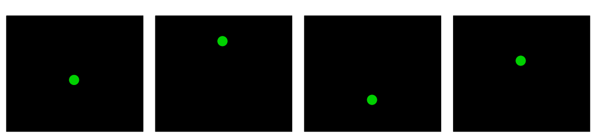
HCC: Small: Computing Systems with Intrinsic Semantics for Enhanced Human-Computer Communication

1 Introduction

Goal: Our goal is to answer two scientific questions: (1) How can semantics become intrinsic to (or arise intrinsically from) computing systems, rather than such systems depending on human users to provide meaning to their internal computations and resulting outputs? (2) What impact does computing artifacts with intrinsic semantics have on human-computer interaction.

Background and Gap in Our Knowledge: With the exponential growth in the amount of information in the society and ultra-fast global networks to deliver such content, information quality (as opposed to quantity or speed of delivery) is becoming an increasing necessity (cf. "data deluge"; Gershon and Miller 1993). A large part of the quality issue arises due to the lack of intrinsic meaning to the information exchanged by computing systems. Semantic computing (Sheu et al. 2010) and semantic web (Berners-Lee et al. 2001; Shadbolt et al. 2006) are beginning to answer these issues, but even in these latest efforts, meaning is not intrinsic to the computing systems (see Section 2 for details); meaning only resides in the brain of the observer. The problem here is that of *grounding*. According to Harnad (1990) who coined the term: "How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?". (Also see the Chinese Room Argument by Searle that prompted the subsequent discussions on grounding [Searle 1980].) Existing research in cognitive science addresses this problem through perception- and/or action-based (in other words, "embodied") approaches (Barsalou et al. 2003; Cangelosi and Riga 2006; Glenberg and Kaschak 2002; Varela et al. 1991; Ziemke 1999) (see Section 2 for more details). However, in these existing works, (1) the key aspect of *intrinsicness* (as opposed to extrinsically provided semantic content) has been overlooked, and/or (2) an *operational recipe* (concrete formulation and algorithms rather than abstract theoretical propositions) for grounding has not been provided. Finally, most prominently, (3) *human attitude* toward grounded vs. ungrounded computing systems has not been evaluated empirically.

Approach: This project draws its main inspiration from the biological brain which is the only known grounded system with intrinsic semantics. Figure 1 gives a glimpse of the problem faced by the brain (e.g., think of the green lights as neuronal spikes). First, the problem of grounding is framed from the point of view of the brain itself, leading to a novel set of assumptions and constraints (e.g., brain's operation is solely based on neuronal spikes that are internal to the brain). Next, an operational recipe (in this case a reinforcement learning framework) is proposed to solve the grounding problem within the framework in the first step. Finally, the resulting system, which



(a) (b) (c) (d)

Figure 1: **A Clueless Situation.** What is the meaning of the green "lights" in these figures? We can tell certain things about these lights: (1) they are round in shape, (2) they have a green appearance, and (3) they seem to be occupying four distinct locations. Beyond that, nothing can be inferred about the meaning of these lights. In other words, what we have here is a genuine grounding problem. Figure 2 will show what these lights mean, and Figure 3 will show how the meaning can actually be inferred.

is expected to exhibit simple behavior indicative of grounding, will be used to evaluate human attitude toward grounded vs. ungrounded computing artifacts and to measure the enhancement in the efficiency of human-computer communication in systems with intrinsic semantics.

Objectives: The objectives of this project are as follows:

1. Derive an operational recipe (exact framework and algorithms) for grounding of simple geometric concepts based on an internal, intrinsic perspective of the biological brain.
2. Extend the operational recipe for grounding to include compositions of geometric concepts.
3. Evaluate human attitude toward grounded vs. ungrounded computing artifacts.
4. Measure the enhancement in efficiency of human-computer communication in systems with intrinsic semantics.

Scientific Merit: Semantic grounding based on embodied principles is not new (Cangelosi and Riga 2006; Varela et al. 1991; Ziemke 1999). The main novelty of this project lies in a concrete operationalization of such principles, based on inspiration from the biological brain. Furthermore, studying the effects of grounded vs. ungrounded computing systems in human-computer interaction can generate valuable new insights on the role and importance of computing systems with intrinsic semantics.

Broader Impacts: Two graduate and three undergraduate students (through REU supplement) will be trained. Data and code will be shared. The research will lead to a natural, more accessibile computer interfaces. Please refer to the project summary and Section 6 for details.

2 Background

In this section, we will briefly survey works that provide motivation to this project. We will see that all pre-existing work have one or more aspects missing from the list of important ingredients of grounding we presented in Section 1: (1) intrinsicness, (2) operational recipe, and/or (3) human-computer interaction dimension.

2.1 Semantic Grounding in Cognitive Science

The initial interest in grounding was ignited by John Searle's provocative Chinese Room Argument (Searle 1980), motivated by the Turing Test Turing (1950), where he criticized the computational approach to cognition and language. Searle argued that any apparently intelligent artifact (e.g., the Chinese Room) that is based on purely symbolic manipulations cannot have true understanding. A direct response to this led to the coining of the term "symbol grounding" by Stevan Harnad (Harnad 1990), where he argued that symbols can be "grounded" on the sensorimotor capabilities of the intelligent artifact, thus taking on meaning. Someone trying to understand Chinese by studying a Chinese-Chinese dictionary was given as an example of an ungrounded system, where the entire operation is based on the exchange of arbitrary symbols in the system.

The theoretical discussions on grounding led to experimental works in cognitive psychology. Naturally, the approaches taken were generally sensorimotor and embodied, although in cognitive theories, emphasis is mostly on the sensory side (as pointed out by Ziemke 1999). A good example of the sensory- (or perceptually) oriented approach can be found in Barsalou et al. (2003). Although Barsalou et al. mention sensorimotor grounding, their approach is perceptual, focusing on multiple sensory modalities that define meaning. Action in their case is simply to reconstruct the perceptual image based on the encoded symbols. A direct contrast to this is the work by Glenberg and Kaschak (2002) where action is identified as the primary source of grounding, not perception. Glenberg et al. demonstrated the “action-sentence compatibility effect” where they showed that a verbally made judgment response that is against the direction of action of the human subject results in hampering of the judgment. The results and analysis suggested that meaning of sentences are grounded in human action.

2.2 Semantic Grounding in Robotics

Robotics is a fertile ground for grounding research since grounding is fundamentally sensorimotor in its nature (Ziemke 1999), and robots provide just that: sensorimotor capability in physical form. In this section, we will discuss more broadly to include embodied cognition approaches.

Important theoretical works in this area include that of Varela et al. (1991), Beer (2000), and Brooks (1991) (note that Brooks' work is also quite empirical, as well as being theoretical) although some of these do not explicitly discuss grounding. Varela et al. (1991) advanced the enactive approach, where action, embodiment and agent-environment interaction was emphasized. Interestingly, the authors of this volume advocated the “groundlessness”, in the sense that the structural coupling between cognition and the world are constantly in flux so that there is no stable foundation to ground anything on. Beer (2000) presented a dynamical system view of cognition, which is closely linked to the enactive approach. Brooks (1991) proposed behavioral robotics as an alternative to then dominant cognitive approaches. However, the perspective of grounding is missing in both Beer and Brooks' work. One exception is the work of Philipona et al. (2003), where the problem of grounding is explicitly defined and an operational recipe is given (also see O'Regan and Noë [2001] for a theoretical foundation of this work). In this work, a sensorimotor agent was supposed to infer the dimensionality of the external space solely based on encoded internal sensor readings. This is by far the closest to our proposed framework, but the work, being more theoretical in nature, does not address human-computer interaction issues.

On the empirical side, there are also many relevant works but we will review only a small number of representative approaches. Pierce and Kuipers (1997) was an early work that recognized the issue of building from ground-up sensory and motor representations. Concrete learning algorithms were used to organize unorganized sensory and motor signals into coherent units. However, in their work, the link between the two that is crucial for grounding (as we will see later) was missing. Work from the same research group include more recent work that addresses grounding more explicitly (Kuipers et al. 2006; Modayil and Kuipers 2007). Another work in this area that directly deals with the issue of grounding is that of Cangelosi and Riga (2006). In this work, a backpropagation neural network is use in a sensorimotor grounding and grounding transfer task involving simulated robots.

2.3 Semantic Web and Semantic Computing

Semantic web (Berners-Lee et al. 2001; Shadbolt et al. 2006) aims to make the web of knowledge more organized and machine readable, by providing a standardized structure to share semantic content. Extensive Markup Language (XML) and various forms of web ontologies were put into use to address this semantic content issue. The aim of semantic web from the outset was to allow machines to more efficiently represent and process semantic content, but these systems are not grounded in the true sense of the word. The final call for meaning still rest on the shoulders of humans.

Semantic computing, a more recent entry into grounding-related research, is similar in spirit

Table 1: **Comparison of Representative Approaches Toward Semantics.** X: no, △: somewhat, O: yes. Related approaches are group together (see the Section number column that indicates where these citations appear).

Citation	**Proposal Section**	**Intrinsicness**	**Operational Recipe for Grounding**	**Linking with Human-Computer Interaction**
Barsalou et al. (2003)	2.1	X	△	X
Glenberg and Kaschak (2002)	2.1	X	X	△
Ziemke (1999)	2.1	O	X	X
Philipona et al. (2003)	2.2	O	O	X
Brooks (1991)	2.2	X	O	X
Beer (2000)	2.2	△	△	X
Kuipers et al. (2006)	2.2	△	O	X
Modayil and Kuipers (2007)	2.2	△	O	X
Cangelosi and Riga (2006)	2.2	O	O	X
Berners-Lee et al. (2001)	2.3	X	X	△
Shadbolt et al. (2006)	2.3	X	X	△
Sheu et al. (2010)	2.3	X	△	△
Traum (1999)	2.4	X	△	O
Proposed Approach	4.x	**O**	**O**	**O**

with semantic web (in many cases leveraging on semantic web), but it extends the scope quite a bit. The Institute for Semantic Computing, an organization set up to steer the field, defines semantic computing as "a field that addresses the derivation and matching of the semantics of computational content and that of naturally expressed user intentions to help retrieve, manage, manipulate or even create the content, where 'content' may be anything including video, audio, text, process, service, hardware, network, community, etc." (Sheu et al. 2010). This area is quite broad, with examples ranging from content-based retrieval of image and video retrieval (Yan et al. 2003) to ontology mining for the web (Tao et al. 2008). See Sheu et al. (2010) for a broader survey of the field. Semantic computing suffers from the same limitation as semantic web: semantics not being intrinsic to the system.

2.4 Semantic Grounding in Human-Computer Interaction

Not much work exists on grounding in the human-computer interaction area. A rare exception to this is the work by Traum (1999). (This is perhaps the only paper that comes up after extensive literature search for works that talk about both grounding and human-computer interaction.) Traum discusses computational models of grounding in collaborative systems. The focus is not on semantic grounding in machines, but rather on how collaborating humans can find a *common ground* (or *mutual belief*, *shared conception*). (So, in a sense, it is not about exactly the same type of grounding that we have talked about so far.) Since computers mediate such collaborative work, the ideas presented in Traum's paper have direct relevance to human-computer interaction. Follow-up work citing this paper are more or less in speech and communication theory, e.g., Jokinen (2009), due to the emphasis on shared grounding.

2.5 Summary

In summary, there is a large body of work dealing with semantic grounding but they all lack in one or more aspects of grounding we wish to address in this project: (1) intrinsicness, (2) operational recipe, and (3) relevance to human-computer interaction. Table 1 in the next page gives an overall comparison of representative approaches surveyed in this section, along with our proposed approach shown at the very bottom.

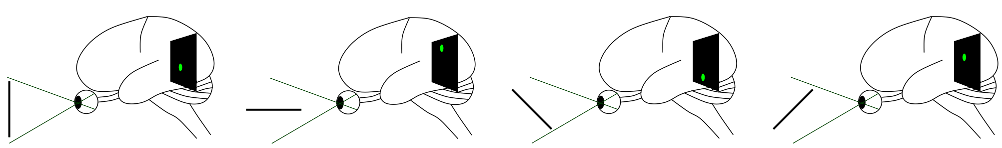
3 Prior Work

3.1 Setting the Framework: Going Inside the Brain

We have seen in Figure 1 a strange scene that is hard to comprehend: Green lights on a black background. However, once we show the full context, as seen in Figure 2, then the meaning of the green lights become evident. The figure reveals that the green lights represent neuronal spikes in the visual cortex of a brain, responding to four different orientations. We, as an observer of this whole system from the *outside*, can easily derive the meaning of the spikes (green lights). However, as we have seen in Figure 1, it seems impossible to comprehend the meaning of the spikes just by looking at them. This is the immediate consequence of "going inside the brain": We lose *direct access* to the outside world.

However, this is the *modus operandi* for the biological brain and the brain has to deal with it. If we consider the fact that the brain operates solely on spikes exchanged internally, it seems impossible that any meaning can be derived from neuronal spikes at any level except for those connected to the immediate sensory surfaces (retina, skin, etc.).

For example, primary visual cortical (V1) neurons respond to visual stimulus features such as orientation, spatial frequency, and phase (Miikkulainen et al. 2005). However, if the input is not given together with the spikes, we cannot in principle decode the spikes. Since secondary visual cortical (V2) neurons receive spikes from V1 but they do not have direct access to the input that triggered spikes in V1, we can say that V2 (or any subsequent area in the visual pathway) has no means to infer the meaning of V1 spikes, just like we are unable to make anything out of the green lights shown in Figure 1.



(a) (b) (c) (d)

Figure 2: **The Clueless Situation, Explained.** The green lights in Figure 1 are revealed as neuronal spikes in the visual cortex, in response to four different orientations.

This raises the question of how the (mammalian) brain was able to exhibit visually guided behavior flawlessly for hundreds of millions of years prior to the famous experimental discovery of V1 neuronal responses to orientated lines by Hubel and Wiesel (1959)? There must be some mechanism in the biological brain to figure out *what to do* when the spikes activate in the visual cortex.

3.2 Solving the Grounding Problem Through Action

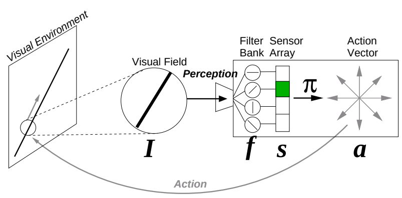
In the previous section, we have observed a strange problem where it appeared that the brain, in principle, cannot infer the meaning of spikes within itself. However, this is not true as it turns out, and the reason for the false conclusion above is due to the fact that more than half of the brain (i.e., the motor areas of the brain) has been disregarded during the discussion.

For the purpose of discussion (and further development later), we present a simple sensorimotor agent, shown in Figure 3(a). The agent has a limited visual field where part of the input (I) from the environment is projected. A set of orientation-tuned units f (sensory primitives) receive that input and transform it to generate a pattern of activity in the sensory array s (green marks active). In the example shown here, the 45° unit is turned on by the input. Based on the sensory array pattern, a mapping π to motor action vector a is determined, resulting in the movement of the visual field in that direction, and then a new input is projected to the agent. Note that the agent is assumed to be aware of only its internal sensory state s, thus it has no knowledge of I, nor that of f. How can this agent attach meaning to such a sensory state? One possible way is to utilize its motor capability, but how?

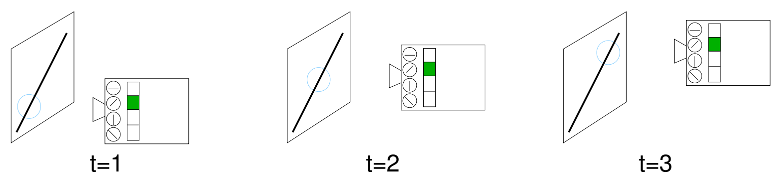
Figure 3(b) shows a strange event where the agent's internal sensory state does not change over time despite the agent's gaze movement along the diagonal direction. The important point here is that the diagonal property of the specific motion exactly corresponds to the environmental property conveyed by the orientation-tuned units. Although this seems trivial, it has profound importance:

This example tells us that even without a single glimpse outside of the box, the agent can internally infer the meaning of its encoded internal representations that are triggered by external stimulus properties.

In terms of Figure 1, if we had a joystick and jiggled around the stick trying to find what directional motion would make a specific green light to be kept on, we can figure out the exact correspondences shown in Figure 2. Thus, grounding *is possible* in this case, but only when action is allowed. This observation not only shows the importance of action for grounding, *but it also prescribes an exact (and very simple) objective to follow* that can lead to grounding. This latter part is the key contribution of this work. Simple, yes, but powerful and novel.



(a) Sensorimotor agent



(b) Internal state invariance shown over time (t=1 to t=3)

Figure 3: **A Sensorimotor Agent.** (a) A sensorimotor agent with visual input I, sensory primitives (orientation detectors) f, sensory state s, and action vectors a is shown. π is the policy, mapping sensory state to action. (b) Invariance in the internal state activity over time is shown, in which case the property of the behavior (diagonal gaze movement) reveals the encoded orientation information.

3.3 PI's Prior Publications on Grounding

In several prior publications, we explored the ideas discussed right above. Choe and Bhamidipati (2004) was our first paper on this topic that used an ad hoc algorithm to learn to maintain internal state invariance. Subsequent works used a more rigorous approach, using Q Learning (Choe and Smith 2006) and SARSA (Choe et al. 2007). In Choe et al. (2008), we presented pilot results on how the sensory filters (receptive fields) can also be learned, as well as the action policy. Other works that can be relevant to the proposed project is Mann and Choe (2011), which explored transfer learning in a reinforcement learning framework. These publications appeared in:

- top conference venues (two AAAI papers and one ICDL paper) (Choe and Smith 2006; Choe et al. 2008; Mann and Choe 2011), one workshop (Yang and Choe 2007), and in
- one journal (*International Journal on Humanoid Robotics*). (Choe et al. 2007).

Results from these prior works will be featured throughout the Research Plan (Section 4).

3.4 Results from Prior NSF Support

The PI was funded by one NSF grant (CRCNS #0905041; Choe, PI; Now expired [9/2009-12/2011]) on a different line of research. The topic of the project was high-resolution serial sectioning imaging of the mouse brain and neuroinformatics infrastructure for data dissemination. Publications resulting from this project are three journal papers (Choe et al. 2011b; Chung et al. 2011; Mayerich et al. 2011b) and six full conference papers (not abstracts) (Choe et al. 2011a; Kwon et al. 2011; Mayerich et al. 2011a; Yang and Choe 2011a,b) over the past two years. A direct outcome of the CRCNS grant is the web-based KESM brain atlas, now serving three whole mouse brain data sets (`http://kesm.org`). The PI also received REU supplements, training two undergraduate students, and one publication resulted from this (Mayerich et al. 2011a) (co-author Aaron Panchal is the student).

4 Research Plan

The research plan is organized into four subsections that correspond directly with the four objectives in the introduction (Section 1). The first two sections are computational in nature, while the two remaining sections are human-oriented.

4.1 Grounding of Simple Geometric Concepts

The task here is similar to the one shown in Figure 1. An agent shown in Figure 3(a) will be adopted as a concrete framework. The task of the agent is clear: Given the sensory state s, figure out the encoded property by mobilizing motor primitives (the action vectors).

Input preparation and response generation: Consider the agent described in Figure 3(a). First, we will preprocess the raw input I_{R} by convolving it with a difference-of-Gaussian (DoG) filter (Rodieck 1965). A small area in the resulting input I_{D} will be sampled (the input is 640×480 and the sampling area is 31×31), producing I.

In order to determine the sensory state, we will use oriented Gabor filters. (Gabor filters are known to resemble the receptive fields in the primary visual cortex of mammals Daugman [1980].) Using the Gabor filters, we calculate the filter response. The filter response is a column vector $\mathbf{r}$. Finally, based on the filter response $\mathbf{r}$, a scalar value representing the sensory state s is calculated as $s = \arg\max_{i=1..n} r_i$, where each value of s corresponds to a unique orientation $\theta = \lfloor (i-1)\pi/n \rfloor$ for $i = 1..n$. The set of all possible s values constitutes the set of sensory states S. For each orientation, there are two matching directions of motion. For example, for θ of 0°, the two directions are 0°and 180°. Thus, the set of actions A has $2n$ movement-direction (or action) vectors as members.

Learning algorithm: Given a particular sensory state $s(t)$ at time t, taking an action $a(t)$ takes the agent into sensory state $s(t+1)$. The state transition depends on the particular edge feature in the visual scene, thus it is probabilistic. The reward $\rho(t+1)$ is simply the degree of invariance achieved across states $s(t)$ and $s(t+1)$.

Note that "invariance" here simply means that the state does not change over time (Figure 3(b)). It is not about "perceptual invariances" which are abstract properties that do not vary across different viewing conditions, e.g., translation/rotation/scaling invariance in vision, or those relating to Gibsonian affordances (Gibson 1979).

One way to measure the *degree* of invariance $\rho(t+1)$ is to calculate the dot-product across the successive sensory filter responses: $\rho(t+1) = \mathbf{r}(t)^{\mathrm{T}}\, \mathbf{r}(t+1)$, where $\mathbf{r}(t)$ is the filter response vector at time step t, and $\mathbf{x}^{\mathrm{T}}$ the transpose of $\mathbf{x}$. With this formula, when the previous filter response is the same as the current, the reward becomes maximized ($\rho(t) = 1$) and in the opposite case minimized ($\rho(t) = -1$). The benefit of using the vector $\mathbf{r}$ instead of directly comparing the scalar values s is that a graded measure of invariance can be obtained instead of a hard *invariant* or *not-invariant* judgment.

The task of the agent is to learn a state-to-action mapping so that it maximizes the reward $\rho(t)$ at time t. However, because of the probabilistic nature of the state transition (which heavily depends on the edge features in the input image), a deterministic state-to-action mapping is not viable. Given the current state $s(t)$, we can think of the conditional probability $P(a(t)|s(t))$ so that if we choose action $a(t)$ with this probability, the probability that the next state $s(t+1)$ being the same as $s(t)$ is maximized. Given an estimate of $P(a(t)|s(t))$, which we will call the *reward probability function* $R(s(t), a(t))$, we let the agent execute the following policy π at each time step t:

1. Given the current state $s(t) \in S$, randomly pick action $a(t) \in A$.
2. Perform action $a(t)$ with probability $R(s(t), a(t))$.
3. Repeat steps 1 to 3 until exactly one action is performed.

To reflect the fact that eye movements follow a fairly straight to slightly curved trajectories between targets (Doyle and Walker 2002), the policy above will be augmented with a momentum mechanism where, with a 30% chance, the action from the previous time step $t-1$ will be repeated, bypassing the steps above. In practice, for step 2 above, the action will be performed if a random draw from $[0..1]$ is less than $cR(s(t), a(t))$, where the parameter c controls the strictness of this check. When the location of gaze reached the image boundary in I_{D}, the movement will be wrapped around and continued on the opposite edge of the input.

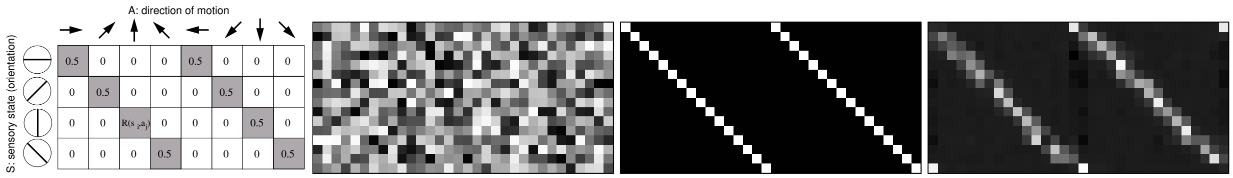
The remaining question is how can $R(s(t), a(t))$ be learned? For that, we will use a simple update rule:

$$R^{<t+1>}(s(t), a(t)) = R^{<t>}(s(t), a(t)) + \alpha\ \rho(t+1), \tag{1}$$

where $R^{<t+1>}(\cdot, \cdot)$ is the reward probability function at time $t+1$, and α the learning rate parameter. Finally, $R^{<t+1>}(s(t), a)$ values are normalized by $\sum_{a' \in A} R^{<t+1>}(s(t), a')$ for all $a \in A$. The algorithm above takes elements from reinforcement learning (Sutton and Barto 1998; Werbos 1987), especially the SARSA algorithm ($R(s, a)$ is similar to $Q(s, a)$). Figure 4 shows pilot results of learned $R(s, a)$ values, and Figure 5 behavioral results.

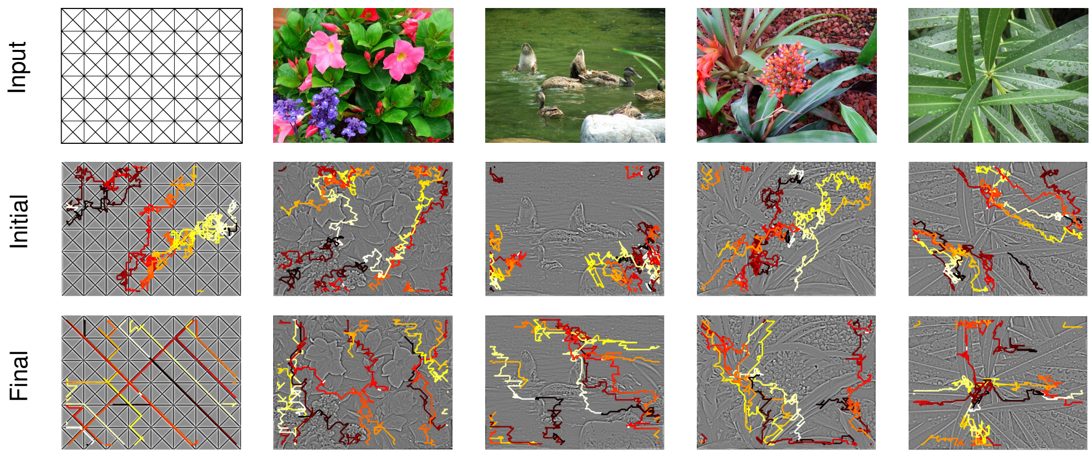
Research Tasks: Based on these preliminary results, we propose to carry out the following: (1) test with an extended set of natural images to assess the effect of image statistics and noise on learned sensorimotor mapping (see Choe and Miikkulainen [2004], Miikkulainen et al. [2005], Park et al. [2009a], and Park et al. [2009b] for PI's related work on natural image statistics); (2) implement the algorithm on a pan-tilt webcam (Logitech Orbit) for physical grounding (see preliminary implementation and results reported in Choe et al. [2007]); (3) add lateral inhibition in $\mathbf{r}$ so that the response becomes more sharply tuned, leading to faster convergence; (4) increase the number of filters/action primitives used (i.e., scale up); (5) extend the learning rule to include reward look-ahead (short-range prediction); and (6) explore agent setup where sensory and motor primitives are yet not formed (see our pilot study in Choe et al. [2008] and Yang and Choe [2007]).

Evaluation: The effectiveness of the learning algorithm will be based on two criteria: (1) K-L divergence (Kullback and Leibler 1951) between the reward probability distribution $R(s, a) = P(a|s)$ and the ideal distribution (figure 4c); and (2) windowed average of reward values over time



(a) R table (b) Initial R (c) Ideal R (d) Final R

Figure 4: **Initial, Ideal, and Learned $R(s, a)$ Values.** (a) The reward probability function $R(s, a)$ is shown as a table. The rows represent sensory states and the columns motor primitives. In an ideal, predictable environment with straight lines only, the table should have a diagonal structure as shown: two directions of motion (e.g., 0°and 180°) per one orientation (0°), marked in gray. Natural images do exhibit a collinear structure (Geisler et al. 2001), thus the ideal case shown here can be a good approximation. (b-d) The grayscale representation of the (b) initial, (c) ideal, and the (d) learned R tables are shown. White represents the maximum value, and black the minumum value. The learned R in (d) shows a clear diagonal structure, similar to the ideal case (c). (Unpublished pilot results are shown.)



(a) Grid (b) Flowers (c) Ducks (d) Plant (e) Oleander

Figure 5: **Gaze Trajectory at the Beginning and at the End of Learning (Pilot Result; Color Figure).** Gaze trajectories near the initial and the final stages of training are shown. The color represents the flow of time, where black $\rightarrow$ red $\rightarrow$ yellow $\rightarrow$ white is repeated every 768 iterations. (a) The initial trajectories are more like random walk. (b) The final trajectories show straighter motions, and those along strong edge features in the input image. Note that such motions maximally maintain invariance in the internal perceptual state of the agent, and the property of those motions (oriented motion in a particular angle θ) exactly reflects the stimulus property of the current perceptual state (i.e., the current input has orientation θ). The trajectory wraps around the borders since a toroidal boundary condition was used. **NOTE:** *The DoG-filtered grayscale image in the background is shown only as a reference. The learning algorithm had access only to the orientation filter response.*

and reward distribution. Even though the algorithm itself is purely unsupervised, in the sense that it does not receive particular $R(s, a)$ value as a target, as shown in figure 4(c), we can assume a theoretically ideal target. Note that the ideal $R(s, a)$ is not used in the training; it is only used for measuring the accuracy of the learned $R(s, a)$. The second measure will show whether the algorithm converges.

4.2 Grounding of Compositions of Geometric Concepts

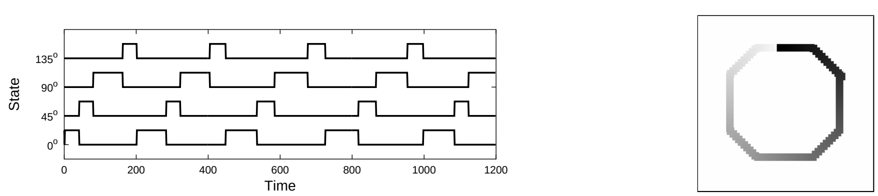
So far we have seen how simple geometric concepts that are encoded in the internal state of an agent can be grounded using action. However, it is not straight-forward how to extend this to more complex geometric concepts such as shapes or arbitrary combinations of the simple geometric concepts. In symbol grounding, compositionality of meaning is an important requirement (Harnad 1990), and thus as the second objective, we propose to extend our action-based grounding approach in objective 1 to include compositions of simple geometric concepts.

Extending to Spatio-temporal Invariance: The task of the agent is to understand, based on the changes in its internal state over time, properties of the input that the internal state is reflecting. The same principle of invariance can again be used in this situation to recover meaningful action sequences that reflect the geometric properties of the stimulus object.

The first approach is to extend the detection of invariance from that of a single sensory unit to that of a spatio-temporal pattern in the sensory array. Let us consider an example. Figure 6(a) shows an internal state activity pattern over time. We can ask the same question: What does this spatio-temporal response mean? The problem is that we cannot use the same internal state invariance as we used in Section 4.1. At the level of a single orientation detector, internal state invariance is only maintained along each of the eight sides of the octagon, but if we look at the spatio-temporal pattern in the orientation responses, we can see that a fixed pattern repeats itself (Figure 6(b)). We can treat this spatio-temporal pattern as an entity of its own and ask the same question of grounding: What does this pattern mean? The answer is the same as before, based on the same argument followed in Figure 4.1: the meaning of the internal state pattern is identical to the property of the motor action sequence that keeps the internal spatio-temporal pattern invariant over time.

The second approach is more complex, but in this case the same single-unit-level invariance criterion from Section 4.1 can be used, if the input encoding becomes more elaborate. For example, if there is a higher level of encoding that responds to a specific spatio-temporal pattern in the low level sensory state (note, not in the input itself), we can ask the same question of grounding: What does the higher level unit's spike mean? The answer would be the same as before: A sequence of motor actions that maintains invariance in the higher level unit. The sequence of actions that maintains such an invariance at a higher level would be able to reflect a more complex stimulus property (i.e., structure) that gave rise to a particular spatio-temporal pattern in the first-order sensor array.

Now we can think of two main tasks, corresponding to the two approaches discussed above: grounding of compositions of geometric concepts by (1) using an extended notion of internal state invariance, and (2) maintaining single-unit internal state invariance while making the architecture

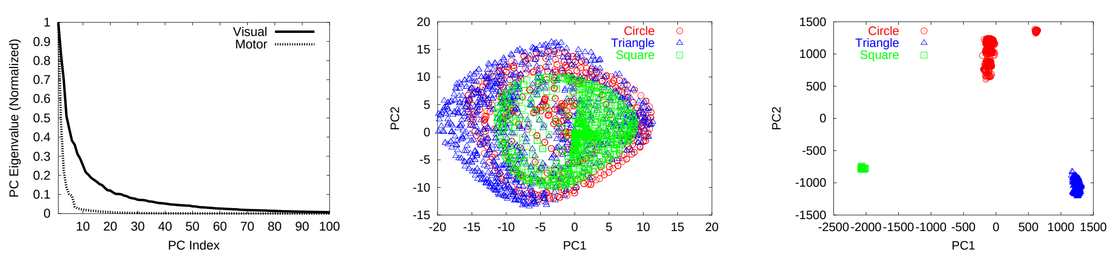


(a) Internal state (b) Gaze trajectory

Figure 6: **Inferring Stimulus Properties of a Complex Object.** The agent's dynamically changing internal state and the corresponding gaze trajectory are shown. (a) The activation state of the four neurons in the agent is shown over 1200 time steps. (b) The agent's position is plotted over 500 time steps, where the grayscale indicates the time step (white is time 0, and black is time 499). As the agent repeatedly directs its gaze around the octagon, the internal state also shows a repeating pattern, from 0°, 45°, 90°, 135°, back to 0°, for example (the pattern in the interval $[200, 449]$). Note that the entire trace in (b) corresponds to the interval $[0, 499]$ in (a). The plots show actual data from a preliminary simulation using a deterministic exploration strategy.

hierarchical so that higher level units can respond to specific spatio-temporal patterns in the sensory array dynamics. However, as we will see below, it is not possible to integrate task 1 into the reinforcement learning framework in Section 4.1. So, we will focus on task 2 and its various subtasks. In both tasks (let us consider both for now), the meaning of a sequence of internal state transitions (spatio-temporal pattern) will have to be inferred based on the corresponding motor sequence. The immediate problem for both tasks is that it is difficult to use the same discrete tabular representation of $R(s, a)$ used in Section 4.1 (this is a common problem in scaling up reinforcement learning). To address this problem, we will collect a large data set of gaze trajectories using the trained agent from Section 4.1 (see Figure5 for example gaze trajectories). We can then cut them up in random length and stretch or compress them to fit a predetermined unit length. Then, we can do cluster analysis (e.g., using PCA) to come up with a small set of candidate motor patterns that will serve as the "gesture" pool, upon which the internal sesory state will be grounded. Our initial results with gaze trajectories from three stereotypical objects of various size and location show that these gaze trajectories can be very tightly clustered (Figure 7). Incidentally, by stitching together individual sensory states that correspond to each gaze location in the gestures, we can also construct the candidate sensory array pattern pool. If we get n of these gestures and sensory patterns, we can have a $n \times 2n$ reward table, by doubling the gesture pool by flipping the direction, and thus reuse the same reinforcement learning paradigm in Section 4.1. This method is a good fit for the second approach above, but it would not go well with the first approach since the state to action pairing is unclear.

Research Tasks: Based on the general method above, we will investigate the following: (1) Analyze local gaze trajectory statistics, to aid in the construction of the gesture pool. Such stereotypical motion sequences are similar to whole gestures encoded in the premotor cortex (Graziano 2009). For this work, we will leverage on our extensive body of work on the computational model of cortical development (Choe 2001; Choe and Miikkulainen 2004; Miikkulainen et al. 2005; Park et al. 2009a,b). (2) Based on the result from (1), construct sensory internal state pattern pool. Such patterns may be similar to higher-level object features encoded in the temporal cortex (Tanaka 2003). (3) Run reinforcement learning algorithm in Section 1 to learn the grounding. This step will simulate action-based grounding of higher level object feature encoding (composite geometric features). (4) We will also assess the possibility of mechanisms like this accounting for "mirror neurons" (neurons in the prefrontal cortex that respond to visually observed gesture and also activate analogous motor behavior; Rizzolatti et al. [2001]). (According to Fuster [1997], such sensorimotor coupling exists at all levels of the brain hierarchy, and our model provides an explanation of why



(a) PC Eigenvalues (b) Visual Representation (c) Motor Representation

Figure 7: **Comparison of Visual vs. Motor Representations.** Principal components analysis results are shown for visual vs. motor representations of shapes (circle, triangle, and square). (a) First 100 eigenvalues are shown. Projections onto the first (PC1) and second (PC2) eigenvectors are shown for the (b) visual and (c) motor representations. Motor representations show clear class separability, while the visual representations are mostly overlapping.

it may be so, based on arguments from grounding.) For the actual training, at first, we will test the steps above using synthetic images (Figure 5(a) top row). For this input, compositions of features would be limited (angles, multiple angles, etc.), so it will help establish feasibility. Once the attempt is successful, we will move on to natural images, with or without basic shapes embedded in the scene.

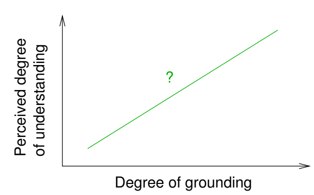
Evaluation: We will use the same evaluation metrics as proposed in Section 4.1 to evaluate the quality of grounding for compositions of geometric features. We will also conduct analysis on the learned gesture pool and sensory internal state pattern pool. For example, we can measure how much variance in the sensory/motor space the first few primitives in the pool account for.

4.3 Evaluate Human Attitude Toward Grounded Computing Artifacts

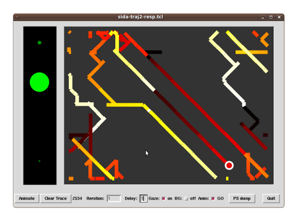
There are important implications of our theory of grounding on human-computer interaction. First, we will assess human attitude toward grounded computing artifacts, while leaving the linkage with human-computer interaction to objective 4 (Section 4.4). In this section, we will assess the following: (1) How much "understanding" humans attribute to computing artifacts that exhibit varying degrees of grounding? (2) Can humans tell apart manually grounded vs. autonomously grounded computing artifacts?

(1) Attribution of Understanding to Computing Artifacts that Have Grounding: How much "understanding" do humans attribute to computing artifacts? For example, when a user punches in "1+1=" into one's calculator and get back "2", would the user say "wow, this machine understands arithmetic and the concept of numbers!"? (See Searle [1997] on intrinsic intentionality.) For this first study in this section, we will measure human users' perceived degree of understanding (dependent variable) as we vary the degree of grounding (independent variable). We expect that humans will attribute more understanding to computing artifacts that have higher degree of grounding.

The key issues here are how to control the degree of grounding and how to measure the perceived degree of understanding. For the former, for a given bit of geometric information (say orientation or shape), we will display, one of the following at a time, for the same duration of time: (1) symbols with their forms unrelated to the actual meaning (star, heart, Greek symbols, etc.), (2) numerals, (3) static graphical displays, (4) dynamic graphical displays (animation version of (3)), and (5) embodied gestures, expressing the information. These different incarnations corresponding to an increasing degree of grounding. To measure the perceived degree of understanding, we will simply ask the human user to rate between a numerical scale of 1 to 5, with 1 being the least degree of understanding (equal to an explicit lookup table), and 5 being the most degree of understanding (close to that of a human). Human users will be prompted by the information, shown the five different alternatives, and for each alternative, asked to rate them on the 1 to 5 scale. For each user, 45 different stimulus conditions will be shown (6 orientations and 3 shapes $\times$ 5 degrees of grounding) in a random order. We expect the correlation to be positive (Figure 8(a)). For this study, we will recruit 20 graduate and undergraduate students in the Department of Computer Science.



(a) Grounding vs. Understanding



(b) Grounding Demo App

Figure 8: **Measuring Human Attitude Toward Grounding.** (a) An illustration of expected correlation between degree of grounding and degree of perceived understanding. (b) Demo app showing agent internal state (left) and gaze trajectory (right).

Prior to the experiment, to rule out bias, we will not explain to the study participants the concept of grounding. Data collection, retention, analysis, and specific experimental setup will follow IRB guidelines.

(2) Telling Apart Manually vs. Autonomously Grounded Computing Artifacts: If we consider the original formulation of the Turing Test proposed by Alan Turing (Turing 1950) we are left with the question of whether human users could distinguish between computing artifacts that are manually grounded vs. those that achieve grounding autonomously. For example, consider the reward table $R(s, a)$ in Figure 4(c) and (d). One is hand-coded (c), and one is learned (d). However, the resulting behavior of the agent driven by the policy (state to action mapping) would look almost the same. So, the question is, can humans tell apart the two? This is an important question since our emphasis so far was on *learning* grounding, and if it is the case that hand-coded grounded systems are perceived as equal to autonomously learned grounded systems, then the necessity for learning diminishes.

We will follow the same experimental design explained right above, by replacing the five alternatives (for degree of grounding) for the independent variable with three alternatives: (1) manually grounded, (2) autonomously grounded (learned), and (3) no grounding (random), to serve as the baseline. The dependent variable will be the same and will be measured using the same 1 to 5 scale. We will use a visual agent running a policy based on the (1) ideal reward table (Figure 4(c)) to show the manually grounded case, (2) the learned reward table (Figure 4(d)), and (3) random reward table (Figure 4(b)) for the three alternatives, respectively. Figure 8(b) shows an app that will display the gaze behavior based on these three policies.

Research Tasks: The main research tasks here will be data analysis and calibration of the experimental protocol. (1) Different users may have a different general attitude toward grounding. That is, the max (and min) response would differ among individuals. So if this is found to be a prevalent trend, we will scale the response between min and max to a fixed scale. (2) There is a possibility that the interval between the alternatives in the independent variable are not linearly scaled. If that turns out to be the case, we will perform nonlinear fitting to correct it.

Evaluation: For the first task, we will use the correlation coefficient to quantitatively measure the correlation between the independent and the dependent variables. We expect to find a strong positive correlation. For the second task, we will use t-test to quantify the pair-wise differences between manual, learned, and random grounding. Here, we expect to find *no difference* between manual vs. learned conditions, but the grounded cases will give higher scores compared to the random case. Objective 4 will investigate under what conditions manual and learned grounding can actually make a difference, despite the expected negative finding here.

4.4 Measure Enhancements in Human-Computer Communication Due to Grounding

The last objective is to measure the enhancement in efficiency of human-computer communication in computing systems with intrinsic semantics.

Identifying an Application Domain where Grounding Matters: First, we will identify an application domain where grounding (especially the autonomous kind) does matter. As discussed in the background section, Traum (1999) is perhaps the only work that discusses grounding (albeit in a slightly different sense) as relevant to human-computer interaction. Traum's work builds up on existing theories of collaboration/cooperation in the cognitive science and artificial intelligence literature, such as the contribution model by Clark and Schaefer (1992), joint intentions/teamwork model by Cohen and Levesque (1991), and shared-plans model by Grosz and Kraus (1993). In Traum (1999), the focus is on establishing common ground among users of a collaboration platform. The users need to establish a "shared belief" so that their individual efforts contribute to the success of the common task.

Although Traum's usage of the term grounding is different from ours, the application framework is quite relevant: collaborative systems. In the work, each user in collaboration has a shared grounding but the grounding can change dynamically and refined over time, through "grounding acts" such as acknowledge, repair, request repair, request ack, etc. This gives us a key insight on why grounding can be important, especially the autonomous kind, as opposed to the fixed, manual kind. When the grounding is not fixed but changing, manually grounded systems will not be able to adapt. Thus in this kind of environment, human users would be able to tell apart manually grounded vs. autonomously grounded computing artifacts, and potentially benefit from the autonomously grounded ones.

(1) Human Response to Dynamically Changing Grounding: Based on the above insight, first we will repeat experiment 2 from objective 3, where as in this case the grounding changes over time. That is, the mapping between symbol and its meaning will change over time while the human user should give the same rating on the 1 to 5 scale. Here, it will be critical to allow enough time (to be empirically determined) to the user so that they can get used to the mapping before we change the mapping. Also, we will have to let the users know when the mapping changed and how. For the random and the manually grounded case, fixed policy will be used to generate the gaze in the app (Figure 8(b)), but for the autonomously grounded case, we will use policy taken at a fixed interval over the retraining (of the agent) period of the new mapping (this retraining will be done off-line and made ready for each different mapping so that the retrained policy can be used during human trial without delay).

(2) Human Performance in Collaborative Systems with Dynamically Changing Grounding: Next, we will build a simple collaborative platform where human users can exchange information with other human users while trying to reach a common goal. The users will be (1) unaided, or aided by computer agents that are either (2) ungrounded, (3) manually grounded, or (4) autonomously grounded. Each task trial will involve two human users, each with the same set of concepts represented in two different encodings (different set of symbols). Each user will know what the symbol means (grounding), but initially, the user will not know the partner's grounding. The task of the users is to produce a "shared grounding" by communicating with the partner. The main idea is to have assistive agents for each user, and have these agents interact with each other to help shared grounding of the human users. The humans will be allowed to observe the interaction between the agents (showing off and trying to match their gestures) that will be shown as an animation. The same protocol we used for various degrees of grounding and grounding in a static vs. dynamic environment will be applied to this domain. We expect to find that human users assisted by grounded agents will learn the shared grounding faster than when unassisted or when assisted by ungrounded agents. In a changing environment, autonomously grounded assistive agents would be found most effective.

Research Tasks: Again, the main research tasks would be data analysis and calibration: Various experimental parameters need to be determined for both methods outlined above. For example, for the dynamic case, how often can the grounding be changed so that human users can catch up? Furthermore, for the second method above, the specific collaborative work environment will need to be refined over time based on pilot experiments.

Evaluation: We will primarily measure the time it takes until shared grounding is reached between the two human users. It is yet unclear how much variation in performance exists, so it is hard to determine beforehand how many study participants we will need for sufficient statistical power. We will run pilot experiments to measure the variability across pairs of individuals, and use that the select an appropriate number. We will use t-test or other statistical tests to compare between different experimental conditions.

5 Curriculum Development Activities

Based on the proposed work, the PI will develop specialized course content for the open seminar course on "Intelligent User Interfaces" (CSCE 634) offered in the Department of Computer Science and Engineering. This course was specifically designed from the inception to accommodate a wide range of topics taught by different instructors, so it is ideal for the purpose of this curriculum development plan. The course content will include significant expansions of the materials presented in the background (Section 2), prior work (Section 3), and the main research resulting from this project (Section 4). The course will also include rigorous coverage of research methods in human-computer interaction (Lazar et al. [2010] could serve as an excellent supplementary textbook). Initially, students will be given the user interface framework developed in this project so that they can have a concrete standing. The students will experience grounded vs. ungrounded interfaces and assess their relative merits. Subsequently, students will be asked to identify new forms of grounding, and design user interfaces of their own based on those new groundings.

6 Broader Impact Plan

Student training and under-represented groups: This project will support the training of two Ph.D. students (through direct funding from this grant) and three undergraduate students (through a separate REU supplement) over the course of three years. The PI has extensive experience mentoring Ph.D. students (9 graduated, 2 current; 2 female), M.S. students (10 graduated, 2 current; 1 female) and undergraduate students (total of 19; 3 female, 3 under-represented minority; NSF REU site, REU supplement, and CRA DREU funding was utilized). We will continue our effort in this direction through REU supplements. The framework resulting from this project (including the code, data, and user interface) also gives us an opportunity to reach out to K-12 students. PI Choe has had significant experience in interactions with high-school level groups (lab tours and organizing high school contest). We will utilize the contacts from these prior engagements to reach out to the K-12 students. Specific activity will include a fun online user interface demo based on the apps from objective 3 and 4. Our past efforts have demonstrated that we are able to incorporate under-represented groups in the project, such as women and minority students (primarily at the undergraduate level). We will continue to involve and reach out to these groups in our proposed project. **Data and code dissemination:** Please refer to the Data Management Plan. **Benefits to the society:** For long-term benefits, please refer to the project summary and introduction. We will organize a workshop on intrinsic semantic systems and their potential impact on human-computer communication, and co-author a white paper for publication in popular science and technology magazines for broader access by the public. The PI has extensive experience in organizing workshops and symposia (co-charing a Society for Neuroscience minisymposium in 2008 and a workshop at the Computational Neuroscience conference in 2010), and his research was introduced in popular magazines such as *New Scientist* (Williams 2011) and a popular science book *Portraits of the Mind* (Schoonover 2010). The PI also contributed to two science documentary film projects. The PI will continue with such efforts to make available to the public exciting scientific advances resulting from this project.

REFERENCES CITED

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7:84–91.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4:91–99.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*, May, 28–37.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

Cangelosi, A., and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30:673–689.

Choe, Y. (2001). *Perceptual Grouping in a Self-Organizing Map of Spiking Neurons*. PhD thesis, Department of Computer Sciences, The University of Texas at Austin, Austin, TX. Technical Report AI01-292.

Choe, Y., and Bhamidipati, S. K. (2004). Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action. In Ijspeert, A. J., Murata, M., and Wakamiya, N., editors, *Biologically Inspired Approaches to Advanced Information Technology*, Lecture Notes in Computer Science 3141, 176–188. Berlin: Springer.

Choe, Y., Mayerich, D., Kwon, J., Miller, D. E., Chung, J. R., Sung, C., Keyser, J., and Abbott, L. C. (2011a). Knife-edge scanning microscopy for connectomics research. In *Proceedings of the International Joint Conference on Neural Networks*, 2258–2265. Piscataway, NJ: IEEE Press.

Choe, Y., Mayerich, D., Kwon, J., Miller, D. E., Sung, C., Chung, J. R., Huffman, T., Keyser, J., and Abbott, L. C. (2011b). Specimen preparation, imaging, and analysis protocols for knife-edge scanning microscopy. *Journal of Visualized Experiments*, 58:e3248. doi: 10.3791/3248.

Choe, Y., and Miikkulainen, R. (2004). Contour integration and segmentation in a self-organizing map of spiking neurons. *Biological Cybernetics*, 90:75–88.

Choe, Y., and Smith, N. H. (2006). Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone. In Gil, Y., and Mooney, R., editors, *Proceedings of the 21st National Conference on Artificial Intelligence(AAAI 2006)*, 936–941.

Choe, Y., Yang, H.-F., and Eng, D. C.-Y. (2007). Autonomous learning of the semantics of internal sensory states based on motor exploration. *International Journal of Humanoid Robotics*, 4:211–243.

Choe, Y., Yang, H.-F., and Misra, N. (2008). Motor system's role in grounding, receptive field development, and shape recognition. In *Proceedings of the Seventh International Conference on Development and Learning*, 67–72. IEEE.

Chung, J. R., Sung, C., Mayerich, D., Kwon, J., Miller, D. E., Huffman, T., Abbott, L. C., Keyser, J., and Choe, Y. (2011). Multiscale exploration of mouse brain microstructures using the knife-edge scanning microscope brain atlas. *Frontiers in Neuroinformatics*, 5:29.

Clark, H. H., and Schaefer, E. F. (1992). Contributing to discourse. *Cognitive Science*, 12:259–294.

Cohen, P. R., and Levesque, H. J. (1991). Teamwork. *Nous*, 35.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856.

Doyle, M. C., and Walker, R. (2002). Multisensory interactions in saccade target selection: Curved saccade trajectories. *Experimental Brain Research*, 142:116–130.

Fuster, J. M. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and the Frontal Lobe*. Lippencott-Raven. Third edition.

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge Co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724.

Gershon, N. D., and Miller, C. G. (1993). Dealing with the data deluge. *IEEE Spectrum*, 30:28–32.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9:558–565.

Graziano, M. (2009). *The Intelligent Movement Machine: An Ethological Perspective on the Primate Motor System*.

Grosz, B. J., and Kraus, S. (1993). Collaborative plans for group activities. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 367–373.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148:574–591.

Jokinen, K. (2009). Gaze and gesture activity in communication. In Stephanidis, C., editor, *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, vol. 5615 of *Lecture Notes in Computer Science*, 537–546. Springer Berlin / Heidelberg.

Kuipers, B., Beeson, P., Modayil, J., and Provost, J. (2006). Bootstrap learning of foundational representations. *Connection Science*, 18:145–158.

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Kwon, J., Mayerich, D., and Choe, Y. (2011). Automated cropping and artifact removal for knife-edge scanning microscopy. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 1366–1369.

Lazar, J., Feng, J. H., and Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. West Sussex, UK: Wiley.

Mann, T. A., and Choe, Y. (2011). Scaling up reinforcement learning through targeted exploration. In *Proceedigs of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 435–440.

Mayerich, D., Kwon, J., Panchal, A., Keyser, J., and Choe, Y. (2011a). Fast cell detection in high-throughput imagery using gpu-accelerated machine learning. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 719–723.

Mayerich, D., Kwon, J., Sung, C., Abbott, L. C., Keyser, J., and Choe, Y. (2011b). Fast macro-scale transmission imiging of microvascular networks using kesm. *Biomedical Optics Express*, 2:2888–2896.

Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Berlin: Springer. URL: http://www.computationalmaps.org.

Modayil, J., and Kuipers, B. (2007). Autonomous development of a grounded object ontology by a learning robot. In *Proceedings of the 22nd National Conference on Artificial Intelligence*, 1095–1101.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):883–917.

Park, C., Bai, Y. H., and Choe, Y. (2009a). Tactile or visual?: Stimulus characteristics determine receptive field type in a self-organizing map model of cortical development. In *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Multimedia Signal and Vision Processing*, 6–13. **Best Student Paper Award**.

Park, C., Choi, H., and Choe, Y. (2009b). Self-organization of tactile receptive fields: Exploring their textural origin and their representational properties. In *Advances in Self-Organizing Maps: Proceedings of the 7th International Workshop, WSOM 2009*, 228–236. Heidelberg: Springer.

Philipona, D., O'Regan, J. K., and Nadal, J.-P. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation*, 15:2029–2050.

Pierce, D. M., and Kuipers, B. J. (1997). Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:162–227.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2:661–670.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(11):583–601.

Schoonover, C. (2010). *Portraits of the Mind: Visualizing the Brain from Antiquity to the 21st Century*. New York, NY: Abrams.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–424.

Searle, J. R. (1997). The explanation of cognition. In Preston, J., editor, *Thought and Language*. Cambridge, UK: Cambridge University Press. Reprinted in (Searle 2002), Chapter 7.

Searle, J. R. (2002). *Consciousness and Language*. Cambridge, UK: Cambridge University Press.

Shadbolt, N., Hall, W., and Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 21:96–101.

Sheu, P. C.-Y., Yu, H., Ramamoorthy, C. V., Joshi, A. K., and Zadeh, L. A., editors (2010). *Semantic Computing*. Piscataway, NJ: IEEE Press.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, 13(1):90–99.

Tao, X., Li, Y., and Nayak, R. (2008). A knowledge retrieval model using ontology mining and user profiling. *Integrated Computer-Aided Engineering*, 15:313–329.

Traum, D. R. (1999). Computational models of grounding in collaborative systems. Technical Report FS-99-03, AAAI.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, LIX:433–460.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Werbos, P. (1987). Learning how the world works: Specifications for predictive networks in robots and brains. In *Proceedings of 1987 IEEE International Conference on Systems, Man, and Cybernetics*.

Williams, C. (2011). Eight arms, big brain: What makes cephalopods clever. *NewScientist*, 2816:36–39.

Yan, R., Hauptmann, A. G., and Jin, R. (2003). Negative pseudo-relevance feedback in content-based video retrieval. In *Proceedings of the eleventh ACM international conference on Multimedia*, MULTIMEDIA '03, 343–346. New York, NY, USA: ACM.

Yang, H.-F., and Choe, Y. (2007). Co-development of visual receptive fields and their motor-primitive-based decoding scheme. In *Proceedings of the International Joint Conference on Neural Networks 2007 Post conference Workshop on Biologically-inspired Computational Vision (BCV) 2007*. [Online] `https://umdrive.memphis.edu/iftekhar/public/IJCNN/BCV.htm`.

Yang, H.-F., and Choe, Y. (2011a). Ground truth estimation by maximizing topological agreements in electron microscopy data. In *Proceedings of the 7th International Symposium on Visual Computing (LNCS 6938)*, 375–384.

Yang, H.-F., and Choe, Y. (2011b). An interactive editing framework for electron microscopy image segmentation. In *Proceedings of the 7th International Symposium on Visual Computing (LNCS 6938)*, 404–413.

Ziemke, T. (1999). Rethinking grounding. In Riegler, A., von Stein, A., and Peschl, M., editors, *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?*, 177–199. New York: Kluwer Academic/Plenum Press.