these affordance are always available to the agent: the agent can place objects inside other objects at any time. Our affordance representation does not deny that latent affordances may exist in objects, but merely attaches uncertainties to their potential for actualization. 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 infinitely many more affordances for containers, 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 infinite, 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 an infinite number of affordances, we will contend that only a finite subset of them are relevant in any given set of contexts, applicable at a particular moment in time.

D. 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). We will use $m_{r_{i,j}}$ to denote the candidate mass values of the rules, 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 four rule aspects (representing four rules), each aspect with a rule and its negation. Generally, the rules are of the form:

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

Below, we show each of the 4 rules for this example, presenting the uncertainty intervals for each of the rules. For ease of reading, we have omitted the index $j = 1$.

$$r^1_{[\alpha_1,\beta_1]} := hasOpening(O_1) \wedge isSmaller(O_2, O_1) \wedge domain(kitchen) \implies containWith(X, O_2, O_1)$$
$$r^2_{[\alpha_2,\beta_2]} := hasOpening(O_1) \wedge isSmaller(O_2, O_1) \wedge domain(office) \implies containWith(X, O_2, O_1)$$
$$r^3_{[\alpha_3,\beta_3]} := isCylinder(O_1) \wedge hasHandle(O_1) \wedge hasOpening(O_1) \wedge isLiquid(O_2) \wedge domain(kitchen) \implies containWith(X, O_2, O_1)$$
$$r^4_{[\alpha_4,\beta_4]} := isCylinder(O_1) \wedge hasOpening(O_1) \wedge domain(office) \implies containWith(X, O_2, O_1)$$

Rules r^1 and r^2 relate to the agent's general commonsense understanding of the physical properties of objects as containers in any domain. Objects that have openings can, simplistically, serve as containers.

Rules r^3 and r^4 relate to a more domain specific set of social rules. Rule r^3 essentially represents our intuitive social expectation that cylindrical objects with handles that have an opening can serve as a container for liquids. Rule r^4 represents our intuitive social expectation that cylindrical containers in an office can serve as containers for objects, not necessarily liquids (i.e., pen-holders).

When modeling an infant or a toddler, we may choose to include rules r^1 and r^2, but exclude rules r^3 and r^4. An infant or a toddler is unlikely to have learned rules r^3 and r^4 as they involve more subtle social norms that they might subscribe to at a later date through trial and error, for e.g., by receiving reprimand for pouring water into the pen holder.

VII. Experiment

We consider the example of a robot tasked with finding a container for (1) stowing away a pen, and (2) for storing water. The environment contains real objects consisting of a Mug, Pen and Water, across two domains, namely kitchen and office. The robot must decide if the Mug is a suitable container for $Water$ or the Pen in each context of a $Kitchen$ and an $Office$. The robot can only perform one action, $PlaceIn$, which allows it to place one object inside another.

We evaluated our set of rules across both contexts (office and kitchen) for all values of DS-theoretic masses ranging from 0 to 1 (step-size of 0.005). For each combination of (mass $m_{c_{1,1}}$) representing whether the robot is in the kitchen or in the office or it is unsure if it is in the kitchen or office, we computed the values of the ambiguity measure λ of the $containWith()$ affordance and plotted our results in Fig. 1. Each of the four plots shows the value for an ambiguity measure, λ, for the affordance of $containWith()$ using a Mug. Values close to 1 represent high certainties and values close to 0 represent maximum uncertainty.

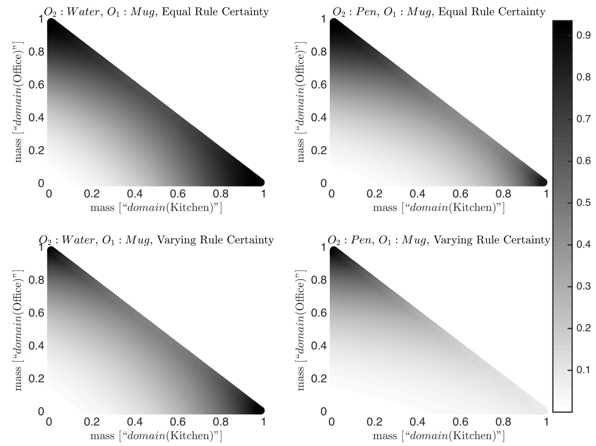

Fig. 1. Plot showing that coffee mugs are more likely to be perceived as pen-holders in an Office than in a Kitchen (bottom-right). But, the same mugs are perceived as suitable water containers in both domains.

The first column of plots represents the cases when the robot is reasoning whether a mug is a suitable container for water. The second column of plots represents the cases when the robot is reasoning whether a mug is a suitable holder for a pen. The top row of plots corresponds to the cases where each of the four affordance rules r^1 - r^4 have equal and maximum certainty. The bottom row of plots corresponds to the cases where rules r^1 and r^2, which are the more general of the rules, have lower certainties than rules r^3 and r^4, which are the more specific of the rules.

Our results demonstrate that mugs are generally good containers for water, in both the kitchen and office domains. However, for storing pens, mugs are deemed to be better-suited in an office than in a kitchen (bottom-right). Our affordance framework allows the robot to reason about social expectation together with functionality, while simultaneously accounting for uncertainty in knowledge and information.

VIII. Discussion and Conclusion

In this paper, we developed a novel logic-based framework and algorithm using Dempster-Shafer (DS) theory for inferring object affordances. We demonstrated how our framework can handle changing contexts, uncertainties and be extended to include the 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. Our framework allows for a body-independent and agent-capability-independent language with which to reason about action possibilities. Perceptions of these types of higher-level affordances are important both for agents that have to use novel or unknown objects as tools to perform tasks, and also for agents that operate in social spaces where they must recognize the social affordance of the situation (e.g., where to stand, where to move).

One advantage of such an approach, developmentally, is that the same framework - representation and algorithm - can be used across various developmental stages. During infancy, a child learns simple commonsense and naive physics rules about objects. They learn that these objects have simple affordances like graspable, liftable, turnable and reachable. As they develop into toddlers, the child begins to learn affordance rules for more functional aspects. That is, they learn that sharp objects can be used to cut with and that objects with openings can be used to drink water. Child development of affordance perception develops as they grow [25]. Finally as children and later as adults, they learn social etiquettes, and learn to appreciate the role of context in affordance reasoning. Children learn affordances through exploratory actions and through instruction (e.g., learning how to hold a spoon) [26]. They learn that mugs have the affordance of being used as a paperweight or a decorative piece depending on context. Children perceive more and more complex invariants as they grow and are tuned into more social aspects [27]. In each of these stages, while the rules are different, the representation and inference process remains the same. Our approach suggests that there is no additional cognitive machinery needed to support the higher-order reasoning about functionality and social norms that appears in later stages of development.

More generally, the proposed framework displays consistency with ecological and socio-cultural approaches to perceptual learning and development: that perceptual development is a process of increasing one's ability to detect and select relevant, increasingly differentiated and specific perceptual information that is already available in the environment [28]. Similarly, in the proposed framework, perceptual learning involves updating and refining the set of learned affordance rules, and selecting the appropriate subset of rules over which to reason in a particular situation. As the agent develops, the rule's premises evolve from murky undifferentiated percepts F and sociocultural contexts C to more specific ones. Affordances A also evolve as new action capabilities are developed by the agent. Uncertainty measures account for other developmental variables such as perceptual sensitivity.

Our framework has shown some potential, but we will still need to address various challenges. For example, we will look into a framework for learning various affordance rules from observation, exploration, and demonstration. In doing so, we will need to incorporate action formalisms to allow a more dynamic reasoning process. Towards this end, we are extending our logic to reason about the effect of acting on affordances. Finally, we will be integrating our framework into a robotic cognitive architecture and grounding our representation to work with embedded cognitive systems.

IX. Acknowledgements

This research was funded in part by grant #N00014-14-1-0149 from the US Office of Naval Research.

References

[1] M. Olmos, J. A. Carranza, and M. Ato, "Force-related Information and Exploratory Behavior in Infancy," *Infant Behavior and Development*, vol. 23, no. 3-4, pp. 407–419, 2000.

[2] F. Guerin, N. Kruger, and D. Kraft, "A Survey of the Ontogeny of Tool Use: From sensorimotor experience to planning," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 1, pp. 18–45, 2013.

[3] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[4] J. J. Gibson, *The Ecological Approach to Visual Perception*, 1979, vol. 39.

[5] M. T. Turvey, "Affordances and Prospective Control: An Outline of the Ontology," *Ecological Psychology*, vol. 4, no. 3, pp. 173–187, 1992.

[6] E. Reed, *Encountering the world: toward an ecological psychology*, 1996, vol. 34, no. 10.

[7] D. Norman, "The Psychology of Everyday Things," in *The Psychology of Everyday Things*, 1988, pp. 1–104.

[8] A. Chemero, "An Outline of a Theory of Affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[9] 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.

[10] A. Scarantino, "Affordances Explained," *Philosophy of Science*, vol. 70, no. 5, pp. 949–961, 2003.

[11] 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.

[12] R. C. Schmidt, "Scaffolds for Social Meaning," *Ecological Psychology*, vol. 19, no. 2, pp. 137–151, 2007.

[13] 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.

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

[15] 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.

[16] 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.

[17] W. P. Chan, M. K. X. J. Pan, E. A. Croft, and M. Inaba, "Characterization of Handover Orientations used by Humans for Efficient Robot to Human Handovers," *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[18] E. Ugur, Y. Nagai, E. Sahin, and E. Oztop, "Staged Development of Robot Skills: Behavior formation, affordance learning and imitation with motionese," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 2, pp. 119–139, 2015.

[19] K. Varadarajan, "Topological Mapping for Robot Navigation using Affordance Features," *2015 6th International Conference on Automation, Robotics and Applications (ICARA). Proceedings*, 2015.

[20] 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.

[21] V. Sarathy and M. Scheutz, "Cognitive Affordance Representations in Uncertain Logic," in *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2016.

[22] R. R. Yager, "On the Dempster-Shafer Framework and New Combination Rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[23] 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.

[24] E. Davis and G. Marcus, "Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence," *Communications of the ACM*, vol. 58, no. 9, pp. 92–103, 2015.

[25] E. J. Gibson and M. A. Schumuckler, "Going Somewhere: An Ecological and Experimental Approach to Development of Mobility," 1989.

[26] A. Costall, "Socializing Affordances," *Theory & Psychology*, vol. 5, no. 4, pp. 467–481, 1995.

[27] M. Kyttä, "Affordances of Children's Environments in the Context of Cities, Small Towns, Suburbs and Rural Villages in Finland and Belarus," *Journal of Environmental Psychology*, vol. 22, no. 1-2, pp. 109–123, 2002.

[28] H. L. Pick, "Eleanor J. Gibson: Learning to Perceive and Perceiving to Learn." *Developmental Psychology*, vol. 28, no. 5, pp. 787–794, 1992.