Symbolic approach to affordances in SGOMS, a paper for the Affordances in Vision for the Cognitive Robotics Workshop

Sterling Somers, Carleton University, Canada

Abstract—**This paper discusses the use of a symbolic cognitive architecture for symbolically representing affordances. Initial models of counter-steer affordances in which symbolically representing affordances facilitates context sensitivity are discussed. This paper proposes the use of geometric affordances in which body schema are matched against the geometric properties of the environment to facilitate action decisions.**

Keywords-symbolic; affordances; cognitive modeling; ACT-R; SGOMS; GOMS;

I. INTRODUCTION

Due to the theoretical ties with the concept of *direct perception,* affordances are, prima facie, non-symbolic, non-representational, and non-inferential [1], [2]. References [3], [4] are among the first authors to use an affordance approach in robotics. Despite their claim to the opposite, I agree with the suggestion by [2] that their approach is representational. Once the robot has been trained to associate certain feature vectors with certain "affordances", the feature vector mappings are, strictly speaking, representational. References [3], [4] attempt to maintain Gibson's original notion of affordances but their redefinition to include representations is seen as decoupling the agent from its environment.

The robot presented in [3], [4] associates feature vectors with traversability-related action capabilities. Reference [4] assesses the robot for ramp, gap, and aperture passage. The presented robot learns the traversability affordances well in a simple environment, with only a slight drop in performance where novel, complex objects are introduced. That said, the changes made within the complex environments were simple changes such as object orientation or aperture height adjustments. While [2] suggests that the robot presented in [3], [4] should not implement representations, the present work proposes that not only are the current representations appropriate, but that an advantage might be gained by adding *body schema* as an additional representation.

This article presents an implementation of what I term, *geometric affordances* – traversability affordances that are triggered by the geometric properties of space between objects. The use of symbolically representing these affordances as part of a symbolic context, used within the SGOMS [5] (Socio-technical GOMS) (Goals, Operators, Methods, and Selection rules) [6] modeling framework as implemented within ACT-R

(Adaptive Control of Though—Rational) [7] will be described hereto.

II. ACT-R AND SGOMS

A. ACT-R

ACT-R [7] is a cognitive architecture used by scientists to model the mind. ACT-R is typically used to implement models of humans performing psychology experiments as a means of validating theories about the functional capabilities of the brain. The ACT-R architecture consists of several modules - meant to represent the theorized modules of the mind (memory, perceptual systems, a motor module, and a central production system) - that communicate with one another, via the central production system, through the use of buffers. These buffers facilitate communication through the use of *chunks*, slot/value pairs representing properties. ACT-R is, therefore, a symbolic architecture. The central production system matches on the contents of the various buffers each production cycle, resulting in actions which could include: changing the contents of buffers, motor actions, and memory read/write requests. ACT-R also employs a sub-symbolic system to help account for human processing and memory limitations. While ACT-R is well-suited for building models of humans performing tightly-constrained psychology experiments, it is not clear how well it is suited for complex tasks that involve cooperation, communication, complex planning, and interruption.

B. SGOMS

SGOMS [5] is framework for implementing a hierarchical control structure within ACT-R to address agent performance in *macro*-cognitive tasks. Hierarchical control in SGOMS is achieved through the implementation of three levels: the planning unit, the unit task, and the operators.

At the lowest level, the operators represent low-level actions such as memory requests or brief motor commands. As demonstrated in [8], SGOMS planning units facilitate, what might be considered, supervised vision-action loops where perceptual systems and motor systems work in a semi-autonomous fashion. Since the central production system in ACT-R is intended to fire only one time in any production per production cycle, the implementation of an operator level reduces the strain on the system's ability to plan and re-plan, facilitating a (closer-to) real-time perception/action cycle.

Operators are implemented in ACT-R through the use of an operator buffer. The contents of the operator buffer is set top-down from unit tasks. Because higher-levels have access to the operator buffer, the operators are only semi-autonomous as the contents of the buffers can be changed top-down resulting in a change of low-level instructions (as a result of implementing a new task).

The unit task is the second level of control in SGOMS. Unit tasks are meant to link operators into coherent groups (e.g. complex movements, movements that often occur together). Specifically unit tasks represent groups of low-level tasks that are expected to occur without interruption or expected to be interrupted in predictable ways that are easily managed by the system. Like operators, unit tasks have their own buffers, but unlike operators are not autonomous to any degree as they occur entirely within the production cycle.

Finally, the highest level of control is the planning unit. The planning unit is the slowest level of control and, as the name suggests, consists of a group of unit tasks, therefore functioning as a plan. Similar to the above layers, the planning unit has a dedicated buffer. Collectively the buffers represent what would be, in non-SGOMS ACT-R, the agent's goal. The mapping between planning unit and unit tasks is modulated by the overall context such that any given plan can have different branching solutions, depending on what the context presents. The overall structure supports both ordered/successive unit tasks or unordered unit tasks.

III. PREVIOUS WORK

While SGOMS has been applied to a number of domains, the domain of interest to the present audience is that of affordances in steering control of a flight simulator. The model by [8], [9] is an affordance inspired model which uses optical flow to modulate counter-steer. Relevant for the remainder of this paper is the model's use of symbols to represent the optical flow information. By representing the optical flow information symbolically, the overall system can make actions decisions accordingly.

The model presented in [8], [9] is a model of a pilot reacting to a semi-expected left-hand veer often experienced by small, single engine aircraft during take-off. The tendency to turn left is usually caused by one or more of the following factors: torque steer, spiraling slipstream, asymmetrical loading (p-factor), or gyroscopic procession. This left-hand veer is semi-expected because although a reasonably educated pilot can expect it to happen most of the time, when it occurs, and the degree to which it occurs is a result of physical factors unmeasurable to the pilot. The pilot therefore can only react to the situation as opposed to implementing some static plan.

The model uses a simple motion detection vision system inspired by the retinal flow affordance work in [10] who found that retinal flow information is sufficient for steering control. The vision system described in [8], [9] was implemented using Python bindings to OpenCV using screen captures from the monitor as input. Movement was detected when a least-squared-difference calculation detected a group of pixels in a new location (indicating movement). In the work described in [8], lateral movement was represented as the either 'left' or

'right' indicating left- or right-hand movement. In [9], 'normal' and 'extreme' changes in rotation are also represented, leading either to normal response (counter-steer) or an abrupt abort (respectively).

Counter-steer was accomplished in [8] dynamically with an increase in rudder control opposing either left- or right-hand motion. In the context of SGOMS, the agent's main task was to take-off, carrying out planning units and unit tasks as would be expected in a normal take-off. The appearance of bottom-up visual information indicated rotation (left/right) results in a re-plan to respond to the more pressing need of controlling the steer of the aircraft. At the point of interruption a semi-closed loop between the vision and action systems occur until the symbol 'left' (or 'right') no longer appear in the visual buffers, indicating the plane is tracking straight. Once the plane is tracking straight, the 'normal' take-off hierarchy is resumed. This interruption and resumption may occur multiple times during a single take-off procedure.

Important for the present discussion is the model's ability to naturally switch between tasks using the SGOMS framework. Contradictory to the advocates of *direct perception*, symbolically representing affordances seems to, at least, be sufficient for controlling context. In [9], for example, direction and degree were both represented and the simulated pilot could make decisions of whether to counter-steer as normal or to take emergency action (as the strength of the rotation, in some situations, may be unsafe to try to recover using a standard counter-steer technique). While the models [8], [9] by no means prove that symbolically representing affordances is a necessary condition of exploiting affordances, it is unclear how a model could effectively account for different contexts.

IV. CURRENT WORK

While current work is primarily intended for a cognitive science audience, the current workshop encourages interdisciplinary discussion for an affordance-based robotics audience. The affordance work under investigation is of more controlled studies of aperture passage, particularly [11]. In this study and related studies [12]–[17] the authors identify various features of the animal-environment system (usually purported to be directly perceived), which modulate affordance passage. The research findings in this area and a brief discussion about the relation to the concept of *geometric* affordances is discussed below.

A. Geometric Affordances

The experiment in [11] involved participants walking through apertures, rotating their shoulders as needed. Having participants in different height groups, the authors manipulated the aperture width and concluded that degree of shoulder rotation is modulated by eye-height. In [12] eye-height, head-sway, and stride-length are implicated in aperture traversability judgments. In both [11], [12] the features are said to be "calibrated" to shoulder width in such a way that aperture passage is directly perceived. While these two studies are committed to the Gibsonean notion of direct perception, studies [13]–[17] are somewhat less stringent.

For example, in [13] participants were grouped in a dyad before making aperture passage judgments. In [16], aperture traversability is judged when participants are carrying objects, including cases where the object extended the total width of the person (the object was wider than the participants shoulders). The authors in neither [13] nor [16] discuss how any feature of the environment could be "calibrated" to the agent. Of course, width is not the only feature of an aperture. In [14], participants are asked to judge traversability *under* a barrier. In these experiments, participants made traversability judgments for height both in simple conditions (without augmentation) and augmented conditions in which the participants wore either blocks on their feet (affecting apparent eye-height) or helmets on their head (affecting the accuracy of eye-height as a guide).

Although out of scope of the present discussion, it is questionable how the affordances can be directly perceived in these cases because affordance is not simply between the agent and the environment but, rather, manifest in a system that includes an agent and some other object (another human, an object carried). Intuitively, it cannot be that the affordance is perceived directly (in the Gibsonean sense) because the action capabilities are not the capabilities of the singular animal.

Taken together, however, [11]–[17] are suggestive of a capacity to extract the geometric properties (e.g. width & height), of an aperture and make traversability judgments by making a comparison to a representation of the estimated size of the system passing through (a single agent, an agent plus another object, or an agent in a dyad). The *geometric affordance* hypothesis proposes that there is a cognitive mechanism used for comparing geometric properties of the environment with the geometric properties of the agent (or agent-plus-object) system and that the same mechanism is used regardless of whether width or height is being estimated. Such a mechanism may use information such as eye-height, head-sway, or stride-length to make perceptual estimations, but ultimately, the output of the processes responsible for detecting these features is used as input for the comparison mechanism.

B. ACT-R Implementation

The proposal being presented is to implement a geometric affordance system in ACT-R by making use of ACT-R's declarative memory system as well as the structure of SGOMs to attend to appropriate contextual details. There are two aspects to aperture traversability judgments: a yes/no judgment that decides whether there is any body posture that would suit traversability; and action-oriented portion in which a current body posture must undergo transformation (e.g. shoulder rotations, ducking) in order to achieve as suitable posture. The following will be a brief outline of how the system will accomplish both the action decision and performance at a high-level in ACT-R.

One of the main features in the proposed model is to implement *body schema* in ACT-R's declarative memory. Body schema are intended to be chunk-based representation of possible body postures as well as their associated (approximate) geometries. While in the model these body schema will be static, it is expected that these schema can be created over time through proprioception and haptic feedback (feedback about your body posture in conjunction with feedback about the extensions of your body such as may occur when you bump into things), or through visual estimation (simply estimating the size of an object).

The model will be developed in the Python version of ACT-R [18] and will use MORSE (Modular Open Robot Simulation Engine) [19] as the simulation environment. Because the focus of the model is to build a theory of cognitive processing, at first pass the details of low-level vision will be largely ignored. The model assumes that the geometric properties of an aperture is a reasonable output of low-level vision processes and will focus instead on the cognitive processes supporting action-judgments and performance. Currently under development is a version of MORSE's semantic camera which can directly retrieve objects within a camera's field of vision. This version of MORSE's semantic camera will output geometry properties of the environment (e.g. distances to objects, distances between objects) in the "chunk" format appropriate for ACT-R. The visual chunks can be matched to body schema in the declarative memory. A failure to match results in a chunk representing the failure, which can be used by the model to make a negative judgment. A match will result in the body schema chunk being represented in the internal context of the agent (as an action possibility).

In this model, one can think of the body schema chunks as representing the context match-able by unit tasks. The unit tasks, in cases like these, would act like mappings between body postures. Furthermore, unit tasks also act like attentional filters. For example, in a cluttered environment it would not be long before multiple body schemas are activated bottom-up (simply as a result of the perceptual system). However, only the unit tasks relevant to the current plan will match. Even in the case where multiple unit tasks match both the present context and exist within the same planning unit, ACT-R has built-in conflict resolution, including a utility system which chooses productions (in this case at the level of unit tasks). Utility is one of the learning mechanisms in ACT-R for choosing certain productions over others based on their success history. It is at least plausible that to use the built-in utility system of ACT-R can be used to learn affordances.

V. CONCLUSION

The fields of cognitive robotics and computational psychology are tightly linked. In computational psychology many assumptions about both perception and action are simply not dealt with (or, at least, are poorly dealt with). There is a significant amount of work to be done to create cognitive models (which make psychological and neurological predictions) of complex behavior. Work in affordances has intriguing applications in robotics yet due to ties with the notion of *direct perception* it has an unclear position in computation. This paper presented ideas on how to use the affordance research within a symbolic paradigm.

R<small>EFERENCES</small>

[1] J. J. Gibson, *The ecological approach to visual perception*. Hillsdale, NJ: Erlb, 1986.

[2] A. Chemero and M. T. Turvey, "Gibsonian Affordances for Roboticists," *Adapt. Behav.*, vol. 15, no. 4, pp. 473–480, Dec. 2007.

[3] E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk, "To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control," *Adapt. Behav.*, vol. 15, no. 4, pp. 447–472, Dec. 2007.

[4] E. Ugur and E. Sahin, "Traversability: A Case Study for Learning and Perceiving Affordances in Robots," *Adapt. Behav.*, vol. 18, no. 3–4, pp. 258–284, Jul. 2010.

[5] R. L. West and S. Somers, "Scaling up from Micro Cognition to Macro Cognition : Using SGOMS to build Macro Cognitive Models of Sociotechnical Work in ACT-R," in *Proceedings of the 33rd Annual Cognitive Science Society*, 2011, pp. 1788–1793.

[6] S. K. Card, T. P. Moran, and A. Newell, "The keystroke level model for user performance with interactive systems," *Commun. ACM1*, vol. 23, pp. 396–410, 1983.

[7] J. R. Anderson, *How can the human mind occur in the physical universe?* New York, New York, USA: Oxford University Press, 2007.

[8] S. Somers and R. West, "Steering Control in a Flight Simulator Using ACT-R," in *Proceedings of the 12th International Conference on Cognitive MOdelling*, 2013.

[9] S. Somers and R. L. West, "Macro Cognition: Using SGOMS to Pilot a Flight Simulator," in *19th Annual ACT-R Workshop*, 2012.

[10] L. Li and W. H. Warren, "Retinal flow is sufficient for steering during observer rotation.," *Psychol. Sci.*, vol. 13, no. 5, pp. 485–91, Sep. 2002.

[11] W. H. Warren and S. Whang, "Visual guidance of walking through apertures: body-scaled information for affordances.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 13, no. 3, pp. 371–83, Aug. 1987.

[12] A. J. Fath and B. R. Fajen, "Static and dynamic visual information about the size and passability of an aperture," *Perception*, vol. 40, no. 8, pp. 887–904, 2011.

[13] C.-H. Chang, M. G. Wade, and T. A. Stoffregen, "Perceiving affordances for aperture passage in an environment-person-person system.," *J. Mot. Behav.*, vol. 41, no. 6, pp. 495–500, Nov. 2009.

[14] J. K. Stefanucci and M. N. Geuss, "Duck! Scaling the height of a horizontal barrier to body height.," *Atten. Percept. Psychophys.*, vol. 72, no. 5, pp. 1338–49, Jul. 2010.

[15] B. R. Fajen and J. S. Matthis, "Direct perception of action-scaled affordances: the shrinking gap problem.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 37, no. 5, pp. 1442–57, Oct. 2011.

[16] J. B. Wagman and K. R. Taylor, "Perceiving Affordances for Aperture Crossing for the Person-Plus-Object System," *Ecol. Psychol.*, vol. 17, no. 2, pp. 105–130, Apr. 2005.

[17] J. B. Wagman and E. A. Malek, "Perception of Whether an Object Can Be Carried Through an Aperture Depends on Anticipated Speed," *Exp. Psychol. (formerly "Zeitschrift für Exp. Psychol.*, vol. 54, no. 1, pp. 54–61, Jan. 2007.

[18] T. C. Stewart and R. L. West, "Python ACT-R: A New Implementation and a New Syntax.," in *12th Annual ACT-R Workshop*, 2005.

[19] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan, "Modular open robots simulation engine: MORSE," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 46–51.