

Visual Affordances and Symmetries in *Canis habilis*: A Progress Report

Thomas E. Horton, Lloyd Williams, Wei Mu, and Robert St. Amant

Department of Computer Science
North Carolina State University
Raleigh, NC 27695

tehorton@ncsu.edu, lwillia@ncsu.edu, wmu@ncsu.edu, stamant@csc.ncsu.edu

Abstract

This paper describes the visual system of *Canis habilis*, a prototype tool-using robot under development. *Canis* identifies affordances in its environment in terms of symmetries between itself and objects and between objects. The implementation is incomplete, but shows promising preliminary results. The paper briefly sets the context of our research in tool use, based on the study of animal behavior, describes a symmetry-based approach to identifying a limited set of visual affordances, and presents an example of the application of the system to a simple task.

Introduction

The concept of affordance has considerable appeal for researchers in robotics and artificial intelligence. Gibson's familiar definition is as follows (Gibson 1979):

The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb to *afford* is found in the dictionary, but the noun *affordance* is not. I have made it up. I mean by it something that refers both to the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.

One reason for the popularity of the concept is its generality: it is non-committal about the specifics of an agent's architecture (physical or cognitive) and its environment, putting the focus on complementary relationships between the two. This generality has led, perhaps not surprisingly, to considerable disagreement about exactly what affordances are, with controversies to be found in the literature of cognitive science (Vera & Simon 1993), human-computer interaction (St. Amant 1999), and even the home discipline of the concept, ecological psychology (Jones 2003).

A number of projects in AI, robotics, and machine vision have been influenced by the concept of affordances; Şahin et al. (2007) provide the most comprehensive technical survey we are aware of. Most work on affordances in these areas has concentrated on how to identify relationships between properties of the environment and functions that a given agent

is capable of. An agent that can easily distinguish between what it is possible to do and what it is not possible to do can, in principle, make better decisions about more general courses of action.

One idea in ecological psychology that has received less attention in AI and robotics is direct perception, the possibility that "meaning is in the environment" (Chemero 2003). While we do not accept direct perception, we find a related issue interesting: If we are building a tool-using robot that should be able to recognize visual affordances, what mechanisms should we engineer into its visual system?

In this paper, we outline an approach to detecting visual affordances based on the concept of symmetry (Thimbleby 2001). Our test platform is a prototype tool-using agent we call *Canis habilis* (Wood, Horton, & St. Amant 2005; St. Amant & Wood 2005). In the remainder of this paper, we briefly survey examples of relevant research in the robot and machine vision literature, describe our approach, and give a summary of the status of the system.

Affordances and tool use

Our work examines affordances in the context of a general class of activities, those related to tool use. We define tool use as follows (St. Amant & Horton 2008, p. 1203):

Tool use is the exertion of control over a freely manipulable external object (the tool) with the goal of (1) altering the physical properties of another object, substance, surface or medium (the target, which may be the tool user or another organism) via a dynamic mechanical interaction, or (2) mediating the flow of information between the tool user and the environment or other organisms in the environment.

This definition is based on a survey of animal tool-using behavior; it attempts to integrate and improve on a historical set of definitions of tool use in the field of animal behavior (van Lawick-Goodall 1970; Alcock 1972; Beck 1980).¹

Tool use provides a fertile domain for studying affordances. A tool-using agent may need to recognize affordances that relate to how the agent interacts with a tool

¹Our exchanges with philosophers of mind and animal behavior researchers suggest that some of our interpretations may be controversial.

(proximal affordances) as well as affordances that relate to how the tool interacts with other elements of the environment (distal affordances).

There is a natural affinity between the concepts of affordance and tool use. Because of the enormous differences in the cognitive and physical architectures of agents capable of tool use (e.g., from wasps to humans (Oswalt 1973)), it makes sense to consider how the environment might provide opportunities for an agent to use tools, even if the agent has extremely limited or non-existent cognitive processing capabilities.

It is easiest to see examples of the application of affordance concepts to animal tool use in laboratory experiments. For example, Visalberghi and Limongelli (1996) describe experiments in which capuchin monkeys try to obtain a reward within a transparent tube; reeds are provided as tools that can be inserted into the tube, which is too narrow to allow entry of a hand. Povinelli (2000) has carried out more detailed experiments that require chimpanzees to push a rod through a narrow opening to retrieve a reward. The rod can be grasped on either end; one end is too large to fit through the opening, but is also hard to grasp because of its shape. Chimpanzees have difficulty solving this problem, choosing to grasp the end of the rod that is easy to grasp, which is not part of the solution.

We also find examples of affordances in tool use in natural environments, though the application of the concept is less obvious. Fishing for termites, probing logs with sticks, and using leaves to wipe fur by chimpanzees and orangutans are examples (Whiten *et al.* 1999; van Schaik *et al.* 2003); the use of gloves for foraging by dolphins is another (Krützen *et al.* 2005). Affordances do not capture all aspects of effective tool use, but they seem to play a role in many cases. Our particular interest is in affordances that can be recognized visually.

Related work

A number of projects in robotics and machine vision bear on affordances, specifically in the context of tool use.

Of relevance is the work of Bogoni and Bajcsy, who examined ways of representing tools in terms of function (Bogoni & Bajcsy 1993; 1994). Suggesting that object representations should go beyond the description of intrinsic properties such as shape and material, their goal was to include empirically determined descriptions of an object's functional features. To this end, they developed a system for automatically discovering the functional properties of objects through task-oriented interactions using active perception.

GRUFF (Generic Recognition Using Form and Function) was a system developed for the visual recognition of novel objects (Stark & Bowyer 1996). Rather than using explicit geometrical models of specific objects, GRUFF tried to assign objects to generalized categories based on their functional characteristics (it is assumed that these functional characteristics can be extracted from an object's geometry). This functional analysis involves determining whether a given object satisfies the constraints of the object category (for example, a container must be graspable to be a cup).

The functional characteristics GRUFF looks for are similar in concept to Gibson's affordances.

Kemp and Edsinger (2005) observe that for many tools and tasks, a basic, yet essential, component of tool-use is the identification and control of the tool's endpoint. To take advantage of this observation, they designed a robotic system for automatically detecting a tool's endpoint and estimating its position in space relative to the robot's hand, using optical flow techniques. With this information, the robot has control over the position of the endpoint, which might enable the use of very simple tools (a hook, for example) as well as support further automated exploration and testing of more complex tools.

Stoytchev's work on the autonomous learning of tool affordances by a robot is of particular interest (Stoytchev 2005). The robot, consisting of a five d.o.f. manipulator arm and gripper mounted on a mobile base was able to use tools for basic extension-of-reach tasks. The tasks were comparable to those performed by chimpanzees in experiments by Povinelli (2000). Through exploration and interaction with a tool and a target, the robot was able to autonomously learn the tool's affordances.

An earlier version of *Canis habilis* was developed by Wood *et al.* (Wood, Horton, & St. Amant 2005; St. Amant & Wood 2005). Based on the Sony Aibo, the original version of *Canis* was a proof of concept built to solve a version of the classic monkey and bananas problem. The main contribution of the project was architectural, demonstrating tool use in a mobile robot; visual processing and low level motor control relied on functionality provided by the Tekkotsu framework (Touretzky & Tira-Thompson 2005).

Affordances as symmetries

A promising approach to capturing the nature of visual affordances is suggested by Harold Thimbleby, who has proposed that *symmetries* give rise to affordances (Thimbleby 2001). Thimbleby, working in the field of human-computer interaction, observes that in many cases, affordances result from the visual symmetries displayed by objects in an agent's environment. For example, door knobs, radio dials, and screwdrivers all feature a rotational symmetry that affords turning.

Weyl (1983) presents a detailed exploration of the concept of symmetry, and provides an informal definition as follows:

A thing is symmetrical if there is something we can do to it so that after we have done it, it looks the same as it did before.

Using Weyl's work as a basis, Thimbleby (2002) developed a more formal definition of symmetry:

In general, if $p(v) = p(T(v))$ then a property p is symmetric under the transformation T . Further, T must be invertible, so $p(v) = p(T^{-1}(v))$ as well – e.g., if T is a rotation, the property is unchanged through a rotation, and a rotation back.

This definition is further extended by Cairns and Thimbleby (2008) to emphasize the role of perception. This new formulation introduces a perception function ψ which may

be used to reduce a system, S , to a set of psychological attributes. Thus, S is symmetric under T , if, after the transformation, it seems *to an observer* to remain unchanged, i.e., $\psi(S) = \psi(T(S))$.

In describing affordances, Thimbleby (2001) notes, “We say that an object may afford some or several sorts of action, and when it does so this is in some sense a set of natural or ‘easy’ relations.” Intuitively, symmetry seems a good candidate for a basis of affordances, as visual symmetries are often immediately obvious (Cairns & Thimbleby 2008): “many visual symmetries leap out at us without prompting.” Thimbleby (2001) suggests, “Affordance can then be defined as those symmetries that apply under the actions relevant to the activities or tasks that are performed with the object.”

We should note that Thimbleby often uses visual symmetries as inspiration for applying the concept of symmetry to more abstract realms such as state spaces (Thimbleby 2001). Conversely, we have taken Thimbleby’s approach as inspiration for the highly visual tasks involved in object interaction in general, and tool use in particular.

Thimbleby’s approach to visual symmetries also differs slightly from ours. A typical example from Thimbleby (2001) is a plain wooden pencil. In Thimbleby’s analysis, the pencil has two main symmetries: a hexagonal or circular symmetry around its length, and an approximate reflective symmetry about its center (i.e., both ends are roughly similar). The significance of the rotational symmetry is that the angle at which the pencil is grabbed does not affect its use. The significance of the reflective symmetry is that it may be easy to make errors by grabbing the wrong end, with the eraser end down and the writing point up. Thimbleby concludes that symmetries provide freedom in how an object is used, while approximate symmetries lead to errors when the symmetry expected by the user does not hold.

In this example, Thimbleby focuses on symmetry inherent in a single object; the pencil is self-symmetric under a rotational transform. For our purposes, we are more interested in symmetries between two or more objects. In effect, if one object were “transformed” into another, what properties would be unchanged (i.e., symmetrical)? The rotational symmetry of the pencil may indicate to a user that rotation is irrelevant to its use, but what properties of the pencil might indicate to a user that a pencil can be grasped in the first place? In this case, the user might perceive symmetries between the curvature and diameter of the pencil and the shape of the user’s hand when forming a precision grip, indicating that the pencil can be grasped, as well as suggesting possible effective orientations. Such matching symmetries are a natural analog, in part, for what Gibson refers to as the complementarity of the animal and the environment. Similarly, if the user wanted to use the pencil not as a writing implement, but as a tool to poke a target through a narrow tube, symmetries in curvature, diameter, and length between the pencil and the tube would all be relevant to the task. It is these kinds of visual symmetries *between* objects that are the focus of our work.

Thimbleby makes two important points about the limitations of a symmetry-based approach. First, symmetry cannot be used as a basis to formalize everything that affor-

dance has been used to describe (Thimbleby 2002). In other words, a symmetry-based definition of affordance may include examples some people do not consider to be affordances, while excluding other cases that some think should be included. Of more immediate relevance to our work, is Thimbleby’s second observation that while “symmetries *generate* affordances”, not all affordances are the result of symmetries (Thimbleby 2001). There may well be affordances that are important to a tool using agent, but that are the result of properties other than symmetry. We hope that by exploring the concept of affordance from the perspective of symmetry, our findings may also shed light on affordances in general, not only on those resulting from symmetries.

Current work

We have implemented a prototype system capable of performing a basic comparison of objects for evaluating symmetry. The system takes as input two image masks generated from *Canis*’s video stream using color segmentation. The 2D array of pixels for each mask is then translated into a occupancy array-based representation, similar to the one used by Glasgow and Papadias (1992). The occupancy arrays support representations of objects at multiple resolutions, which allows for fast, but coarse comparisons, as well as more precise, but slower analyses. While the prototype supports the use of 3d representations, we are currently only using it to compare the cross-sectional areas of objects in two dimensions.

Each object in a comparison is associated with a viewing region that can be scaled to encompass all of the object or to isolate a particular region for analysis (roughly analogous to an attention window). This supports the search for symmetries between sub-parts of objects, for example between the tip of a tool and a slot on the target. Various transforms may also be applied to the region selected by this volume, including translation, rotation, and scaling. This allows the system to be used to search for the orientation of the objects that yields the best match (for example, rotating the representation of a wrench around the representation of a bolt until the two are in alignment). At present, the search for an optimum match must still be controlled by a human. Automating this process is the current focus of our work, and we are testing the use of existing methods of shape matching (e.g., shape context (Belongie, Malik, & Puzicha 2002)) to identify potential matches.

When evaluating two representations, the system compares the sets of occupied regions in each of the transformed views. We are interested in the intersection of these two sets and the differences between them and with the bounding region. By comparing the relative areas of intersecting and non-intersecting regions, we can order different orientations by how good a match they provide. Similarly, we can also order the comparisons of one object to several potential matching objects to find the pairings with the highest degrees of symmetry.

Consider the objects in Figure 1. When comparing the ball (a) to the hole (b), if the hole completely contains the ball (all occupied regions in the representation of the ball are also within the representation of the hole), then we can

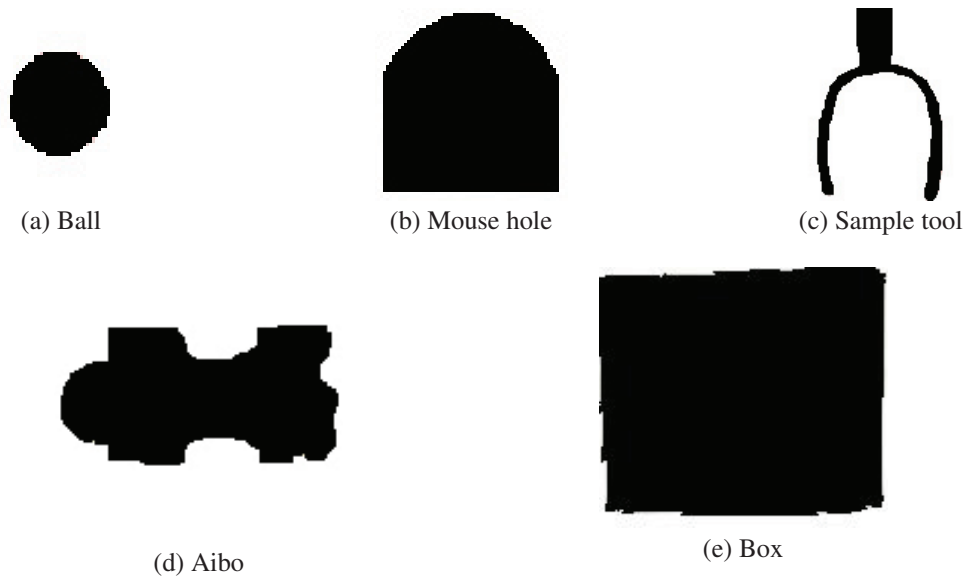


Figure 1: Object masks

assume that the ball can pass through the hole. If the size of the ball were increased, we would notice that the ratio of the area in the intersection to the area in the difference between the two sets increases - the ball fits closely into the hole. In some cases, such as the symmetry between a screwdriver and a screw, a close fit is important, and could be an important indication to a tool using agent that the two objects are associated. If the size of the ball were increased further, eventually the set of occupied regions inside the ball but not inside the hole becomes non-empty, indicating that the ball can no longer pass through the hole.

A void, such as an opening in a surface, may be represented as a filled area (as in the case of the mouse hole) or as an empty, bounded region (as in the case of the curved push tool (c)). Depending on the scenario, one approach may be more appropriate than the other, and it may be advantageous to support both approaches in *Canis*. Thus, the case in which the ball is compared to the push tool is similar to the case of the ball and hole, but here we are looking for an empty set of intersecting occupied regions, indicating that the tool's prongs can completely contain the ball. The push tool is a prototype tool that *Canis* can hold in its mouth and use to push a ball along in front of it as it walks. A close fit (which influences how well *Canis* can control the ball's motion with the tool) is indicated by the ball occupying the entire space within the span between the two prongs of the tool, with no overlap across the prongs.

A comparison of the cross-section of an Aibo (d) and the bounds of a box (e) gives an example of a way in which symmetry between an agent's body schema and elements in the environment could be used in navigation. The symmetry analysis shows that *Canis* can be contained within the box if oriented along the box's length, but not across its width. Also, while *Canis* could fit in the box, once inside, its motion would be tightly constrained. Further, if our analysis were extended to three dimensions, we might also be interested in

a comparison between the ball and the box. For example, if *Canis* were searching for the ball and encountered the box, it might compare the perceived dimensions of the box to a stored representation of the ball to determine whether the ball might possibly be hidden within the box.

To illustrate how this visual analysis fits into the larger context of tool use, consider the task of pushing a ball along a path over a slightly uneven surface, a task we have used for informal testing purposes. Robocup matches demonstrate that Aibos have some ability to dribble a ball; *Canis* has no such programming. In this example task, *Canis* simply moves forward, bumps the ball, and repeats. The plots on the right side of Figure 2 show points on the path of the ball for three representative trials of this task (i.e., points represent the centroid of the image of the ball over a sequence of time steps, on a Cartesian coordinate system overlaid on and oriented to the visual field.) The plots on the left show points on the path of the ball when *Canis* approaches it with the push tool in its mouth. A few differences are clear, and can easily be distinguished by conventional statistical tests. The ball is visible to *Canis* for a much greater proportion of time in trials with the tool; this can be seen in the greater number of data points in those trials. That is, the ball escapes *Canis*'s control more quickly in the trials without the tool. The ball remains closer to the center line of *Canis*'s field of view when pushed with the tool, and the standard deviation of its lateral position is much smaller than in trials without the tool.

Of course, this task is artificial in the extreme. For example, a square or triangular push tool cross-section would require a more relaxed evaluation of symmetry; the visual analysis here is simplified because the tool is held in a relatively constant position with respect to *Canis*'s camera; we have focused on specific, artificially constrained episodes, neglecting the larger structure of tool-using behavior (e.g., tool acquisition); and we have not discussed control or mo-

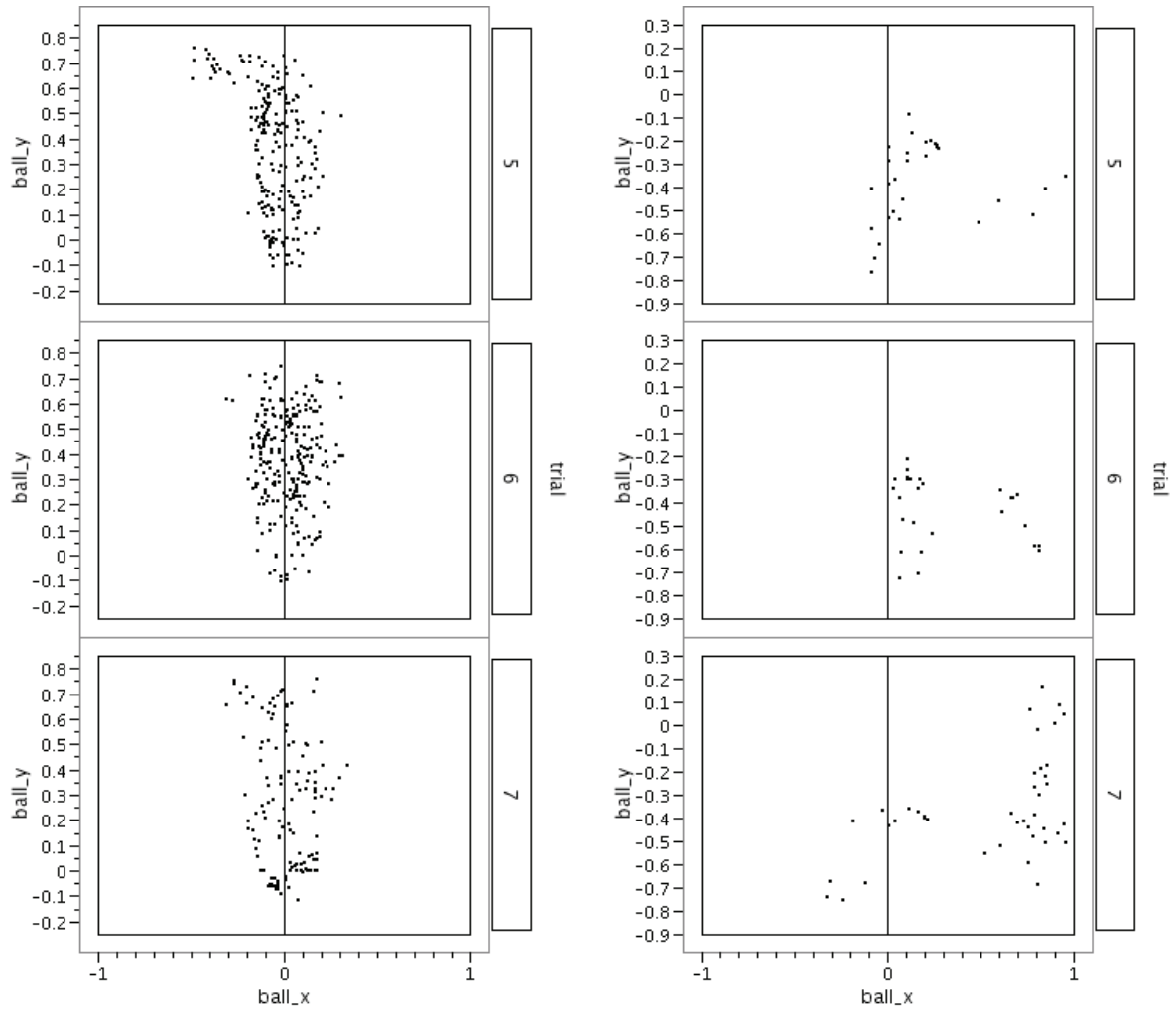


Figure 2: A ball-pushing task, with (left) and without (right) the curved push tool

tor issues (these are a separate part of our work). Other researchers have produced results that give insight into comparable tasks: Stoytchev's robot arm actually learns affordances, and it's possible to see Kemp and Edsinger's work with endpoint identification (2005) being applied to this example and giving interesting results. Nevertheless we believe that the example task provides at least a useful illustration of how visual analysis can produce results that allow an agent to identify patterns that might be used to guide interpretation or behavior appropriate for its environment.

Discussion

We hope to gain more from this approach to visual processing than simply the introduction of a layer of abstraction on top of standard image processing techniques.

For example, a related issue is problem solving based on computational imagery. It is easy to imagine an agent that can recognize symmetries not only in its immediate environment, but also across stored representations of environments encountered in the past or internally generated representations of environments that might be encountered in the future. For example, an agent's choice of action might be guided by recognizing a symmetry that cannot yet be seen in the environment, but that exists between the representation of a goal state and a tool in the physical environment that could be used to achieve the goal. Thus, an agent whose goal is a state in which a hole has been dug in the ground might recognize the symmetry between the blades of a nearby post-hole digger and the sides of the hole in the representation of its goal state. Such an ability ties in with concepts of mental imagery such as those described by Kosslyn (1994). Thus a tool-using agent may need to recognize affordances in stored representations of previous experiences to properly apply a tool in a new context, or in self-generated representations of potential states that have never been encountered before.

Another issue moves in the opposite direction with respect to cognitive processing. Specific symmetry comparisons might plausibly be hardcoded rather than computed more generally and more flexibly on the fly. We are not familiar enough with the literature on non-human vision to tell whether such comparisons, compiled into reflex behaviors, could play a role in models of other visual systems, but this is a possibility that remains for future research.

To summarize, our preliminary exploration of symmetry recognition suggests that the approach holds some promise. Although symmetries are currently matched only in 2D, we still see applications to interaction with a 3D environment. For example, 2D representations such as those in Figure 1 can serve to constrain searches in visual processing, which is an important concern in achieving effective interaction between embodied agents and their environment. A good deal of work remains, including a full implementation and evaluation of its performance; this work is in progress.

Acknowledgments

This effort was supported by the National Science Foundation under award IIS-0534398. The information in this paper does not necessarily reflect the position or policies of

the U.S. government, and no official endorsement should be inferred.

References

- Alcock, J. 1972. The evolution of the use of tools by feeding animals. *Evolution* 26:464–473.
- Beck, B. B. 1980. *Animal Tool Behavior: The Use and Manufacture of Tools*. New York, NY: Garland Press.
- Belongie, S.; Malik, J.; and Puzicha, J. 2002. Shape Matching and Object Recognition Using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 509–522.
- Bogoni, L., and Bajcsy, R. 1993. An active approach to characterization and recognition of functionality and functional properties. In *AAAI Workshop on Reasoning about Function*, 9–16.
- Bogoni, L., and Bajcsy, R. 1994. Functionality investigation using a discrete event system approach. *Journal of Robotics and Autonomous Systems* 13(3):173–196.
- Cairns, P., and Thimbleby, H. 2008. Affordance and Symmetry in User Interfaces. *The Computer Journal*. to appear.
- Chemero, A. 2003. An outline of a theory of affordances. *Ecological Psychology* 15(2):181–195.
- Şahin, E.; Çakmak, M.; Doğar, M. R.; Uğur, E.; and Üçoluk, G. 2007. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior* 15:447–472.
- Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Glasgow, J., and Papadias, D. 1992. Computational imagery. *Cognitive Science* 16(3):355–394.
- Jones, K. S. 2003. What is an affordance? *Ecological Psychology* 15(2):107–114.
- Kemp, C. C., and Edsinger, A. 2005. Visual tool tip detection and position estimation for robotic manipulation of unknown human tools. Technical Report AIM-2005-037, Massachusetts Institute of Technology, CSAIL.
- Kosslyn, S. M. 1994. *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, Cambridge, MA.
- Krützen, M.; Mann, J.; Heithaus, M. R.; Connor, R. C.; Belder, L.; and Sherwin, W. B. 2005. Cultural transmission of tool use in bottlenose dolphins. *Proceedings of the National Academy of Sciences* 102(25):8939–8943.
- Oswalt, W. H. 1973. *Habitat and Technology: The Evolution of Hunting*. New York, NY: Holt, Rinehart, & Winston.
- Povinelli, D. 2000. *Folk Physics for Apes*. NY: Oxford University Press.
- St. Amant, R., and Horton, T. E. 2008. Revisiting the definition of tool use. *Animal Behaviour* 74(4):1199–1208.
- St. Amant, R., and Wood, A. B. 2005. Tool use for autonomous agents. In *Proceedings of the National Conference on Artificial Intelligence*, 184–189. Menlo Park, CA: AAAI Press.

- St. Amant, R. 1999. User interface affordances in a planning representation. *Human Computer Interaction* 14(3):317–354.
- Stark, L., and Bowyer, K. 1996. *Generic Object Recognition using Form and Function*, volume 10 of *Series in Machine Perception and Artificial Intelligence*. World Scientific Press.
- Stoytchev, A. 2005. Behavior-grounded representation of tool affordances. In *IEEE International Conference on Robotics and Automation (ICRA)*.
- Thimbleby, H. 2001. Affordance and Symmetry. *Interactive Systems: Design, Specification, and Verification, 8th. International Workshop, DSV-IS*.
- Thimbleby, H. 2002. Reflections on symmetry. *Proceedings on Advanced Visual Interfaces, AVI2002* 28–33.
- Touretzky, D. S., and Tira-Thompson, E. J. 2005. Tekkotsu: A framework for aibo cognitive robotics. In *Proceedings of the Twentieth National Conference on Artificial Intelligence*. AAAI Press.
- van Lawick-Goodall, J. 1970. Tool-using in primates and other vertebrates. In Lehrman, D.; Hinde, R.; and Shaw, E., eds., *Advances in the Study of Behavior*, volume 3. New York: Academic Press. 195–249.
- van Schaik, C. P.; Ancrenaz, M.; Borgen, G.; Galdikas, B.; Knott, C. D.; Singleton, I.; Suzuki, A.; Utami, S. S.; and Merrill, M. 2003. Orangutan Cultures and the Evolution of Material Culture. *Science* 299(5603):102–105.
- Vera, A. H., and Simon, H. A. 1993. Situated action: A symbolic interpretation. *Cognitive Science* 17:7–48.
- Visalberghi, E., and Limongelli, L. 1996. Acting and understanding: Tool use revisited through the minds of capuchin monkeys. In Russon, A. E.; Bard, K. A.; and Parker, S. T., eds., *Reaching into Thought: The minds of the great apes*. Cambridge, UK: Cambridge University Press. 57–79.
- Weyl, H. 1983. *Symmetry*. Princeton University Press.
- Whiten, A.; Goodall, J.; McGrew, W. C.; Nishida, T.; Reynolds, V.; Sugiyama, Y.; Tutin, C. E. G.; Wrangham, R. W.; and Boesch, C. 1999. Cultures in chimpanzees. *Nature* 399:682–685.
- Wood, A. B.; Horton, T. E.; and St. Amant, R. 2005. Effective tool use in a habile agent. In *Proceedings of the IEEE Systems and Information Engineering Design Symposium*.