Creating Affording Situations with Animate Objects

Chris Baber(✉), Sara Al-Tunaib, and Ahmed Khattab

University of Birmingham, Birmingham B15 2TT, UK
c.baber@bham.ac.uk

Abstract. In this paper, we report the design, development and initial evaluation of the concept of animate objects (as a form of Tangible User Interface, TUI) to support the cueing of action sequences. Animate objects support user actions through the provision of affordances, and developments in actuator technologies allow those devices to change their physical form. Each object communicates (in a small-scale 'Internet of Things') to cue other objects to act, and these, in turn, provide cues to the user. The set of cues creates affording situations which can be understood in terms of Forms of Engagement. A user trial was conducted in which participants had to either perform a simulated tea-making task or search for letters hidden under objects to spell words. The first task involved a familiar sequence (and we also included an unfamiliar sequence), and the second task involved either real or false words. We show that, when cues provided by the objects correspond to a 'logical' task sequence (i.e., familiar and/or easily interpretable), they aid performance and when these cues relate to 'illogical' task sequences, performance is impaired. We explain these results in terms of forms of engagement and suggest applications for rehabilitation.

Keywords: Tangible User Interface · Animate objects · Affordance

1 Introduction

1.1 Tangible User Interfaces

Tangible User Interfaces (TUIs) employ physical objects to collect data from, or display information to, users [1–3]. With the decreasing cost of sensors, actuators and processors, and access to 3D printing, it is easy to design and build all manner of things that have the appearance of familiar objects combined with the capability to sense and respond to user activity, and communicate with other objects. The networking of objects, the recognition of user activity and prediction of user intent, and the use of intent prediction to cue specific actions create new challenges for Ergonomics. In a previous paper, we described the underlying concepts relating to activity and intent, and the use of the Blynk protocol for managing networked communications between smart objects [4]. We also, in that paper, reported development of an initial set of objects and conducted a user trial which showed that, when the handle on a jug rose

D. D. Schmorrow and C. M. Fidopiastis (Eds.): HCII 2019, LNAI 11580, pp. 3–13, 2019.
https://doi.org/10.1007/978-3-030-22419-6_1

automatically, most (22/23 participants in the trial) people would use the hand adjacent to the handle to pick it up rather than their dominant hand – even if, during interview after the experiment, they did not notice that they had done this. We believe that this finding implied that their behaviour was influenced by that of the object, and that the influence could well be occurring at a pre-conscious level. Encouraged by this, we have developed more objects and an experimental protocol that allows us to explore such interactions.

In this paper, familiar objects are fitted with sensors (to support ‘awareness and monitoring’) and with actuators and other means of display to provide cues to a user. Of particular interest are questions relating to the interpretation of an appropriate action in response to the cues presented by the object. So, if one is performing a familiar task, will the cues be irritating or distracting? If the cues are erroneous (in that they cue a sequence that is different to the one that you intended), will you continue to respond to the cues? In order to hypothesize why users might either ignore cues or respond erroneously, we propose that the interactions between users and smart objects can be considered in terms of affording situations.

1.2 Affording Situations

Gibson [5] introduced the term affordance into psychology, suggesting that we perceive the world in terms of opportunities for action. What an object affords is determined by the physical properties of the objects (e.g., shape, orientation, size), by the action capabilities of the agent, and by the intention that the use of this object will support. This means that ‘affordance’ cannot simply be a property of an object (so it does not make sense to simply state that a ‘cup affords drinking’). Rather, one needs to be situate the use of the object in the context of an ongoing goal-driven activity being performed by an individual with sufficient ability to use *that* object to achieve *that* goal. Thus, an affordance is the relationship between an individual’s ability to act and the opportunities provided to that person in the given situation in pursuit of a given goal. That is, a cup of particular dimensions can be grasped by a person with particular physical abilities (e.g., hand size, motor skills etc.) in the context of performing a task with a particular goal: a person laying the table will pick up the cup differently than a person who intends to drink from it; picking up a cup that is full to the brim will be different from picking up a half empty cup. These simple observations lead to the proposal that affordance emerges from the interactions between person and object in a given environment and in pursuit of a given goal and this relationship is captured by the idea of Forms of Engagement, illustrated by Fig. 1 and developed in [6–8].

In Fig. 1, interaction (between person and object) involves several Forms of Engagement. Responding to the specific features of an object in an environment in order to perform an action defines the ‘affordance’ (indicated by a dotted line around *Environmental* and *Motor Engagement).* This results in a change of state of the object, which the person attends to through *Perceptual Engagement* (i.e., interpreting visual, tactile, kinaesthetic, auditory etc. information as feedback from the performance of an action). So, this proposes two forms of perception by the person: in *environmental* *engagement* the person responds to features of the object that can support actions (i.e., perception-action coupling) and in *perceptual engagement*, the person responds to the changing state of the object (and environment) as an action is performed. As the object (and environment) change, so this can produce new opportunities for *environmental engagement*. In order to act on an object, there is a need to respond to the 'information' (in a Gibsonian sense) that it conveys. In other words, people are able to 'see' aspects of the object, or the environment, in terms of an action that they want, and are able, to perform in pursuit of a goal. We further separate *Motor Engagement* (the ability to perform an action) from *Morphological Engagement*, the disposition of the person, e.g., in terms of the size of the hand. We have a two-way link between these because hand shaping will be influenced by subsequent actions.

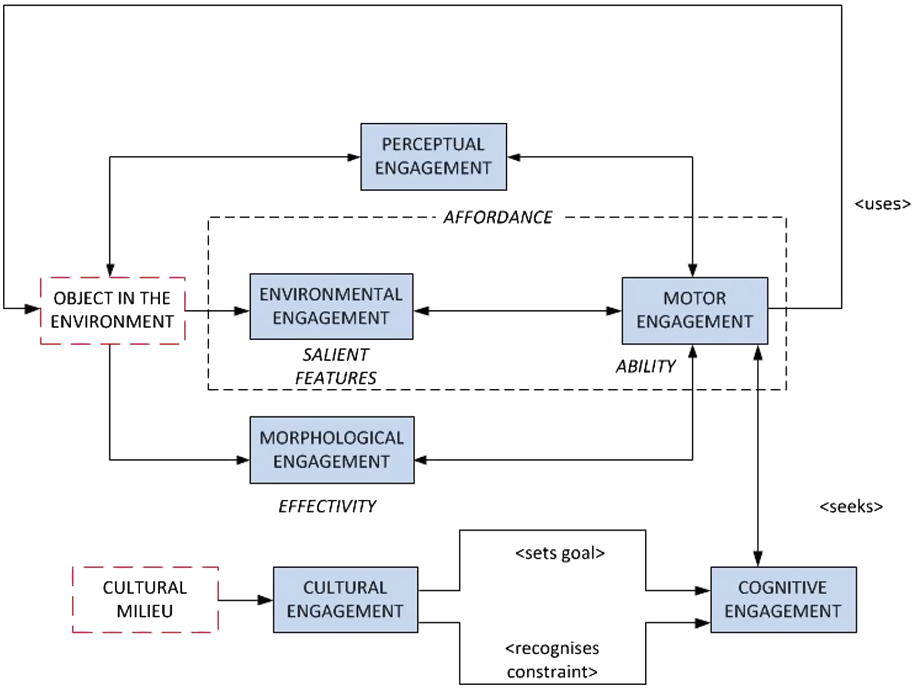

Fig. 1. Forms of engagement [9]

The role of *Cognitive Engagement*, in this description, is two-fold. First, it provides high-level management on ongoing actions (checking for lapses, slips, mistakes etc.) and second, to manage the actions in terms of an overarching goal (e.g., in terms of the Anticipative Actions of the Adaptive Control Model). Of particular interest to this paper, is the extent to which these goals might not be fully-defined prior to performing an action but might evolve in response to the opportunities presented by the objects.

Finally, the notion of an 'acceptable' goal (or 'acceptable' actions to perform on objects) could relate to the culture in which one is acting and is characterized as *Cultural Engagement*.

1.3 Specifying Animate Objects

Having proposed that interaction comprises Forms of Engagement, one can relate these to the possible inferences that animate objects could make (about user intention) as they are being interacted with. At the most basic level, sensors on the object could provide data to characterize the motion, orientation, position, etc. of the object. These data would define the *motor engagement* with which the person was interacting with the object. The object, assuming that it can modify its appearance, could encourage *Environmental Engagement* through changes that emphasize specific features. So, when a handle rises on the side of a cup, people are more likely to use the hand on that side of the cup to pick it up [9]. Thus, the action that is selected in anticipation of motor engagement, could be cued by the appearance of the object, and the *morphological engagement* that is necessary for the action would correspond to the physical appearance of the object. Having some knowledge of where the object is being used could also influence the definition of appropriate actions, through *Cultural Engagement*, e.g., one might anticipate that drinking from a tea cup in the Savoy Hotel is not identical to drinking from a tea cup in one's kitchen at home. Combining inferences drawn from Motor and Morphological Engagement, the object could infer the most likely goal of the person, and use this inference to provide additional cues and guidance of action [10]. A mapping of Forms of Engagement to the cues provided by animate objects in shown in Table 1.

Table 1. Mapping of forms of engagement to cues from animate objects.

Form of engagement	Cues
Environmental	Spatial layout of objects; form of objects
Morphological	Grasp permitted by object
Motor	Manipulation supported by object
Perceptual	State and appearance of object
Cognitive	Intention (of user or 'system')
Cultural	Conventions governing object manipulation

Let us assume that the Animate Object looks like something familiar which has been fitted with sensors [9]. For example, Fig. 2 shows a jug with its sensor unit (developed for the CogWatch project[1]).

[1] https://www.birmingham.ac.uk/research/impact/original/cogwatch.aspx.

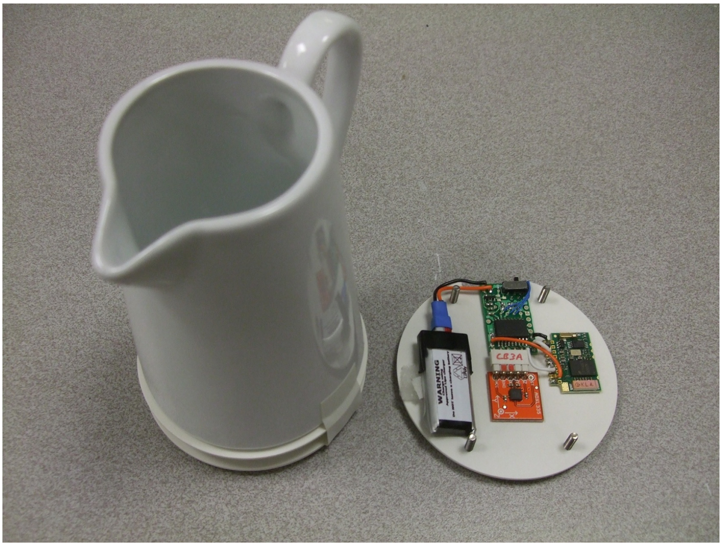

Fig. 2. Jug and sensor unit

On the one hand, a jug is an object that we 'know' how to use, but on the other hand, this is an alien object that is capable to doing things that we do not, necessarily, fully understand. The jug could, for example, be part of a system that monitors our daily cream intake and the system could have a 'intention' of ensuring that we drink a specified quantity of liquid, or it might be part of a system that has the 'intention' of reducing our caffeine intake (using the pouring of cream as a proxy for drinking coffee). One way in which such 'intentions' could be communicated to the user would via the objects themselves (through lights, sounds, movement etc.), providing feedback and cues to the person. In this way, the form of the objects could display their function and we can create 'affording situations' in which the appropriate action is cued by the objects that the person needs to use.

Relating this to Forms of Engagement and the cues provided by Animate Objects, we could suggest that the motion of an object (part or whole) would be compatible with a movement to be made by the person, e.g., if the handle on the object moved, then one might expect the person to move their hand to that handle. Alternatively, a light on the object turned on (or the object made a noise), one would expect the person's attention to turn to that object. From this, Cue Compatibility draws on environmental and perceptual Forms of Engagement, and can influence Morphological and Motor Engagement. Cognitive Engagement relates to selecting which action to perform, and Cultural Engagement constrains selection of action in terms of 'normal' behavior. From this one can propose a set of tasks that reflect the relationship between the Task that the person ought to perform and the State of the Object (Table 2).

Table 2. Mapping user action to object state

Action	State (object after action)	Pre-condition (object prior to action)
Pick up	Raised	*Jug.Lowered = true ^ Jug.Handle = up ^ Jug.LED = on ^ Jug.Audio = 'pick_me_up'*
Put down	Lowered	*Jug.Raised = true*
Move	Location changed	*Jug.Raised = true*
Pour	Object tilted	*Jug.Raised = true ^ Jug.Location(proximal.cup) = true*

1.4 Constructing Animate Objects

The objects used in this study were designed and 3D printed to look like familiar, everyday objects, e.g., mugs, jugs, kettle, drawer handle, spoon handle. They were designed to incorporate electronics, i.e., sensors, a microprocessor (Arduino), Wi-Fi communications. Communications were managed using the Blynk Bridge Application, which allowed each object to communicate on a network and also allowed an Apple iPhone to be used to send commands to each object and log their status. Figure 3 shows some of the objects used in this study. Each object is powered by a rechargeable power pack and all objects have multi-colour Light Emitting Diode (LED) strips. Commands (from the iPhone or from other objects via the Blynk Bridge network) can turn the LEDs on or off, or change the colour of LED on the objects. Some objects had small motors that could vibrate the object (e.g., the spoon) or lower and raise the handles (e.g., jug, kettle), and small audio chip to replay recorded voice messages.

Fig. 3. Animate objects arranged for the 'tea-making' task

2 User Trial

The aim of the user trial was to explore the interaction between cues offered by Animate Objects and actions provided by users. A basic question was whether cueing from Animate Objects could over-ride participants' expectations of familiar tasks. So, participants were asked to perform two tasks. That is, whether the top-down control of action implied by cognitive engagement could be over-ridden by environmental engagement. In the first task ('simulated tea-making') we assumed that participants would have a well-learned 'script' that defined a sequence of tasks. In this case, the cues from the Objects should have little impact (other than, perhaps, slowing the participants as the waited for the next cue). Conversely, if the Objects cued a sequence that was unfamiliar or illogical, participants would be likely to ignore the cues and rely on their prior knowledge. In a second task ('letter finding'), participants could not rely on prior knowledge. In this case, individual letters were hidden under the Objects and participants had to lift each Object to find the letter in order to spell a 3-letter word. In this case, good performance would involve lifting only correct Objects.

2.1 Participants

Twenty participants (Female = 7, Male = 13, age = 26±, 3.2 years) were recruited from Undergraduate students in the School of Electronic, Electrical and Systems Engineering. None of those recruited had any previous experience interacting with the previous project.

2.2 Procedure

The design of evaluation trials was approved by the School of Engineering, University of Birmingham ethics committee and followed the Declaration of Helsinki. Once they had signed an informed consent form, each participant was shown a printed copy of instructions, to be read prior to commencing their experiments. The instructions were removed from the participants after they had been read and understood. This was done to create a situation to resemble natural interaction with objects.
Three restrictions were imposed on the participants:

1. There was an upper time limit of one minute
2. Only the use of one hand was permitted.
3. Participants were asked to imagine there was water in the jugs in the tea making task.

Two sets of tasks were used. The first set relied on participants being able to define a goal and use this to complete a familiar task sequence. We used simulated tea-making, in which participants acted out the tasks involved in making tea. In the second set, we required participants to learn an unfamiliar sequence. In this case, letters were hidden under each object and participants had to lift the object to find the letters and then spell out a 3-letter word. In this condition, one sequence of cues resulted in the real word 'TIN' and another sequence resulted in the false word 'TEI'. The order in which participants performed these conditions were counter-balanced to minimise learning effects.

Prior to each task, participants placed both hands on the table in front of the Objects. In the no cue condition (for ‘simulated tea making’), participants were asked to interact with the Objects as if they were making a cup of tea. In the cue conditions (for both tasks), Object states were controlled by the Experimenter according to a defined script. This allowed motor, audio and LEDs to be turned on or off, so that the State of the Objects changed to provide cues to the participants. The Dependent Variable in the experiment was Total Task Time, which began when one or both hands were lifted and ended when the participant declared that the task was complete.

2.3 Results

The corpus of data is collected from 320 trials, together with 10.5 h of video (which was analyzed for timing of actions; 2 analysts viewed the video and agreed ($r = 0.92$) on timings), and 40 feedback forms completed by participants. The results are organized by the dependent variables: total time taken and recall of LED colors.

As there were three conditions in which tea-making was performed, a one-way Analysis of Variance (no cues x familiar x unfamiliar) was run. This showed a significant main effect [$F(2, 59) = 8.79$, $p = 0.0005$]. Subsequent pairwise comparison, using paired t-tests, showed a difference between no cues and familiar [$t(19) = 2.23$, $p < 0.05$], between no cues and unfamiliar [$t(19) = 3.6$, $p < 0.005$], and between and familiar and unfamiliar [$t(19) = 5.8$, $p < 0.0001$]. This is illustrated by Fig. 4.

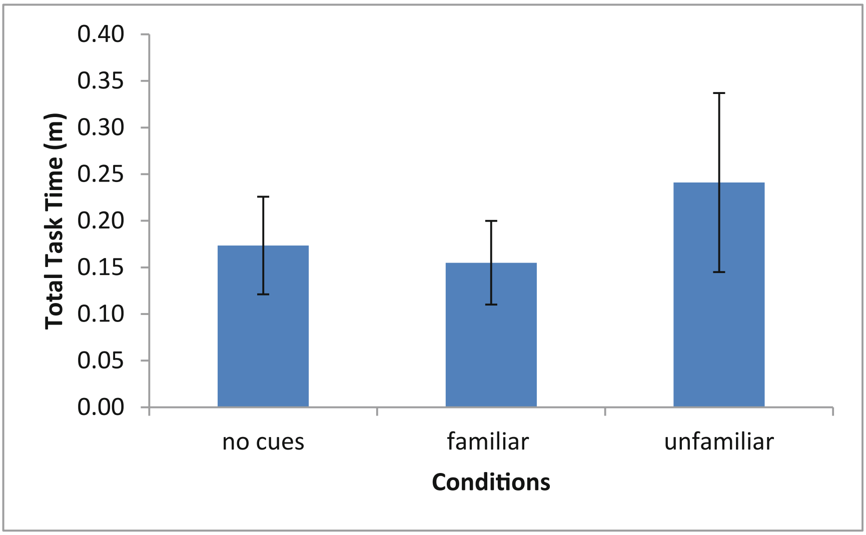

Fig. 4. Average time to complete tea-making tasks

For the spelling task, there were two conditions (real word and false word). A paired t-test showed a significant difference [$t(19) = 2.65$, $p = 0.05$]. This is illustrated by Fig. 5.

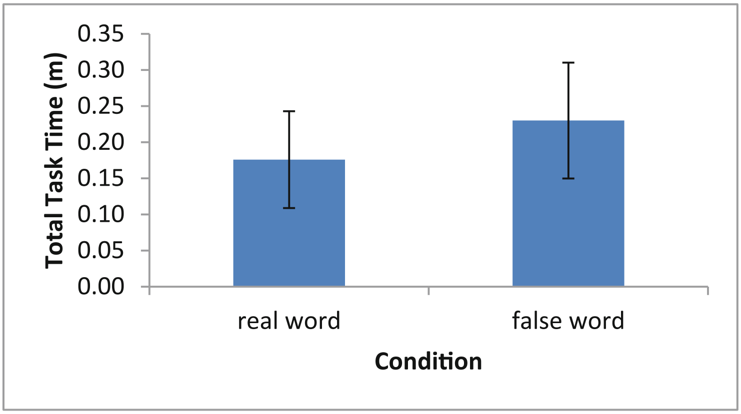

Fig. 5. Average time to complete 'spelling' task

3 Discussion

As might be expected, participants took longer to complete the tea-making task when they were prompted to follow an unfamiliar sequence. More interestingly, participants were faster when they were prompted (in the tea-making task) than when there were no cues. This suggests that, even with familiar task sequences, the provision of cues could be beneficial (at least, in terms of task completion time). For the 'spelling' task, performance on the 'real word' was faster than on the 'false word' – as if participants were trying to align the sequence with a goal but struggling to determine what this might be.

From observation, and discussion with participants during debrief, the use of cues in the unfamiliar tea-making sequence or for the false word task caused participants to interrupt the task sequence or led to them becoming confused. They might, for example, ask for clarification, e.g., 'is it meant to do that?' before continuing with the task. Thus, in the tea-making task, the vibrating spoon was a distraction from the intended sequence but led to a pause while the participant noticed it and then decided whether or not to respond.

3.1 Cues for Action

The cues that were employed in the experiment (audio, LED, moving parts) were attended to and commented upon by participants. Whether or not the cues were responded to depended on when the cues occurred and how the participant would incorporate this into the task sequence that was being followed. If the cue accorded with a logical sequence, it was followed, but if not then it was questioned (or, at least, responded to more slowly).

In the 'spelling' task (in which there might not be so clear a script to define the logical sequence), the cues were more likely to be followed. However, even in this task, participants would seek to make sense of what sequence the cues were guiding them towards. Further, the LEDs were not the only source of information used. Several participants mentioned that the position of objects (for the 'spelling' task) was felt to be important (even though the position of the objects was randomized across trials and so did not provide such information). We infer from this that the association of LED to object to letter was not made as consistently as we had expected, particularly if participants regarded the task as one of managing the spatial arrangement of objects. On reflection, if the only reliable source of information available is spatial arrangement (because they are not certain what the LEDs are telling them), then this strategy makes some sense.

3.2 Forms of Engagement

The role of animate objects in user activity varied across the different tasks. Thus, it is incorrect to assume that the cues would always provoke a response, but equally wrong to assume that these are over-ridden by prior knowledge. More interestingly, it appears that the cues were more beneficial when they support a logical (or familiar) sequence of tasks, and exacerbated confusion on illogical (or unfamiliar) sequences. According to the notion of Forms of Engagement, environmental engagement (in terms of the state of the object) can guide a sequence of actions, with perceptual engagement (as a way of interpreting the state of objects) being mediated by cognitive engagement (as a way of defining a logical sequence) but only when the sequence begins to feel wrong.

3.3 Affording Situations

The notion of 'affording situations' is intended to highlight the importance of context in understanding affordance; it is not simply a matter of saying a 'jug affords lifting or pouring'. While these actions are, of course, possible with a jug, in order to know that this particular jug *could* be lifted or poured at this moment in time, one needs to know the capabilities of the users. Furthermore, in order to know whether either lifting or pouring is appropriate, one also needs to know the intention of the person (or the intention that is plausible for the predicted task sequence). Our aim is to develop technologies that can recognize and interpret contextual factors (based on the analysis of data defining the Forms of Engagement) and use these to discern the plausibility of actions in a given sequence. We can then adapt the animate objects to encourage (or discourage) specific actions.

3.4 Implications for Medical Aids

While the focus of this paper has been on the design and testing of animate objects, it is clear that there could be scope for exploring this concept in the medical domain. Connecting smart objects into a 'rehabilitation internet of things' [11] could offer benefit for monitoring patient activity. Having such objects present cues to guide users can encourage patients to develop or practice actions, particularly if the cues are auditory [12]. Prior work, in the CogWatch project, suggests that adapting familiar objects can reduce anxiety in stroke patients undergoing rehabilitation and also that this can help recall of previously known sequences (lost as a result of stroke). It is a moot point as to whether the type of animate object developed in this project (in which moving parts are intended to cue specific morphological and motor engagement) will prove beneficial for rehabilitation and this is a matter for future research.

3.5 Conclusions

In this paper, we continue our exploration of the ways in which everyday objects can be modified to become animate, and consider how such objects can encourage specific user actions. We demonstrate that people will respond to cues if these correspond to their expectation and understanding of the task. This suggests that the response is mediated by expectation and interpretation. However, relating this mediation to the notion of forms of engagement, we propose that this implies different levels of control. That is, 'affordance' is not simply a matter of responding to the opportunity to act, but is contextualized in terms of the definition of the context. When the context is clear, there appears to be seamless combination of cue response and user activity, such that the cues actually speed-up performance. When the context is unclear or ambiguous (i.e., a cue that is not related to the task being performed, or an outcome that is not clearly understood), response to the cues *could* be helpful but people expend cognitive effort in making sense of these.

References

1. Hornecker, E., Buur, J.: Getting a grip on tangible interaction: A framework on physical space and social interaction. In: CHI 2006, pp. 437–446. ACM, New York (2006)
2. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: CHI 1997, pp. 234–241. ACM, New York (1997)
3. Ishii, H.: The tangible user interface and its evolution. Commun. ACM **51**, 32–36 (2008)
4. Baber, C., Khattab, A., Hermsdörfer, J., Wing, A., Russell, M.: Coaching through smart objects. In: Proceedings of the 11th ACM/EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 298–306 (2017)
5. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1986)
6. Baber, C.: Cognition and Tool Use. CRC Press, Boca Raton (2003)
7. Baber, C.: Cognitive aspects of tool use. Appl. Ergon. **37**, 3–15 (2006)
8. Baber, C.: Designing smart objects to support affording situations. Front. Psychol. **9**, 292–302 (2018)
9. Baber, C., Khattab, A., Russell, M., Hermsdörfer, J., Wing, A.: Creating affording situations: coaching through animate objects. Sensors **17**, 2308 (2017)
10. Jean-Baptiste, E.M., et al.: Intelligent assistive system using real-time action recognition for stroke survivors. In: 2014 IEEE International Conference on Healthcare Informatics (ICHI), pp. 39–44, September 2014
11. Dobkin, B.H.: A rehabilitation internet of things in the home to augment motor skills and exercise training. Neurorehabilitation Neural Repair **31**, 217–227 (2016)
12. Bienkiewicz, M.N., Gulde, P., Schlegel, A., Hermsdorfer, J.: The use of ecological sounds in facilitation of tool use in apraxia. In: 2nd International Conference on Neurorehabilitation (ICR 2014) (2014)