Affordances in psychology, neuroscience and robotics: a survey

Lorenzo Jamone, Emre Ugur, Angelo Cangelosi, Luciano Fadiga,
Alexandre Bernardino, Justus Piater and José Santos-Victor

Abstract—The concept of affordances appeared in psychology during the late 60's as an alternative perspective on the visual perception of the environment. It was revolutionary in the intuition that the way living beings perceive the world is deeply influenced by the actions they are able to perform. Then, across the last 40 years, it has influenced many applied fields: e.g. design, human-computer interaction, computer vision, robotics. In this paper we offer a multidisciplinary perspective on the notion of affordances: we first discuss the main definitions and formalizations of the affordance theory, then we report the most significant evidence in psychology and neuroscience that support it, and finally we review the most relevant applications of this concept in robotics.

Index Terms—Affordances, ecological perception, cognitive system and development, embodied cognition, robots with development and learning skills.

I. INTRODUCTION

The term *affordances* was introduced by the American psychologist James Jerome Gibson in 1966 [1]–[3]. With his own words [1, p. 285]:

> "When the constant properties of constant objects are perceived (the shape, size, color, texture, composition, motion, animation, and position relative to other objects), the observer can go on to detect their affordances. I have coined this word as a substitute for values, a term which carries an old burden of philosophical meaning. I mean simply what things furnish, for good or ill. What they afford the observer, after all, depends on their properties."

What the objects *afford* the observer are action possibilities; according to Gibson, these possibilities are *directly perceived* by the observer from the incoming stream of visual stimuli, without the need to construct a detailed model of the world or to perform semantic reasoning. Only the minimal information that is most relevant for *action* is picked up from the environment, because "perception is economical" [3, p. 135]. Central to Gibson's theory is the notion that the motor capabilities of the agent dramatically influence perception, and that somehow the external environment is understood

L. Jamone, A. Bernardino and J. Santos-Victor are with the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, Portugal, e-mail: {ljamone,alex,jasv} at isr.tecnico.ulisboa.pt; E. Ugur is with Bogazici University, Turkey, and with the University of Innsbruck, Austria, email: emre.ugur at boun.edu.tr; J. Piater is with the University of Innsbruck, Austria, email: justus.piater at uibk.ac.at; A. Cangelosi is with the University of Plymouth, UK, email: a.cangelosi at plymouth.ac.uk; L. Fadiga is with the Istituto Italiano di Tecnologia and with the University of Ferrara, Italy, email: luciano.fadiga at iit.it.
Manuscript received XXX; revised XXX.

in a self-centered and action-centered way. Since Gibson, a number of explanations have been proposed in psychology, while researchers in neurophysiology and brain imaging have reported evidence of neural representations supporting the validity of his theory. Moreover, the concept of affordances has inspired roboticists to develop computational models aiming to improve the perceptual and behavioral skills of robots.

In this paper we review the main definitions of affordances and the most relevant observations in psychology and neuroscience; then we discuss the related work in robotics, presenting the state of the art and identifying the main challenges and the most promising research directions. A brief survey of works dealing with computational models of affordances in the areas of artificial intelligence and robotics was published a few years ago [4]. Here we report a more comprehensive and updated list of such works, together with a deeper discussion on how they link to relevant research in psychology and neuroscience.

The rest of the paper is organized as follows. In Section II we review the main definitions and formalizations of the concept of affordances. In Section III we discuss the most relevant studies in psychology, and in Section IV we report evidence from neuroscience. Then, in Section V we offer a comprehensive survey of the related works in robotics. Finally, we provide a closing discussion in Section VI and some concluding remarks in Section VII.

II. THE CONCEPT OF AFFORDANCES

Jones discusses in [5] how Gibson's thinking with respect to affordance perception has evolved and changed over the years, from his early studies on visual perception [6], to the first appearance of the word *affordances* [1], until his last work [3], where he writes [3, p. 127]:

> "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 to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment."

Still, this definition of affordances is very broad, leaving room for various interpretations. In Gibson's view, a stable surface affords traversing, a predator affords danger, a stone affords hammering. Also, while he mainly focuses on affordance perception, he does not discuss much how the

ability to perceive affordances is acquired, how affordances are represented, and how are they used by the agent to drive its behaviours.

After Gibson died in 1979, the concept of affordances was further analysed by many psychologists, with various authors trying to better formalize it [7]–[15] and others discussing related issues and implications (as, for example, Michaels [16] and Heft [17]). In the following we review the main reinterpretations and derivations of Gibson's theory.

A. Ecological Psychology

Turvey [7] defines affordances as *dispositional properties* of the environment: properties that become apparent only in some specific circumstances, which in the case of affordances is the presence of an agent. For instance, an apple has the disposition to be edible, which becomes an actual property for an animal that is able to eat it (e.g. a pig). According to Turvey, affordances allow prospective control (i.e. action planning), as they inform about action possibilities to achieve goals. More recently, Caiani [8] has further investigated the dispositional interpretation of affordances. According to Caiani, affordances can be viewed as "sensorimotor patterns in the perceptual stimulus that may or may not be associated with some target suitable for action".

As opposed to Turvey, Stoffregen [9] claims that affordances cannot be defined as properties of the environment only, but instead they are relative to the agent-environment system considered as a whole. He defines affordances as [9, p. 115]:

> "...properties of the animal-environment system, that is, that they are emergent properties that do not inhere in either the environment or the animal."

According to Stoffregen, if affordances were belonging to the environment only, then the agent would have to do further reasoning to infer what is available *to him*. Instead, the agent directly picks-up information that already exists (i.e. the *direct perception* theory), and that is related to the agent-environment system.

On the other hand, Chemero [10] argues that affordances are not *properties* (and for sure not properties of the environment), but instead they are *relations* between particular aspects of the agent and particular aspects of the environment. *The apple affords to be eaten by the pig* is somehow similar to *John is taller than Mary* in the sense that it is a relation between two entities, and not a property of one or the other.

B. Computer Science

Efforts to formalize the concept of affordances arise from the computer science community as well, and in particular from the fields of artificial intelligence (AI) and robotics.

Most notably, the work of Steedman [11] provides a computational interpretation of affordances using Linear Dynamic Event Calculus. Steedman does not focus on the perceptual aspect of affordances; instead, he is interested in the relation between the object-actions pair and the corresponding events in the environment. In other terms, his formalism relates the objects (and the afforded actions) to the pre-conditions allowing for the actions to take place, and to the post-conditions generated by the actions. For example, a door is linked to the actions of 'pushing' and 'going-through', to the necessary pre-conditions (e.g. the door needs to be open to allow 'going-through') and to the consequences of applying these actions to the door (e.g. 'pushing' a door which is closed will result in opening the door). The different actions that are associated with an object constitute the *Affordance-set* of that object, which is populated incrementally through learning and development. The explicit modeling of actions pre-conditions and post-conditions naturally allows action planning, for example with a forward-chaining strategy; this constitutes perhaps the most interesting aspect of this formalization.

Also, Sahin et al. propose a formalization that explicitly addresses the use of affordances in autonomous robots [12]. Similar to Chemero, who views affordances as relations between the agent and the environment, they define affordances as acquired *(effect, (entity, behavior))* relations, such that when the agent applies the behavior on the entity, the effect is generated. According to Sahin et al., affordances can be viewed from three different perspectives, namely agent, environmental and observer perspectives: to be useful in robotics, affordances should be viewed from the agent perspective, and explicitly represented within the robot. This formalism has been extended and implemented in a number of robotic studies [18]–[33] that will be reviewed in Section V.

C. Industrial Design

The idea of applying the concept of affordances to objects design comes from D. Norman's popular book: *The Design Of Everyday Things* [13]. In his book, whose original title was *The Psychology Of Everyday Things* (POET), notions from Ecological Psychology are combined with those of ergonomics, generating the concept of *user-centered design*, which focuses on the needs and expectations of the user, disregarding what he thought were secondary issues like aesthetics. A good designer should "make things visible", making sure that the object interface provides the right messages, so that the user can easily "tell what actions are possible" (i.e. the object affordances). In this context [13, p. 9]:

> "... affordances refer to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used."

As pointed out in [15], Norman's discussion on affordances slightly deviates from the Gibsonian view, mainly because of the different objectives that Gibson and Norman had: Gibson was interested in understanding how humans perceive reality; Norman was interested in designing objects so that their utility could be perceived easily. In [14, p. 9], Norman writes:

> "The designer cares more about what actions the user perceives to be possible than what is true".

D. Cognitive Science

The theory of affordances has been useful in explaining how possibilities provided by the environment are perceived and acted on by the organisms. However, it says very little

about how this process is connected to higher-level cognitive skills such as memory, planning and language. A fruit might indeed suggest "eat me" [34], but we would reason over many different factors such as its price, ownership or hygiene before eating it. Indeed, one important question is how to integrate the concept of affordances within a broader cognitive architecture and how to relate it with other cognitive processes.

The early computational model of visual recognition, proposed by the pioneering cognitive scientist David Marr, views recognition as a top-down and object-centric information processing task [35], [36]. The three stages in this task, namely primal, 2.5D and 3D sketches are required to reconstruct the objects in our brain in order to understand and reason over them. Therefore, there is not much space for an ego-centric direct concept like affordances in such a constructivist approach. Vera and Simon, on the other hand, explicitly address affordances in their system [37]. However, contrary to the Gibsonian view, they argue that affordances might be viewed as relatively fast semantic mappings between symbolic perceptual and action representations, advocating the construction of perceptual symbols for the detection of affordances [38].

The constructive and direct approaches to perception have been integrated by several cognitive scientists, including Ulric Neisser, whose perceptual system is composed of three modules [39]. While the first two modules are responsible for direct perception of action and social interaction possibilities, respectively, the third module covers processes such as pattern recognition, language understanding and problem solving [40]. The direct perception module is responsible for the detection of affordances as "every purposive action begins with perceiving affordances" [39, p. 235]. However, according to Neisser, the Gibsonian view of affordances is inadequate, since "it says so little about perceiver's contribution to the perception act" [41, p. 9]. Exposed to objects with a large number of affordances, the agents are put in an active role through a cyclic perceptual activity which prepares them to search for particular affordances at each moment.

E. Summary and discussion

In summary, despite the details of their different interpretations, ecological psychologists (Gibson, Turvey, Stoffregen, Chemero) seem to agree on one major point:

- an affordance manifests itself in relation to the action and perception capabilities of a particular actor.

Another central pillar in their discussion is that affordance perception is *direct*. But what does this mean in computational terms? The main idea is that low level visual features are associated with action possibilities without the need to reconstruct an intermediate fully detailed model of the objects nor to recognize them semantically, making affordance perception precategorical and subconscious; interestingly, this idea finds a neurophisiological support in the discovery of canonical neurons (as discussed later in Section IV). However, *direct perception* does not mean that no intermediate processing happens; on the contrary, a lot of computation is performed by the brain, mainly in the dorsal stream of the visual cortex,

as discussed later in Section IV. Moreover, the existence of this *direct* link between perception and action does not imply that action execution is automatically generated by affordance perception; such phenomena might occur in less developed animals (e.g. frogs [42]), but this is certainly not the case in humans, where affordance perception comes together with several other cogntive processes (as discussed in Section II-D), including inhibition mechanisms (related experiments will be discussed in Section III-B).

Then, the formalizations of Steedman and Sahin add an important dimension:

- an affordance representation should include the effects of the afforded action (with respect to the object that displays such affordance, and with respect to the agent who perceives such affordance).

Interestingly, this overall view nicely accommodates what is generally accepted in experimental and developmental psychology: that perception is deeply influenced by action [43] and that actions are prospective and goal-directed [44].

The interpretation of Norman somehow "simplifies" the notion of what an affordance is, providing a down-to-earth definition which is currently the more popular among the non-specialists, although it maintains the original traits of "agent-dependent" and "action-suggesting".

Finally, studies in cognitive science (mainly by Neisser) point out how affordances can play an important role in more general cognitive architectures, but also that the existing theories still do not fully support that.

III. EVIDENCE FROM PSYCHOLOGY

We discuss here evidence obtained from different fields of psychology. Overall, the three following sections provide support for three major claims about affordances:

(A) humans (and animals) perceive the environment in terms of their body dimensions;

(B) the visual perception of some object properties that are important for action (e.g. size, position) has a *direct* link to action generation, that allows for fast action execution if there is a will to execute the action;

(C) affordance perception depends on the sensorimotor experience of the agent (or, in other terms, affordances are learned by the agent through the interaction with the environment, in a developmental fashion).

A. Ecological Psychology

Although J. J. Gibson introduced the concept of affordances within his theory of visual perception [3], therefore strongly focusing on the visual processes behind affordance perception, most of the later experimental studies in ecological psychology were dealing with the verification of affordance perception capabilities, without elaborating the underlying perception/pick-up mechanisms and the related visual invariants. They typically measure whether or not humans can detect the action possibilities offered by objects and spatial layouts in different conditions, and do not provide detailed insights on how animals perceive affordances. Still, they provide valuable

information on a broad range of affordances by systematically changing the animal/environment systems in a controlled way.

Prior to affordance perception experiments with humans, it was already shown that many animals, including simple amphibians, can perceive locomotion affordances of varying size barriers, apertures, overhead holes and pits [45], [46]. Ingle and Cook, in particular, show that the ratio between the size of the body and the size of the aperture is important in deciding whether to jump or not through a hole for leopard frogs [47]. McCabe is among the first to propose to measure the existence of affordances using ratios between animal and environment properties [48]:

> "Since affordances constitute unique animal/environment compatibilities, one cannot measure those compatibilities with standardized metric systems. ...For the relationship of interest, inches are superfluous. It is the ratio that is perceptually detected and used."

Then, in his pioneering work, Warren shows that a human's judgment of whether he can climb a stair step is not determined by the absolute height of the step, but by its ratio to his leg-length [49]. Furthermore, he shows that humans could detect not only the critical points that signal the existence of the climb-ability affordances, but also the optimal points that correspond to minimal energy consumption configurations for climbing. Next, the researchers have investigated how perception of affordances is affected by changing the 'body' component of the perceived affordance ratio by tricking the subjects about the size of their bodies. They show that when the perceived eye-height of the subject changes (i.e. the subject views the world from a higher or shorter spot without noticing the changed elevation) her judgement on critical points also changes in deciding sitability and climbability [50]. More interestingly, perception of affordances that are not directly related to the height of the subject, such as pass-through-ability that depends on the ratio between the width of the shoulder and the size of the aperture, also changes when perceived eye-height is modified in a similar way [51]. These experiments show that the visual perception of geometrical dimensions such as size and distance is scaled with respect to bodily properties of humans.

Recently, decisions on reaching attempts and execution of such movements have been analyzed on both children and adults, who are asked to fit their hand through openings of various sizes [52]. While subjects of every age group show sensitivity to changes in the environment by scaling their attempts to opening size, supporting the idea that a basic perception of affordances is present since early on and it affects action decisions, younger children (16-month-olds to 5-year-olds) are more likely to attempt impossible openings and to touch openings prior to refusing, suggesting the presence of a slow developmental trend of affordance learning and refinement during growth.

Other researchers investigated additional factors that affect affordance perception in more complicated settings. For example, haptic information is shown to be used together with vision in order to determine the critical points that signal the transition between walking and crawling in infants that move over sloped surfaces [53], [54]. In another locomotion task, it is shown that humans can reliably detect the critical points that signal gap-crossing in dynamically changing layouts [55], [56] even with monocular vision [57]. Furthermore, not only static properties but also the dynamic state of the humans are shown to be important in detecting critical points in complicated tasks such as detecting the time-gap for safe street-crossing [58]. However, how exactly these additional factors are integrated with more basic affordance perceptions has not been fully explained yet.

B. Psychophysics

Ellis and colleagues [59]–[61] have extensively investigated object affordances using the stimulus-response compatibility effect paradigm, and the more general relationship between affordances and embodied cognition. For example, in [60] they investigate what happens when a visual target object has to be attended in the presence of distractor object. They measure the participants' reaction times to produce the responses requested by the experiment (i.e. to classify as either "round" or "square", by producing a precision or a power grip, a target 3D object presented on a computer screen) and study how they are influenced by the congruence or incongruence of the distractor size with the requested responses. The results show the target-related compatibility effects found in previous experiments without the distractor, and also show an unexpected interesting effect of the distractor: responding to a target with a grip compatible with the action afforded by the distractor produces slower reaction time in comparison to the incompatible case. They interpret these results proposing that the inhibition of the action elicited by the distractor interferes with the execution of similar actions in comparison to different actions.

Recently, a paradigm called spatial alignment effect has been used to measure to what extent object related affordances depend on the space, i.e. reachability boundaries [62]. This paradigm compares the reaction times of an action towards the same object in different configuration, i.e. in different position and orientation. Faster reaction time to a configuration corresponds to 'direct' perception of the affordance, and therefore faster activation of the action. When left or right handled 3D mugs are placed in peripersonal or extrapersonal spaces, the reaction time in response to a left-/right-hand grasp command is fast only when the hand-handle sides are congruent and the mug is within the peripersonal space [63]. In a similar experiment performed with real mugs that spatial compatibility effect is also observed when the object is presented not just in the actual-reaching spaces, but also in near-reaching spaces which are also perceived to be reachable [64]. These experiments suggest that 'direct perception' of graspability depends on both grasp-related properties of the object and 'direct grasp' activation potential without any intermediate movement.

C. Developmental psychology

Affordances provided by the environment are not fixed; instead, they change over the course of the lifetime of the human. New motor skills bring new action possibilities, which,

in turn, should be learned to be detected perceptually. For example, after acquiring independent sitting capability, infants can actively explore the objects with their freed-up hands, and learn about manipulation affordances. Accordingly, they start exhibiting sensitivity to three-dimensional form of objects, and reason about completion of incomplete parts of these objects only after self-sitting experience [65]. At 6 months of age, they manipulate objects in a differentiated manner depending on the object's properties [66]. During the second half year they become sensitive not just to objects and surfaces alone, but to the affordances entailed by the relation between the two [67]. However they need to start crawling and walking, in order to perceive affordances related to the spatial layout of the environment. While crawling children perceive both rigid and non-rigid surfaces as traversable, walking children learn that non-rigid surfaces such as waterbeds are not 'that much' traversable [68].

Although J. J. Gibson himself mentions that affordances are learned in children [3], he does not discuss the learning mechanisms in detail because he is not particularly interested in the development of affordance perception [69, p. 271]. Likewise, all the experiments in Ecological Psychology focus on affordance perception in adult agents in a fixed time point in their lifetime. Instead, the line of research established by Eleanor Jack Gibson, who "nearly single-handedly developed the field of perceptual learning" [70], has been a notable exception in studying the learning of affordances from both developmental and ecological perspectives. As J. J. Gibson writes [71]: "we divided the problems between us, and she has concentrated on perceptual learning and development while I concentrated on the senses. Her forthcoming book (E. J. Gibson, 1969) will take up the story where mine leaves off".

E. J. Gibson believes that "ecological approach to perception is the proper starting place for a theory of perceptual learning and development" [72]. Indeed, in contrast to the main views of her time, she believes that the ambient arrays of light that arrive to the receptors are already structured and carry information related to the affordances of objects: the development corresponds to learning how to pick-up this information. Following the ideas that she developed in her early work on learning of reading in children [73], she argues that learning is neither the construction of representations from smaller pieces, nor the association of a response to a stimulus. Instead, she claims that learning is "discovering *distinctive* features and *invariant* properties of things and events" [74], [75, p. 295]. In learning the letters, for example, children detect different combinations of distinctive features which are invariant under certain transformations such as straight lines and curves [73]. Perception is not constructing a new description of the world; therefore learning is not enriching the input but discovering the critical perceptual information in that input. She names this process of discovery *differentiation*, and defines it as "narrowing down from a vast manifold of (perceptual) information to the minimal, optimal information that specifies the affordance of an event, object, or layout" [75, p. 284]. Her position is related to modern views in Cognitive Science that see visual perception as an active process of discovery, supported by sensorimotor knowledge [76]. Indeed,

she also notes how babies use exploratory activities, such as mouthing, feeling, licking, and shaking, to generate the perceptual information; exploration increases predictability and efficiency of affordance perception [77], and terminates with a significant reduction of uncertainty [78].

In the following we report the most relevant milestones of child development, as observed in a number of psychology studies, and relate them to the progressive acquisition of affordance perception and exploitation skills.

1) Early sensorimotor development: The sensorimotor development in humans already starts in the womb [79], [80] and progressively shapes infant behavior after birth into the childhood [81], [82]. Newborns have several innate reflexes such as pupil reflex to light, sucking reflex to objects in mouth, mono reflex to sudden movements or palmar-grasp reflex to objects inside palm [83]. They also have crude skills such as primitive form of hand-eye coordination, which can direct both eyes and hands of the newborns toward visually detected events and objects [84]. These reflexes and primitive skills transform to more complex intentional sensorimotor programs, and help development of advanced cognitive capabilities. For example, palmar-grasp reflex, which automatically closes infant's fingers in response to objects in the palm, transforms into intentional grasping [81], and disappears by 6 months of age [85]. Rudimentary hand-eye coordination skill, on the other hand, is continuously exercised by frequently moving the hand in front of the eyes between 2 and 5 months of age [86]. Consequently, by 4 months of age, infants learn to perceive reachability affordances [85, p. 199]. By 5 months of age, they learn to slow down the speed while approaching to the object [85, p. 100] with accurate reach trajectories [87, p. 41]. It takes 9 months for infants to reach for objects with correct hand-orientation and adjust their grip size based on objects' size before contact [85]. Affordances related to orientation and size of objects develop later than position related affordances as probably the former requires more complex processing of the visual data.

2) Means-end behaviors, predicting effects and imitation: By 7-8 months of age, infants start exploring the reachable objects with qualitatively distinct manipulative actions such as pinch-grasping, rubbing, holding, and shaking [88], [89]. The actions around this age are relatively simple [90] and mostly involve single objects [91]. Focused exploratory behaviors and multimodal exploration have special importance for learning in this stage [92]. Mostly without intervention of the parents, infant self-explores and self-observes the environment in a goal-free fashion [93]–[96], driven by intrinsic drives such as curiosity, novelty and surprise [97]–[99]. Through such exploration and learning, they start distinguishing actions from their consequences [100]. Observing the consequences of actions and relating these consequences to the visual and other properties of the objects leads to learning of object affordances [72]. After approximately 9 months of age, the infants start using learned affordances to achieve certain goals, predicting desired changes in the environment to achieve the goals, and executing the corresponding actions [101]–[103]. By 10-12 months, they can make multi-step plans using learned affordances and sequencing the learned action-effect mappings

for different tasks. For example, they can reach a distant toy resting on a towel by pulling the towel first or retrieve an object on a support after removing an obstructing barrier [104]. At around 12 months, infants have also become skilled in reproducing the end states of observed actions that involve objects. Goal emulation, i.e. reproduction of the observed goal [96] with a tendency of skipping the demonstrator's strategy [105], becomes possible at this age; one possible explanation might be that at this time infants can predict the effects of their actions on the objects based on the learned affordances, and can therefore obtain a self-centered representation of the observed goals by chaining such predictions. On the other hand, infant's affordance learning strategy changes around this age as well. Provasi et al. [93] showed that while 9 month old infants can learn affordances of containers only through self-exploration, 12 month olds learn more effectively if opening action is demonstrated by others. This data suggests that social learning through goal emulation [96] becomes available with learned affordances, and helps learning further affordances. While younger infants are more inclined to emulate observed goals, older infants tend to exactly imitate the demonstrated action sequence even if they notice that those actions are not causally related to the goal [106]. On the other hand, copying the exact action sequence is not straightforward for young infants as the affordances and observed action trajectories might not match to infants' own sensorimotor repertoire [107], [108]. In order to overcome this challenge, parents support infants by making modifications in their infant-directed actions, i.e. they use "motionese" [109]. For example, they insert pauses while executing complex actions to highlight the subgoals that can be sequentially (and more easily) achieved by infants [110]. Motionese might serve as a bridge from affordance-based goal-emulation mechanisms to complex imitation capabilities [21].

3) Tool use and problem solving: Infants start exploring multi-object affordances from month 12 [111]. They insert rods into circular holes in a box or stack up blocks into towers of two blocks from 13 months [112]. While at 13 months infants can only insert circular rodes into circular holes in a plate, by 18 months they can perceive the correspondence between different shaped blocks and they start inserting different shaped blocks into corresponding holes [113]. Finally, by 18 months of age, they can perceive affordances that involve more than two objects and act accordingly. At the second year, in general, children begin using objects for increasingly complex problem solving and tool use [114]. While a number of researchers have suggested that tool use requires a cognitive leap beyond information that is directly perceived, thus requiring the ability to engage in novel forms of symbolic or relational thinking [115], a new wave of research proposes an alternative view in which tool use is seen as an extension of the perception-action coupling that infants show during the first year of life; therefore, the very concept of tool may emerge from the detection of possible affordances between objects or object parts, based on information that is directly perceived [116]. From this perspective, the trial and error attempts that precede successful tool use can be seen as exploratory behaviors, providing opportunities for affordance learning. Indeed, Lockman sees tool use as gradually emerging

through sensorimotor experience, building-up on objects affordance knowledge [116]; his position is close to the embodied cognition approach [117], which assumes that to use tools people need to activate past sensorimotor experience with them, but no semantic reasoning skills. However, whether tool use emerges progressively through familiarization with experience or it appears through sudden insight is still an unresolved issue. For instance, the recent observations of Fagard et al. [118] seem to support the latter hypothesis. In [118] a longitudinal study on five infants from age 12 to 20 months is reported; children have to use a rake-like tool to reach toys presented out of reach. Their results indicate that it is only between 16 and 20 months that the infants suddenly start to intentionally try to bring the toy closer with the tool. According to Fagard, this sudden success at about 18 months might correspond to the coming together of a variety of capacities, such as the development of means-end behavior. This is in line with the early view of Kohler, which sees tool use as appearing from sudden insight [119]. Recently, an extensive multimodal dataset was collected by recording videos of 124 human subjects performing visual and manual exploration of unfamiliar tools while verbalizing their experience [120]: the participants were exploring lithic tools, for which the affordances/functions were not known. Interestingly, the analysis of this corpus of data could provide further insight about the visual (and multimodal) processes that support both the learning of affordances and the attempts to infer the affordances of unknown objects and tools.

D. Summary and discussion

The experiments described in Section III-A show that humans (and other animals as well) perceive affordances in the environment depending on their body dimensions. However, they say little about what is exactly the visual information that is picked (e.g. what low level features?), and why that specific information is selected (among all the possible choices).

The studies reported in Section III-B clarify some of the concepts proposed by the ecological psychologists, in particular with respect to the relationship between affordance perception and action execution: even if an afforded action is *suggested* upon visual presentation of an object, its execution can be *inhibited* by other cognitive processes, if there is no conscious will to perform the action. If there is a conscious will, instead, then the execution of the afforded action is faster.

While Sections III-A and III-B focus only on affordance perception, in Section III-C we discuss affordance learning: i.e. learning how to perceive affordances, learning how to extract the minimal visual information that is most relevant for action. Research in developmental psychology clearly shows that the human ability to perceive affordances emerges gradually during development, and it is the outcome of exploratory and observational learning; as this ability appears, the children also start to be capable of predicting the effects of the actions, eventually achieving problem solving skills. This body of research offers a clear suggestion to roboticists: in order to effectively *perceive* affordances, robots must first *learn* from their own sensorimotor interactions with the environment.

IV. EVIDENCE FROM NEUROSCIENCE

The evidence from neuroscience reported hereinafter helps to better interpret computationally the observations and intuitions coming from psychology, and more specifically:

(A) that perception of action-related object properties is fast;
(B) that perception and action are tightly linked and share common representations;
(C) that object recognition and semantic reasoning are not required for affordance perception.

A. Visual processing in the primate cortex

According to two visual streams theory [121]–[124], visual information is processed in two separate pathways in primate cerebral cortex. The ventral pathway plays an important role in constructing semantic perceptual information about the objects through categorization, recognition and identification [125]. The dorsal pathway, on the other hand, processes visual information to control object-directed actions such as reaching and grasping [126], and presents shorter latencies (about $100 msec$) with respect to the ventral pathway [127]. In particular, edge detection (visual area V1), depth processing (visual area V3) and surface and axis representations (CIP - Caudal Intraparietal Area) are critical subprocesses of the dorsal visual pathway of the primate cortex leading to the affordance representation in AIP - the anterior intraparietal area [128]. A comprehensive description of the information processing occuring in the primate visual cortex, from a computational perspective, is provided in [129].

According to J. Norman, it is straightforward to conclude that "the pickup of affordances can be seen as the prime activity of the dorsal system" [130, p. 143]. Indeed, more recent anatomical and physiological evidence has led researchers to propose a further subdivision of the dorsal stream into a dorso-dorsal and a ventro-dorsal sub-streams [131]. While the dorso-dorsal sub-stream seems to be more involved in the online control of action through proprioception, the ventro-dorsal one provides somatosensory and visual information to the ventral premotor cortex. Ventral premotor area F4 is in fact involved in coding the peripersonal space for reaching [132], [133] and area F5 contains neurons coding hand/mouth grasping.

B. Visuo-motor neurons

The majority of F5 neurons discharge during goal-directed actions such as grasping, manipulating, tearing, holding [134] but do not discharge during similar fingers and hand movements when made with other purposes (e.g. scratching, pushing away). F5 grasping neurons show a variety of temporal patterns of activation. Some neurons are more active during the opening of the fingers, some discharge during finger closure and some others discharge during the whole movement. More interestingly, many grasping neurons discharge in association with a particular type of grip (precision grip, finger prehension and whole hand prehension). Taken together, the functional properties of F5 neurons suggest that this area stores a set of motor schemata [135] or, in other terms, a 'vocabulary' of motor acts [134]. Populations of neurons constitute the

'words' composing this vocabulary. This vocabulary-like structure is characterized by a syntactic organization. Different levels of generalization are present, from very specific neurons discharging only during grasping of a particular object (e.g. a small piece of banana but not a small piece of apple) to very generalizing neurons responding during movements that share the same goal but are performed with different effectors (e.g., when the monkey grasps an object with its right hand, with its left hand or with the mouth). Interestingly, this organization seems to support the principle of motor equivalence postulated by Bernstein early on [136]. In addition to their motor discharge, many F5 neurons (about 20%) have been shown to fire also in response to food/object visual presentation [134]. More recently the visual responses of F5 neurons were re-examined using a formal behavioral paradigm to separately test the response related to object presentation, during the waiting phase between object presentation and movements onset, and during movement execution [137]. The results showed that a high percentage of the tested neurons, in addition to the 'traditional' motor response, responded also to visual presentation of three-dimensional graspable object. Among these visuo-motor neurons, two-thirds were selective to one or a few specific objects. When visual and motor properties of F5 neurons are compared, it becomes evident that there is a strict congruence between the two types of responses. Neurons that become active when the monkey observes objects of small size, discharge also during precision grip. In contrast, neurons selectively active when the monkey looks at large objects, discharge also during whole-hand prehension.

The finding in macaque monkeys of visuo-motor responses at the two sides of the frontoparietal circuit for grasping, which includes parietal areas AIP/PF/PFG and ventral premotor area F5, strongly supports the affordance idea. Frontoparietal neurons might be devoted to transform object visual information into grasping actions. Premotor cortex, in turn, sends projections both to the primary motor cortex and to the cervical enlargement [138]–[140]. Object-related visuo-motor neurons have been successively named *canonical neurons* [141] to distinguish them from the other class of visuo-motor neurons of area F5, the *mirror neurons*, responding instead to action observation [142]. Further studies confirmed the existence of canonical neurons within ventral premotor cortex [143], and within intraparietal region AIP and posterior parietal cortex [144], [145]. It should be noted, however, that while F5 visuo-motor neurons seem to group actions according to a motor syntax (see above), parietal (AIP) neurons seem more influenced by geometrical features of objects and when some generalization occurs, it seems more dependent on geometrical clustering [144].

These results have interesting parallels in humans. Experiments on human subjects using either fMRI [146] or TMS [147]–[149] have shown sub-threshold activations of specific motor neurons during observation of objects that affords specific actions. Interestingly, these motor activations appear to be more pronounced if the observed objects are within reaching distance [150], of if they are reachable by another agent [151]. EEG studies investigating the time course of affordance activation have also shown that early sensory visual

pathways are modulated by the action associated with objects and by the intentions of the viewer [61].

C. Object recognition and semantic reasoning

The visuo-motor control of manipulation [126] and locomotion [152] actions was shown not to be affected by impairments in the ventral stream. A patient with such impairment was able to successfully avoid obstacles, or insert mail into slots in correct orientation using her dorsal system. However, while performing actions successfully, she did not recognize the objects she was interacting with, and cannot report the related properties of the objects such as orientation of the slots or height of the obstacles. Other experiments measured reaction times towards various familiar and unfamiliar objects, and they revealed that semantic information about the objects (e.g. object category) does not appear to produce detectable effects in priming actions; on the other hand, subjects act faster when the actions are congruent with the perceived visual qualities of even unfamiliar objects [153]. Therefore, it seems that humans do not need to recognize objects in order to perceive and act on their immediate affordances. Still, semantic object recognition information and the corresponding properties were shown to be communicated to the dorsal pathway in a top-down fashion for control of manipulation actions [154].

This hypothesis seems to be confirmed by brain imaging studies as well. In [155], Humphreys showed that, when presented with a tool, some patients, who lacked the ability to name the tool, had no problem in gesturing the appropriate movement for using it. According to Humphreys, this suggested a direct link from the visual input to the motor actions that is independent from more abstract representations of the object, e.g. its name. In another study that Humphreys presented in [155], two groups were shown object pictures, non-object pictures and words. One of the groups was asked to determine if some actions were applicable to what it was presented. The other control group was asked to make size judgments. The brain activities in both groups were compared using functional brain imaging. It was observed that a specific region of the brain was activated more in the first group who were to make action judgments. It was also seen that this specific region was activated more when the subjects were presented with pictures of the objects rather with the names. This showed that action related regions of the brain were activated more when the visual input was supplied (i.e. sensory representation), rather than the linguistic one (i.e. abstract/symbolic representation).

D. Summary and discussion

Overall, research in neuroscience provides important information about the neural representations that seem to permit affordance perception. More specifically, the discovery of canonical neurons (described in Section IV-B) unveils the existence of specific neural structures that are involved in both action execution and action perception (or better, affordance perception) [156], supporting a more general view in neuroscience and cognitive science that sees action and perception as closely related [157]–[159]. These conclusions

are interestingly similar to those of the ecological approach [160, p. 1236]:

> This process, in neurophysiological terms, implies that the same neuron must be able not only to code motor acts, but also to respond to the visual features triggering them. ... 3D objects, are identified and differentiated not in relation to their mere physical appearance, but in relation to the effect of the interaction with an acting agent."

Moreover, several studies support the idea that object recognition and semantic reasoning are not required for affordance perception: different brain areas are involved in the related processing (see Section IV-A) and patients that show inabilities to recognize objects could instead perceive their affordances (see Section IV-C).

V. AFFORDANCES IN ROBOTICS

The concept of affordances is highly relevant to autonomous robots and it has influenced many studies in this field.

First, it is important to underline the similarity of the arguments of J. J. Gibson to those of the reactive and behavior-based roboticists (e.g. R. Brooks [161]). This similarity was initially noted by Arkin [162], and then further discussed by Sahin [163]. Interestingly, both the theory of affordances and that of behavior-based robotics emerged as alternative suggestions to the dominant paradigms in their fields, by opposing direct action-perception couplings to modeling and reasoning. Indeed, Brooks's claim "the world is its own best model" [161, p. 13] puts forward the idea of a robotic *direct perception* in which a complex internal model of the environment can be replaced by a number of simple sensory-motor mappings. Similarly, Murphy [164] suggests that robotic design can draw inspiration from the theory of affordances to eliminate complex perceptual modeling without loss in capabilities. Also, Duchon et al. [165] have invented the term *Ecological Robotics* referring to the application of ecological principles to the design of mobile robots, like for example the use of optic flow for navigation control.

However, when scaling from simple reactive behaviors to more advanced scenarios (e.g. manipulation of different objects, complex problem solving), the solutions proposed by behavior-based robotics seem to be not powerful enough. Still, computational models of object affordances can be extremely helpful to capture the most informative object properties in terms of the actions that a robot is able to perform, and in terms of its available sensors. Also, they might offer an elegant and effective solution to represent that sensorimotor knowledge that is fundamental to obtain meaningful predictions about the consequences of the robot actions, which are needed for planning and problem solving. Indeed, when looking at the different formalizations of the concept of affordances (that we reviewed in Section II and summarized in Section II-E) and at the evidence from psychology and neuroscience (outlined in Section III and Section IV respectively) there seem to be two main aspects which are very relevant for robotics:

i) affordances depend on the perceptual and motor capabilities of the agent (they are not properties of the

environment alone), and the ability to perceive them is acquired by the agent through a long sensorimotor learning process;

ii) affordance perception suggests action possibilities to the agent through the fast activation of sensorimotor (i.e. visuo-motor) neural structures, and it provides also a mean to predict the consequences of actions.

Overall, during the last twenty years many roboticists have been successfully using ideas related to affordances and direct perception for the design of intelligent robots. In the following we discuss the most influential among such works.

A) In Section V-A we introduce pioneering works in which robots are learning action-related object properties by doing actions: this is the central idea behind affordance learning, which then permits affordance perception.

B) Then, in Section V-B we review the works in which the effects of the actions are explicitly considered, leading to affordance models that can be used not just to suggest the afforded action, but also to predict its consequences.

C) Section V-C deals with grasp affordances; we discuss these works in a separate subsection because they are very specific to the grasping problem.

D) The studies included in Section V-D extend the concept of object affordances to pairs of object that can interact together (i.e. multi-objects affordances), leading to the concept of tool use.

E) Then, in Section V-E we report research in which models of affordances (both of single-object and multi-objects) are used to make predictions of the effects of the actions, and such predictions are employed for action planning and problem solving.

F) In Section V-F we highlight the importance of the affordances in human-robot interaction (HRI), by reporting works in which the robot ability to perceive affordances improves the quality of HRI.

G) In Section V-G we review studies of developmental and cognitive robotics in which models of affordances are included in developmental learning processes and cognitive architectures.

H) Finally, in Section V-H we summarize the state of the art in affordance-related robotics research.

A. Learning action-related object properties

Among the very first attempts to exploit robot actions to improve object perception is the work of Krotkov [166]. In particular, in the "Whack and Watch" experiment, the mass and friction of different objects are estimated by poking them with a wooden pendulum while recording their motion with a camera. Although the claimed objective of the study is to detect the materials of the objects, Krotkov suggests that:

> "Perception of material types and properties will contribute significantly to the emerging area of research on reasoning about object functionality. ... The ability to classify objects by their material properties will permit deeper reasoning, for example, recognizing that a hard-heeled shoe could substitute

for a hammer, even though their shapes differ dramatically."

Interestingly, similar experiments have been performed in most of the early works on robotic affordances. Fitzpatrick et al. [167]–[169] propose an ecological approach to affordance learning, putting forward the idea that a robot can learn affordances just by acting on objects and observing the effects: more specifically, they describe experiments in which robots learn about the rollability affordance of objects, by tapping them from several directions and observing the resulting motion. The first step is to understand that tapping an object has the effect of generating an optic flow in the images extracted from the robot cameras, i.e. a perceptual invariant that is directly perceived [167], [168]. Then, during repeated explorations, the robot can learn the relationship between such effects and the action-object pairs, and eventually perform a simple emulation task in which the best action to reproduce an observed motion of the object has to be chosen [169]. Stoytchev presents simulation results in which a robotic manipulator applies random sequences of motor behaviors to different objects and attempts to learns the so-called 'binding affordances', i.e. behavior sequences that generate specific motions of the objects [170].

One limitation of these early studies is that what is learned are the affordances of specific objects. This means that the identity of the object must be recognized by the robotic agent before the affordances could be inferred; however, this strongly limits the generalization capabilities of the system to never-seen-before objects. Moreover, this approach seems to be in contrast with the idea of direct perception from Ecological Psychology, in which perceived visual features are directly associated with an affordance, without an explicit recognition/classification of the object.

B. Representing the effects of the actions

Later works address the problem of relating the perceived properties of the environment to the learned affordances, and also to the effects of the afforded actions.

1) Effects from internal indicators: A number of early simulation studies propose to use internal drives in learning affordances and extracting the related invariants from the continuous sensorimotor experience of the robot. MacDorman [171] studies the extraction of invariant features of different affordances provided to a simulated mobile robot, where affordances are provided as internal feedbacks from interactions with different objects such as tasty or poisonous ones. In his study, the invariant features are defined as perceptual signatures that tend not to vary among the samples of the same affordance category but they do vary among different affordance categories. Cos-Aguilera et al. [172], [173] cluster the sensory space of a wandering simulated mobile robot to capture the regularities in the environment, and learn the relations between the active clusters and the predefined outcomes of interactions with these clusters. Indeed, the learning and extraction of invariant features and the use of internal indicators are interesting ideas, that originate from the observations in Ecological Psychology (mainly of E. J. Gibson

about *differentiation* [74], [75, p. 295], see Section III-C) and Developmental Psychology (about instrinsic motivation and curiosity [97]–[99], see Section III-C2). However, the robotic works described above do not provide information on how to exploit affordance perception in realistic goal-directed tasks.

2) Pre-defined effect categories: Fritz et. al [174] introduce the concept of *affordance cueing*, which consists in detecting salient regions of the visual stream where the perceptual invariants that allow affordance detection are present. Within such regions, only the visual features that support the identification of an affordance are selected. In a later extension of the work [175], the system is applied to a real world scenario; however the perceived affordance (i.e. liftability) depends on a very simple cue: whether the top of the object is flat or not. In order to perceive and act on more complex affordances, such as power- and precision-graspability affordances provided by mugs with handles to an anthropomorphic robot hand, low-level features that are directly extracted from depth images of the objects were shown to be effective [32]. Ugur at al. [22], [176] learn traversability affordances of a robot for navigation in indoor environments in the presence of everyday objects such as balls, tables and doors. After exploration and learning in simulation, given features extracted from grids in the depth image, the real robot is able to predict which directions are traversable without using an intermediate object detection process. The robot is able to detect critical points for traversability through apertures, under overhead obstacles, over slopes, and against cylinders lying in different orientations without explicitly recognizing objects, in scenarios similar to the ones used in the experiments in ecological psychology summarized in Section III-A. Dogar et al. extend this approach and use learned affordances in a goal-directed way by selecting a movement primitive [23] or by blending the pre-coded movement primitives based on their affordances [26]. Cakmak et al. [25] go one step ahead by chaining affordance-based predictions of outcomes of actions and generating plans; however object-free perception of the environment is very limiting to make complicated plans. Movement possibilities that are provided to the end effectors of manipulators can also be considered as traversability affordances as noted by Erdemir et al. [177], who propose a cognitive architecture in which the agent can predict the consequences of its actions based on internal rehearsal. Kim et al. [178], on the other hand, study traversability in more difficult outdoor environments with different objects and layouts such as wet grounds, grass fields, trees and brushes. In an online learning setting, the robot collects image snapshots while traversing the environment, self-labels them based on its collision experience, and learns the relations between low-level visual features and traversability through classification. Eventually, the robot exhibits high performance in difficult terrain, for example by detecting tall grasses, which look like substantial obstacles due to their vertical extent, as traversable.

3) Probabilistic representations: All the aforementioned studies rely on deterministic one-directional mappings. Instead, the works in [179]–[189] use probabilistic networks to capture the stochastic relations between sensorimotor variables. These representations allow for bi-directional inferences, that enable prediction, imitation and planning. More-

over, probabilistic learning greatly improves the ability of the system to deal with uncertainty, redundancy and irrelevant information, efficiently implementing Gibson's idea of *economical perception* [3, p. 135] (see Section I).

Demiris and Dearden [179] propose the use of Bayesian Networks (BN) [190] to learn a forward model that relates the robot motor commands to the visual effect of the resulting motion. Such a model can then be inverted to imitate a visually detected motion of another agent (e.g. a human). However, no object-oriented actions are considered in this study. Hart et al. [180] report experiments in which the humanoid robot Dexter learns the liftability affordance of objects using probabilistic relational models [191], and in particular Relational Dependency Networks (RDN) [192]. Both object properties (e.g. bounding box dimensions) and action properties (e.g. hand approaching direction) contribute to the liftability estimation: for example, a long box used in the experiments was always graspable, but was liftable only when approached from the top. The authors also propose to employ Relational Probability Trees [193] to explicitly analyse the contribution of the different sensorimotor variables encoded in the RDN representation, making evident that only some variables were indeed affecting the object liftability. Sun et al. present in [186] an extension to their previous work [178]. While in their first system the learned affordances are directly associated with the perceived visual appearance of different terrains (i.e. the direct perception approach), in their later study the affordances are related to categories that are formed on top of the visual features, using a probabilistic graphical model that they call the Category-Affordance model. Results obtained with two different mobile robots in indoor scenarios show that the addition of this intermediate categorization step improves the prediction accuracy of the system, especially when a small amount of training data is available. Kroemer et al. [188] implement direct perception of graspability and pourability affordances without using any hand-designed features. For this, they use a non-parametric representation for affordance-bearing parts of objects, which is based directly on the point clouds perceived by the robot. Hermans et al. [189] learn pushability affordances of objects through training regressors that predict straight-line or rotation scores based on local and global visual features. After learning through exploration, the mobile manipulator robot could infer effective push points for novel household objects regardless of whether they belong to a previously encountered object class. Similarly, Kopicki et al. [187] learn probabilistic models of outcomes on pushing manipulations. However, this type of predictive models could only make prediction in a uni-directional way.

The common limitation of many of these studies is that they consider the existence of only one affordance (e.g. liftability, traversability, containability). It is not obvious then how these approaches can generalize to the more realistic case in which multiple affordances are present in the environment, and learning one particular affordance influences the prediction of other affordances. A more general computational model, which supports learning of different affordances, is proposed later on by Montesano et al. [181]–[184]. The model employs a Bayesian Network (BN) to encode the probabilistic

relations between a set of random variables describing three main sensorimotor entities: actions, objects and effects. The data is discretized using k-means clustering to train the BN efficiently; a later version of the model allows to directly use continuous variables, by using a Gaussian Mixture Models representation of the perceived visual features as input of the BN [185]. After learning, the BN can be used to infer the conditional probability distribution of any set of variables with respect to any other variable, allowing several robot abilities: e.g. to predict the outcomes of an action on a visually presented object, to select the action that would generate a desired/observed effect, to estimate what action has generated a visually detected effect. Interestingly, these abilities match a number of behaviours observed in human babies, as reported in Section III-C2.

4) Self-discovered effect categories: A number of researchers focused on affordance learning methods that first categorize the effect space in an unsupervised way, and then learn the relation between object features and the discovered categories, thus achieving a self-discovery of effect categories that is compatible with the modern view in Developmental Psychology [101], [102] and Neuroscience [158]. Griffith et al. [194], [195] describe a learning procedure in which the robot discovers whether objects can be classified as containers. The recorded effects of the action are clustered in an unsupervised way using X-means [196] and used to automatically determine the object category (i.e. container or non container). The perceived visual features of the objects are also clustered through X-means, and associated with the effect category using sparse coding [197]. Using such stored experience the robot is able to then infer the category of unknown objects based on their perceived visual features. While Ugur et al. [27] apply X-means clustering that is followed by SVM classification, Ridge et al. [198] use Self-Organized Maps and Kohonen's learning vector quantization to learn push related affordances. Ridge et al., in a follow-up work [199], show a gain in performance when affordances are learned using features that are particularly defined dynamically with respect to the corresponding manipulation actions. Ugur et al. [19], [33] realize a two-level clustering algorithm that takes into account the representational differences between different perceptual channels and uses a verification step that makes sure that the discovered effect categories can be predicted by the robot. Similar ideas on generating useful categories were also proposed by others, where categorization is also based on the ability to predict the outcome of action execution (Mugan and Kuipers [200]) and categories are used only if they appear in the learned rules (Pasula et al. [201]).

C. Grasp affordances

Because the detection of grasp affordances is essential for robotic manipulation, there is a large body of literature on this topic. The solutions proposed in the literature can be roughly categorized in two groups: analytic and data-driven approaches. Analytic approaches [202], [203] rely on precise geometrical and physical models of the objects and of the manipulators in order to compute the optimal finger placement for stable grasps. Data-driven approaches [204], on the other hand, use self-generated or existing databases of grasp experiences in order to assess grasp affordances of target objects. Bogh et al. [204] divide data-driven approaches into three groups: grasping known, familiar and unknown objects. The approaches in the first category synthesize grasps for known objects, where the identity and pose of the target object are recognized first, and the grasp control for the corresponding object are fetched from the database. While some of these studies use simulators [205] in order to sample and evaluate grasps for objects represented as 3D shape primitives [206], others learn grasps from human examples [207] and refine them with robot experience [208], [209]. The solutions of the second category compare the target object with the known objects using similarity metrics that take into account grasp related visual object features [210], [211], with the assumption that similar objects can be grasped in similar ways. Finally, the studies in the third category directly compute grasp affordances by identifying perceptual structures in the form of local [212] or global [213] features. Analytical approaches provide guarantees of stability and robustness; however, they assume exact knowledge of relevant properties of the object (e.g. 3D geometry, center of mass, weight, surface friction) and of the robot hand (e.g. dynamic and kinematic model), that are typically not known precisely in real environments. Similarly, data-driven approaches for known objects give high performance, but cannot deal with novel objects at all. Data-driven approaches for familiar or unknown objects, on the other hand, can infer grasp affordances of novel objects; however, since they rely on previously acquired knowledge, they require large datasets of object images and robot-object interactions, and might be computationally expensive, making the real-time extraction of object features challenging.

Overall, the approaches that best fit the interpretation of affordances given by Ecological Psychology and Neuroscience are the data-driven grasp synthesis approaches for familiar or unknown objects, which employ datasets that are (at least partially) generated by the robot [210]–[213]. In other terms, these are the approaches in which the graspability affordance is perceived from low level image features, instead of using analytical reasoning over 3D object models, and in which the robot embodiement and sensorimotor capabilities are taken into consideration to learn what are the relevant low level features to look for in the images (i.e. how to pick-up the affordances); these approaches do not include an object recognition step in the visual pipeline, and they can better generalize to novel objects by looking for distinctive low level image features that indicate graspability.

D. Multi-objects models and tool use

A few computational models of affordances deal with multi-objects scenarios, either in terms of tool use [214]–[223] or pairwise object interaction [224], [225], with the long-term objective of obtaining more complex problem solving abilities in autonomous robots. A robot agent specifically tailored towards learning tool use is reported by Wood et al. [214]. In their work, an artificial neural network is used to

learn appropriate postures for reaching and grasping tools, on board the Sony Aibo robot platform. Sinapov and Stoytchev [215]–[217] investigate the learning of tool affordances as tool-behavior pairs that provide a desired effect. However, what is learned are the affordances of specific tools (i.e., considered as individual entities), and no association between the distinctive features of a tool and its affordances is made. The generalization capabilities of the system are limited to dealing with smaller and larger versions of known tools. The work by Tikhanoff et al. [218] focuses on learning a specific affordance (i.e., pulling) during tool use. Although useful for robot operations, this knowledge is specific for the tool that is experienced, and cannot be easily generalized to novel (i.e. previously unseen) tools. In a recent extension [219] visual features extracted from the functional part of the tool are related to the effects of the action, and this allows to adjust the motion parameters depending on how the tool has been grasped by the robot. An interesting approach has been proposed by Jain et al. [220], in which a Bayesian Network is used to model tool affordances as probabilistic dependencies between actions, tools and effects. To address the problem of predicting the effects of unknown tools, they propose a novel concept of tool representation based on the functional features of the tool, arguing that those features can remain distinctive and invariant across different tools used for performing similar tasks. However, it is not clear how those features are computed or estimated, if they can be directly obtained through robot vision and if they can be applied to different classes of tools. Moreover, it is worth noting that in [215]–[220] the properties of the acted objects are not explicitly considered in the model; only the general affordances of tools are learned, regardless of the objects that the tools act upon. The works of Moldovan et al. [224], [225] consider a multi-object scenario in which the relational affordances between objects pairs are exploited to plan a sequence of actions to achieve a desired goal, using probabilistic reasoning. The pairwise interactions are described in terms of the objects relative distance, orientation and contact; however, they do not investigate how these interactions are affected by different geometrical properties of the objects.

An important step toward solving this problem in a more general way, by considering the properties of both tools and affected objects and modeling how those properties influence the effects of specific actions, has been recently made by Gonçalves et al. [221], [222]. In their work, the Bayesian Network (BN) probabilistic model initially proposed in [184] is extended to consider not just actions that are directly applied to an object, but also actions that involve the use of a tool, i.e. an intermediate object. Indeed, this BN model encodes the probabilistic relation between the visual features of both tool and object, the applied action and the resulting effects. A further extension of this work shows how the auto-encoders framework could be applied to this problem to allow for online incremental learning with continuous variables [223]. These works [221]–[223] support computationally the position of Lockman [116] that tool use can emerge gradually from the sensorimotor experience of the agent and from the progressive detection of affordances of the interacting objects (see Section III-C3).

E. Multi-step predictions for action planning

In order to provide the system with the ability to plan a sequence of actions to achieve a desired goal, several one-step predictions have to be chained together to obtain a consistent multi-step prediction, in which the predicted post-conditions of one action match the required pre-conditions of the following action. Interestingly, the computational interpretation of affordances outlined in Section II-B naturally allows for this to be achieved, and indeed many robotics works described hereinafter follow from that formalization.

In [226], a simulated mobile robot learns the environment dynamics in its perceptual space and makes multi-step action plans to achieve goals in a locomotion task. The initial percept space is categorized in an unsupervised manner, i.e. irrespective of the interaction experience of the robot, and the robot learns the *initial-percept → final-percept* mapping. On the other hand, in [227], predicates are discovered from low-level sensory readings during a goal-free exploration and learning phase. However, although objects could be categorized based on their shapes in the sensory level, this information is not used in effect prediction. Moreover, only position features are used to learn "simple affordances of the object" [227, p.886]. Sequences of sensorimotor predictions are used for forward planning in systems based on the SensoriMotor Contingencies Theory (SMCT) [228], [229]; the theory is highly related to O'Regan and Noë's [76] account for visual consciousness, which argues that seeing is a way of acting and vision is a mode of exploration of the world that is mediated by knowledge of sensorimotor contingencies. Although affordances are not explicitly addressed in SMCT, they are "undoubtedly strongly related" [76, p. 945] as both theories oppose detailed internal modeling and emphasize the information pickup aspect of perception [75]. In this context, Hoffman et al. [230], [231] study locomotion affordances provided to quadruped robots and Hogman et al. [232] learn grounded object categories following the same idea, i.e. to use multiple sensorimotor observations obtained from a sequence of actions. Jun Tani's neurorobotics framework [233], [234] has also been effectively used for learning and predicting agent-environment interactions for multi-steps plans, using either multi-component recursive neural networks [233] or stochastic multiple timescale recurrent networks [234]; however, their focus is on learning visuo-proprioceptive sequences rather than learning how to perceive the affordances provided by the environment.

An interesting computational architecture which tackles the problem of using learned affordances for action planning was proposed under the name of object-action complexes (OACs, [235]–[238]). In an attempt to bridge low-level sensorimotor knowledge and high-level symbolic reasoning in autonomous robots, OACs brings together ideas and techniques from behavioral robotics and AI in a coherent architecture. Inspired by Steedman's formalization of affordances [11] (see Section II-B), an OAC is defined as a triple (E, T, M) where E identifies a motor program, T is a function that predicts how the current state of the environment (including the robot) will change after the execution of the motor program, and M

is a statistical measure of the success of the OAC. Both T and M can be learned by the robot through exploration. For each OAC, the function T maps only relevant portions of the state space, and can contain both continuous (sensorimotor) and discrete (symbolic) variables; for example, continuous variables could be the perceptual invariants of an affordance, while discrete variables could be logical predicates. Therefore, within the same control architecture, there can be both low-level OACs related to a basic action (e.g. grasp an object of a certain perceived size) and high-level OACs related to more abstract goals (e.g. take possession of a non reachable object). The hierarchical combination of high-level and low-level OACs could then allow for complex problem solving based on the combination of symbolic planning and learned sensorimotor schemas. Following the OACs approach, Kaiser et al. [239] recently implemented a complete system on the humanoid robot ARMAR-III, in which rule-based whole-body locomotion and manipulation affordances are extracted from segmented RGB-D images. Still, the practical uses of OACs reported in the literature are limited to relatively simple examples, with pre-learned transition rules and pre-defined high-level relations, in which many typical situations that characterize robot behaviors in the real world are not considered (e.g. noisy or erroneous perception, execution failures, unexpected events).

Notably, recent advances in this line of research have been obtained in parallel by two different groups, who propose systems in which the learned affordances and their predictions are exploited to generate and execute symbolic plans in complex real world scenarios. First, Ugur and Piater realize development of the symbolic knowledge in three stages [240]. In the first stage, the robot learns the affordance categories. Then, the system learns logical high-level rules that return a stacking-effect category given the affordances of the involved objects. Finally, these categories and rules are encoded in Planning Domain Definition Language (PDDL), enabling symbolic planning with off-the-shelf AI planners. In a follow-up work, Ugur and Piater close the symbol formation loop by grounding the generated plans in the real-world and by discovering new affordances that appear during plan execution [241]. Second, in the recent work of Antunes et al. [242], one-step predictions based on learned affordances, that are encoded with the Bayesian Network model proposed in [221], [222], are used as probabilistic rules for a probabilistic planner (i.e. PRADA [243]), which can in turn predict the consequences of complex action sequences using approximated inferences over a structured dynamic Bayesian Network representation. In both works [241], [242] the afforded actions are represented as *goals*; this is in line with modern positions in neuroscience (see [134], [135], as discussed in Section IV-B) and developmental psychology [44]. This organization of the motor system allows to separate the action selection process, based on the perceived affordaces, from the physical execution of the action; as an outcome, the robots show a certain degree of flexibility in problem solving, displaying behaviors in the real world that resemble what is observed in human babies (see Section III-C2, and in particular [101]–[104]).

F. Human-robot interaction and communication

Affordance learning and perception can be exploited to improve the quality of natural human-robot interaction, for example by providing robots with a better understading of natural language [244]–[247], by helping the robotic understanding of human actions [248], [249] and by enabling object recognition in terms of the functions that objects afford to humans [250]–[253].

1) Affordances and language: In [244] the robot learns object affordances during verbal interaction with a human caregiver, using Bayesian Networks, and can therefore learn several words-meaning associations. Similar models are proposed in [245] and [246], where the relations between words (nouns, adjectives and verbs) and objects properties (including their affordances) are represented using a set of SVM classifiers. A framework based on Markov Random Field is proposed by Celikkanat et al. [247] in order to model the co-occurrences of actions, object percepts, object affordances and language by constructing a so-called *concept web*; given information in a particular channel, such as an object percept or a word, the corresponding concepts in the web are activated (e.g. the object affordances). In all these systems, the ability to concurrently perceive i) spoken words and ii) objects affordances, helps the robot to remove perceptual ambiguities and to better understand the interaction with a human.

2) Detecting human affordances: A number of works have originated in the computer vision community with respect to the problem of visual detection of 'human' object affordances, i.e. the actions that the objects afford to humans [248]–[253]. In [248] object affordances are detected along with human manipulation actions from video sequences in an integrated framework, using both standard [254] and factorized [255] conditional random fields. The same objective is pursued by the system proposed in Koppula et al. [249], where more complex full-body human activities are considered and object-object interactions are modeled as well. The system acquires RGB-D video streams on the PR2 robot and uses associative Markov networks [256] to model the actions-objects dependencies, where the nodes encode the subactivities and affordances, and the edges correspond to the learned relations between these components. AfNet, the Affordance Network [250], [251], is an open knowledge-base of visual affordances of common household objects, particularly suited for affordance detection from RGB-D images in domestic robots. Image regions with specific properties (e.g. high convexity of a surface) are labeled with specific affordance labels (e.g. contain-ability). In [252] images and meta-data sources extracted from the web are used to learn a knowledge base of object affordances using Markov logic networks [257], a representation that unifies Markov Random Fields and first-order logic. In [253] a large corpus of RGB-D images of over 105 kitchen, workshop and garden tools, in which tool parts are annotated with the relevant affordances (e.g. grasp-ability, pound-ability, cut-ability), is used to train structured random forest classifiers [258] to infer the affordances of local shapes and geometries. While these systems achieve good recognition rates and can be very useful in practice (e.g. for service

and collaborative robots), they do not provide much insights about how humans may use affordance knowledge to better understand the actions of others. Unlike biological agents, these learning systems are only observing the actions of others (i.e. by analyzing images) without doing actions themselves. Indeed, the absence of this mirror mechanism might justify the performance gap which is still present between humans and machines [259], although there is no consensus yet on the existence of such a mechanism in humans [260].

G. Developmental and cognitive modeling

Affordance learning in artificial agents has been included in a number of studies that model either long-term developmental processes (Section V-G1), specific psychology studies (Section V-G2) or cognitive architectures for autonomous robots (Section V-G3).

1) Developmental modeling: In their survey of the ontogeny of tool use, Guerin et al. [82] formulate concrete recommendations on how to approach the modeling of general mechanisms of sensorimotor development and knowledge representation, including actions-object relationships. Cangelosi et al. [261], [262] identify the key challenges in developmental robotics, and propose a practical roadmap which includes a series of milestones such as action and affordance learning, social learning and cognitive integration. Hart and Grupen [263] propose a developmental system in which the sensorimotor space of a manipulator robot is self-organized by applying Piaget's accommodation and assimilation mechanisms [101]. Fitchtl et al. [264] use predictions of action effects as inputs in predicting effects of other actions. They show that a speed boost can be achieved in the initial phases of learning, especially if the learning problem is hard. Ugur et al. [265] show that such a bootstrapping effect can be obtained not only if the pre-learned affordances are used as inputs for predicting more complex affordances, but also if the training objects are actively selected to be maximally different based on their learned affordances. In a follow-up work, Ugur et al. [266] propose an active learning method which uses intrinsic motivation for selecting actions and feature selection for establishing links, in order to let the system discover the hierarchical affordance structure automatically. As a result, the robot learns simple poke affordances first, and uses the learning outputs to acquire more complex stacking affordances later on. This 're-use' idea has been exploited also by Wang et al. [267], who consider the *transfer* of learned affordances from previously explored objects to new similar objects. The system described by Worgotter et al. [268], on the other hand, transfers affordance knowledge between objects and between actions, in order to replace the components that are missing in the immediate environment of the robot during plan execution. Ivaldi et al. [269], [270] increase object recognition performance through a socially guided intrinsic motivation system that actively selects objects to explore, actions to execute and caregivers to interact with; similar strategies that combine robot motor exploration with social guidance could be used to increase the speed and efficacy of affordance learning as well. Indeed, Ugur et al. [21] describe a staged

developmental framework in which a wide range of skills (execution of behavioral primitives, affordance perception, emulation, imitation) emerge through real-world interactions, inlcuding robot-caregiver interactions. They show that the use of simple-to-complex perceptual information (i.e., first tactile, then visual and finally social cues) and the progressive shift from self-exploration strategies (during affordance learning) to observational learning (through emulation) are necessary and sufficient for the progressive development of the targeted sensorimotor skills. Indeed, this study offers a computational account of several observations in Developmental Psychology (see Section III-C2 in particular), supporting the idea that affordance learning allows goal emulation, and that both precede the emergence of imitation abilities in human infants [93], [96], [106].

2) Cognitive modeling: Cognitive and neuro-robotics simulations have been proposed to model the object affordance and stimulus responsibility effects [271], [272]. These models are based on the TRoPICALS [271], a neuro-robotics cognitive architecture that captures the neuro-cognitive mechanisms underlying several distinct affordance compatibility effects. Specifically, the study of Caligiore et al. [272] investigates the affordance competition between target-distractor objects with the iCub robot, reproducing the experimental results of Ellis et al. [60] (see Section III-B). The neuro-robotics model explains these results on the basis of the detailed neural mechanisms that could underlie the interference/facilitation effects caused by the perception of the distractor contextually with the target. This explanation is based on a novel idea according to which the prefrontal cortex might play a double role in its top-down guidance of action selection: (a) producing a positive bias directed to trigger the actions requested by the experimental task; (b) producing a negative bias directed to inhibit the action evoked by the distractor.

3) Cognitive architectures: A number of robotic cognitive architectures that model a wide range of abilities explicitly addressed affordances in encoding perception-action relations. In LIDA (Learning Intelligent Distribution Agent [273]), which models a broad spectrum of cognitive capabilities, affordances are encoded in the connections that link objects, object categories and actions in the Perceptual Associative Memory that is used in the understanding phase, prior to the attending and the action phases. In CLARION (Connectionist Learning with Adaptive Rule Induction On-line [274]) the actions effects are observed to learn low-level connectionist representations using reinforcement learning, and to refine high-level rule-based symbolic representation. The OACs framework [235], described in Section V-E, is part of the three-levels cognitive control architecture that was realized in the PACO-PLUS project [275]. The affordance model proposed in [184], reported in Section V-B3, is used to keep object-action-effect triplets in the Procedural Memory Module of the cognitive architecture [276, p. 156] of the iCub robot [277]. More recently, this architecture has been extended to allow the iCub robot to perform tasks requested by a human user with natural language (as outlined in [242], see Section V-E), bringing together action-centered language understanding [278], learning and perception of the affordances of tools and target objects

[221], [222], and complex action execution routines [218].

H. Summary and discussion

Overall, most of the reported works implement an idea which is central in Ecological Psychology: that performing actions is crucial for the visual perception of important object properties. Actions can be performed either during perception (like in the early work of [166], discussed in Section V-A, or in the more general approaches of active vision [279] and active perception [280]) or, most interestingly for the present discussion, before perception; indeed, this latter approach is identified with affordance learning, i.e. learning what is the most relevant visual information to pick-up, given the sensorimotor abilities of the agent and its sensorimotor experiences. Affordance learning is the necessary pre-requisite for affordance perception and for prediction of the action effects (as suggested also by the observations in developmental psychology, see Section III-C); the challenge for roboticists is to design systems that can perform this learning efficiently. To do so, one important aspect is how to represent the data: in Section V-B we discuss the evolution of such representations in the state of the art, which culminates in the use of probabilistic representations (see Section V-B3) and in the automatic discovery of the representations based on the robot experiences (see Section V-B4). In Section V-C we provide a brief discussion on grasp affordances; this is a huge topic per-se, as it overlaps with robotic manipulation, and what we do in this survey is to highlight what class of approaches has borrowed more from the lessons of Ecological Psychology. Clearly, the complexity of the robot learning experiences determines the range of situations that the robot will be able to deal with: scaling from single-object to multi-objects experiences is an important step beyond, that can permit efficient and flexible robotic tool use, as described in Section V-D. The works reported in Sections from V-A to V-D deal with the learning of increasingly complex affordances; the learned models endow robots with the ability to i) perceive affordances and ii) predict the effects of the afforded actions. But how to put these abilities to use in real tasks? In Section V-E we examine works in which these abilities are employed for action planning, allowing for problem solving in real scenarios. Then, in Section V-F we explore another dimension in which affordance knowledge can be useful for robots: i.e. to better interact with people through an enhanced understanding of verbal communications and of the experienced situations. Finally, in Section V-G we discuss the efforts in the robotics and computer science communities to include affordance learning and perception into more general frameworks. Affordance learning in humans is made efficient by the combination of different learning strategies (e.g. self-exploration, observation, imitation), as discussed in Section III-C; moreover, as pointed out in Section II-D, affordance perception is integrated with several other cognitive processes. Apart from the results in psychology and neuroscience, an important computational support to these hypothesis has been provided by robotics studies as well, that either replicated specific experiments (compare Section V-G2 to Section III-B)

or included affordance perception into developmental (see Section V-G1) and cognitive (see Section V-G3) architectures.

VI. DISCUSSION

Perhaps the most important insight made explicit by the affordance concept is that perception is deeply influenced by action; notably, this is a consequence of the fact that, indeed, the ultimate goal of perception is not to re-construct the environment, but to allow the agent to effectively *act* on the environment. Therefore, from a computational perspective, the goal of a robot vision system should not be to create a fully detailed representation of the environment, but instead to understand what is the minimal information that has to be extracted from the visual stream to allow the robot to successfully perform actions and achieve tasks. Because such minimal information depends on both the agent (i.e. the robot, in this case) and the environment, to discover what that information is (or in other terms, to find out the most appropriate data representations) the agent needs to do actions and to perceive the effects through its own sensors: i.e. affordance learning is necessary. Interestingly, the associative mechanisms that result from affordance learning are tremendously-powerful enablers of sophisticated sensorimotor interaction and cognition because they can shortcut (expensive and brittle) reasoning processes. Initially, it takes a child some effort to learn how to properly grasp a cup by the handle. With time, the child stops thinking about it. On seeing the handle of a cup, a motor schema related to the grasping of that specific type of handles will be subconsciously activated (or, in neuro-physiology terms, activated sub-threshold), before she even realizes that the object is a cup; the motor system prepares for action, allowing for a fast execution in case she consciously decides to grasp the cup. Moreover, these perception-action couplings that are progressively learned during development allow the children to predict action consequences, leading to the emergence of action planning and problem solving skills. Evidence supporting these insights comes from experiments and observations in both psychology (see Section III) and neuroscience (see Section IV).

In robotics, the theory of affordances has been widely used as a source of inspiration. However, only certain aspects have been used, and typically in isolation. Although large-scale projects have studied the application of affordances to robot control and numerous meetings on the topic have been organized in vision and robotics conferences, there is yet no unified view on how (i.e. at what level and to what extent) the theory of affordances should be applied in autonomous robotics. Indeed, a number of open issues are still present. Should affordance-based controllers be considered only at the reactive level of a robot architecture? Or should they be fully integrated in the cognitive architectures, directly affecting more complex behaviors such as imitation, problem solving and verbal communication? Also, how to represent affordances so that they can be effectively used in complex scenarios? Should the sensory percepts be directly linked to the motor coding, or they should be first grouped into symbolic categories? Should the output of affordance perception be

the afforded action (or a set of afforded actions), or the effects generated by the action, or both? And how to represent those effects? Some effects might be salient changes in the environment (e.g. motion of an object), but how to represent the effect of sitting on a chair? How different affordances (e.g. traversability of a space, rollability of an object, sit-ability of a chair) can be represented within the same model? Moreover, it is commonly agreed that affordance relations have to be learned by the agent through interaction with the environment to be really representative of its sensorimotor capabilities. But how? Many different learning algorithms have been applied in different studies; however, it is still not evident whether any of them provides clear advantages with respect to the others. A more interesting dimension, indeed, is that of the exploration strategies employed by the agent (e.g. motor babbling, active learning, reinforcement learning, internal motivation and curiosity, staged developmental learning).

Although we do not dare to provide any conclusive answer to these very broad and challenging research questions, we do hazard to suggest a few promising directions. The first is related to the use of probabilistic representations, which seem to be the best candidate to model the intrinsic uncertainty which characterize robot actions and perceptions. The second is the idea of staged developmental learning, in which what is learned in one stage informs the subsequent stage, both in terms of the exploration strategy to adopt and on how to organize the collected sensorimotor information. The third is the use of affordance predictions to ground the rules of logical reasoning systems, that could lead to effective action planning and problem solving in real robots; in more general terms, this seems to be an interesting direction to bridge the gap between AI and robotics approaches to complex problem solving.

VII. CONCLUSIONS

The concept of affordances has inspired multidisciplinary research in many fields, and the findings have been explored in many diverse (not always consistent) ways. This paper puts many of such research endeavors and results in perspective, particularly in the fields of psychology, neuroscience and robotics, and establishes connections across such diverse disciplines. We offer a structured and comprehensive report of works from these different fields; short summaries and discussions can be found in each section, to help the reader to extract the related core information. Moreover, we conclude the paper with a general discussion in which we also outline a number of open questions related to the use of the affordance concept in robotics and we suggest a few interesting research directions. Overall, from a robotics and computational perspective, the studies of affordances in psychology and neuroscience do once again reveal the extremely tight coupling between perception and action, that is reflected in the existence of shared representations that also integrate goals and effects. Such representations provide a bridge between sensorimotor loops and more abstract and symbolic knowledge, enabling sophisticated cognitive processes that support complex behaviors, learning and memory.

For the future, we believe there is still a lot to be gained from this challenging and multidisciplinary dialogue between psychology, neuroscience and robotics. Indeed, research on affordances can have a considerable impact on society in multiple domains: it can lead to a better understanding of human behaviors and neural functions, it can improve the design of intelligent robots to be employed in human environments, it can offer a new perspective for the realization of agent-centered automated reasoning systems. Despite the important results achieved in the different communities, there are still many open questions; we are confident, however, that the tight collaboration between experts in these different scientific areas can lead to a better understanding of the core principles behind affordance learning and exploitation, that will in turn facilitate the development of useful applications based on this concept.

REFERENCES

[1] J. J. Gibson, *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

[2] ——, "The theory of affordances," in *Perceiving, Acting, and Knowing: Toward an ecological psychology*, R. Shaw and J. Bransford, Eds. Lawrence Erlbaum Associates, 1977, pp. 67–82.

[3] ——, *The Ecological Approach to Visual Perception.* New Jersey, USA: Lawrence Erlbaum Associates, 1986, original work published in 1979.

[4] T. E. Horton, A. Chakraborty, and R. St. Amant, "Affordances for robots: a brief survey," *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, no. 2, pp. 70–84, 2012.

[5] K. S. Jones, "What is an affordance?" *Ecological Psychology*, vol. 15, no. 2, pp. 107–114, 2003.

[6] J. J. Gibson and L. E. Crooks, "A theoretical field-analysis of automobile-driving," *American Journal of Psychology*, vol. 51, pp. 453–471, 1938.

[7] M. Turvey, "Affordances and prospective control: An outline of the ontology," *Ecological Psychology*, vol. 4, pp. 172–187, 1992.

[8] S. Z. Caiani, "Extending the notion of affordance," *Phenomenology and the Cognitive Sciences*, vol. 13, no. 2, pp. 275–293, 2014.

[9] T. A. Stoffregen, "Affordances as properties of the animal environment system," *Ecological Psychology*, vol. 15, no. 2, pp. 115–134, 2003.

[10] A. Chemero, "An outline of a theory of affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[11] M. Steedman, "Plans, affordances, and combinatory grammar," *Linguistics and Philosophy*, vol. 25, 2002. [Online]. Available: ftp://ftp.cogsci.ed.ac.uk/pub/steedman/affordances/pelletier.pdf

[12] E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.

[13] D. A. Norman, *The Psychology of Everyday Things.* Basic Books, 1988.

[14] ——, "Affordance, conventions, and design," *Interactions*, vol. 6, no. 3, pp. 38–42, 1999.

[15] J. McGrenere and W. Ho, "Affordances: Clarifying and evolving a concept," in *Proceedings of the Graphics Interface.* Toronto: Canadian Human-Computer Communications Society, 2000, pp. 179–186.

[16] C. F. Michaels, "Affordances: Four points of debate," *Ecological Psychology*, vol. 15, no. 2, pp. 135–148, 2003.

[17] H. Heft, "Affordances, dynamic experience, and the challenge of reification," *Ecological Psychology*, vol. 15, no. 2, pp. 149–80, 2003.

[18] E. Ugur and E. Şahin, "Traversability: A case study for learning and perceiving affordances in robots," *Adaptive Behavior*, vol. 18, no. 3-4, 2010.

[19] E. Ugur, E. Oztop, and E. Sahin, "Goal emulation and planning in perceptual space using learned affordances," *Robotics and Autonomous Systems*, vol. 59, no. 7, pp. 580–595, 2011.

[20] E. Ugur, Y. Nagai, H. Celikkanat, and E. Oztop, "Parental scaffolding as a bootstrapping mechanism for learning grasp affordances and imitation skills," *Robotica*, vol. 33, no. 05, pp. 1163–1180, 2015.

[21] E. Ugur, Y. Nagai, E. Sahin, and E. Oztop, "Staged development of robot skills: Behavior formation, affordance learning and imitation with motionese," *Transaction on Autonomous Mental Development (TAMD)*, vol. 7, no. 2, pp. 119–139, 2015.

[22] E. Ugur, M. Dogar, M. Cakmak, and E. Sahin, "The learning and use of traversability affordance using range images on a mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 1721–1726.

[23] M. R. Doğar, M. Çakmak, E. Ugur, and E. Şahin, "From primitive behaviors to goal-directed behavior using affordances," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2007, pp. 729–734.

[24] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "Curiosity-driven learning of traversability affordance on a mobile robot," in *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*. IEEE, 2007, pp. 13–18.

[25] M. Cakmak, M. Dogar, E. Ugur, and E. Sahin, "Affordances as a framework for robot control," *Proceedings of The 7th International Conference on Epigenetic Robotics, EpiRob07*, 2007.

[26] M. R. Dogar, E. Ugur, E. Sahin, and M. Cakmak, "Using learned affordances for robotic behavior development," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 3802–3807.

[27] E. Ugur, E. Sahin, and E. Oztop, "Affordance learning from range data for multi-step planning," in *Proceedings of The 9th International Conference on Epigenetic Robotics, EpiRob'09*. Venice, Italy: Lund University Press, 2009, pp. 177–184.

[28] E. Ugur, E. Şahin, and E. Oztop, "Predicting future object states using learned affordances," in *Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on*. IEEE, 2009, pp. 415–419.

[29] N. Dag, I. Atil, S. Kalkan, and E. Sahin, "Learning affordances for categorizing objects and their properties," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3089–3092.

[30] I. Atıl, N. Dag, S. Kalkan, and E. Sahin, "Affordances and emergence of concepts," *10th International Conference on Epigenetic Robotics*, 2010.

[31] E. Ugur, E. Şahin, and E. Oztop, "Unsupervised learning of object affordances for planning in a mobile manipulation platform," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4312–4317.

[32] E. Ugur, E. Oztop, and E. Şahin, "Going beyond the perception of affordances: Learning how to actualize them through behavioral parameters," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4768–4773.

[33] E. Ugur, E. Sahin, and E. Oztop, "Self-discovery of motor primitives and learning grasp affordances," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 3260–3267.

[34] K. Koflka, "Principles of gestalt psychology," *New York: Har*, 1935.

[35] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W. H. Freeman, 1982, (posthumous).

[36] J. L. Bermúdez, *Cognitive science: An introduction to the science of the mind*. Cambridge University Press, 2014.

[37] A. H. Vera and H. A. Simon, "Situated action: A symbolic interpretation," *Cognitive science*, vol. 17, no. 1, pp. 7–48, 1993.

[38] A. Chemero and M. T. Turvey, "Gibsonian affordances for roboticists," *Adaptive Behavior*, vol. 15, no. 4, pp. 473–480, 2007.

[39] U. Neisser, "Multiple systems: A new approach to cognitive theory," *The European Journal of Cognitive Psychology*, vol. 6, pp. 225–241, 1994.

[40] W. Kintsch, *Comprehension: A paradigm for cognition*. Cambridge university press, 1998.

[41] U. Neisser, *Cognition and Reality: Principles and Implications of Cognitive Psychology*, ser. 0716704773. W. H. Freeman and Co, 1976.

[42] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "What the frog's eye tells the frog's brain," *Proceedings of the IRE*, vol. 47, no. 11, pp. 1940–1951, 1959.

[43] J. K. Witt, "Action's effect on perception," *Current Directions in Psychological Science*, vol. 20(3), pp. 201–206, 2011.

[44] C. von Hofsten, "An action perspective on motor development," *Trends in Cognitive Sciences*, vol. 8, pp. 266–272, 2004.

[45] T. Collett, "Stereopsis in toads," *Nature*, vol. 267, pp. 349–351, 1977.

[46] A. Lock and T. Collett, "A toads devious approach to its prey: A study of some complex uses of depth vision," *Journal of Comparative Physiology*, vol. 131, pp. 179–189, 1979.

[47] D. Ingle and J. Cook, "The effects of viewing distance upon size preference of frogs for prey," *Vision Research*, vol. 17, pp. 1009–1019, 1977.

[48] V. McCabe, "Invariants and affordances: An analysis of species-typical information," *Ethology and Sociobiology*, vol. 3, no. 2, pp. 79–92, 1982.

[49] W. H. Warren, "Perceiving affordances: Visual guidance of stair climbing," *Journal of Experimental Psychology*, vol. 105, no. 5, pp. 683–703, 1984.

[50] W. H. Warren and S. Whang, "Visual guidance of walking through apertures: body-scaled information for affordances," *Journal of Experimental Psychology*, vol. 13, no. 3, pp. 371–383, 1987.

[51] L. S. Mark, "Eyeheight-scaled information about affordances: A study of sitting and stair climbing," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, no. 3, pp. 361–370, 1987.

[52] S. Ishak, J. M. Franchak, and K. E. Adolph, "Perceptionaction development from infants to adults: Perceiving affordances for reaching through openings," *Journal of Experimental Child Psychology*, vol. 117, pp. 92–105, 2014.

[53] E. J. Gibson, G. Riccio, M. A. Schmuckler, T. A. Stoffregen, D. Rosenberg, and J. Taromina, "Detection of the traversability of surfaces by crawling and walking infants," *Journal of Experimental Psychology*, vol. 13, no. 4, pp. 533–544, 1987.

[54] J. M. Kinsella-Shaw, B. Shaw, and M. T. Turvey, "Perceiving walk-on-able slopes," *Ecological Psychology*, vol. 4, no. 4, pp. 223–239, 1992.

[55] A. Chemero, "What events are," *Ecological Psychology*, vol. 12, no. 1, pp. 37–42, 2000.

[56] A. Chemero, C. Klein, and W. Cordeiro, "Events as changes in the layout of affordances," *Ecological Psychology*, vol. 15, no. 1, pp. 19–28, 2003.

[57] S. Cornus, G. Montagne, and M. Laurent, "Perception of a stepping-across affordance," *Ecological Psychology*, vol. 11, no. 4, pp. 249–267, 1999.

[58] R. Oudejans, C. Michaels, B. VanDort, and E. Frissen, "To cross or not to cross: The effect of locomotion on street-crossing behavior," *Ecological Psychology*, vol. 8, no. 3, pp. 259–267, 1996.

[59] M. Tucker and R. Ellis, "On the relations between seen objects and components of potential actions," *Journal of Experimental Psychology: Human perception and performance*, vol. 24, no. 3, p. 830, 1998.

[60] R. Ellis, M. Tucker, E. Symes, and L. Vainio, "Does selecting one visual object from several require inhibition of the actions associated with nonselected objects?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 33, pp. 670–691, 2007.

[61] J. Goslin, T. Dixon, M. Fischer, A. Cangelosi, and R. Ellis, "Electrophysiological examination of embodiment in vision and action," *Psychological Science*, vol. 23, no. 2, pp. 152–157, 2012.

[62] M. Costantini, E. Ambrosini, G. Tieri, C. Sinigaglia, and G. Committeri, "Where does an object trigger an action? an investigation about affordances in space," *Experimental brain research*, vol. 207, no. 1-2, pp. 95–103, 2010.

[63] D. N. Bub and M. E. Masson, "Grasping beer mugs: on the dynamics of alignment effects induced by handled objects." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 36, no. 2, p. 341, 2010.

[64] E. Ambrosini and M. Costantini, "Handles lost in non-reachable space," *Experimental brain research*, vol. 229, no. 2, pp. 197–202, 2013.

[65] K. C. Soska, K. E. Adolph, and S. P. Johnson, "Systems in development: motor skill acquisition facilitates three-dimensional object completion," *Developmental psychology*, vol. 46, no. 1, p. 129, 2010.

[66] E. W. Bushnell and J. P. Boudreau, "Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development," *Child Development*, vol. 64, no. 4, pp. 1005–1021, 1993.

[67] K. S. Bourgeois, A. W. Khawar, S. A. Neal, and J. J. Lockman, "Infant manual exploration of objects, surfaces, and their interrelations," *Infancy*, vol. 8, no. 3, pp. 233–252, 2005.

[68] H. L. Pick, "Eleanor j. gibson: Learning to perceive and perceiving to learn." *Developmental Psychology*, vol. 28, no. 5, p. 787, 1992.

[69] A. Szokolszky, "An interview with Eleanor Gibson," *Ecological Psychology*, vol. 15, no. 4, pp. 271–281, 2003.

[70] E. J. Gibson, *An odyssey in learning and perception*. Mit Press, 1994.

[71] J. J. Gibson, "The senses considered as perceptual systems." 1966.

[72] K. E. Adolph and K. S. Kretch, "Gibsons theory of perceptual learning," 2015.

[73] E. J. Gibson and H. Levin, *The psychology of reading*. The MIT press, 1975.

[74] E. J. Gibson, "Perceptual learning in development: Some basic concepts," *Ecological Psychology*, vol. 12, no. 4, pp. 295–302, 2000.

[75] ——, "The world is so full of a number of things: On specification and perceptual learning," *Ecological Psychology*, vol. 15, no. 4, pp. 283–288, 2003.

[76] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behavioral and brain sciences*, vol. 24, no. 05, pp. 939–973, 2001.

[77] E. J. Gibson, "Has psychology a future?" *Psychological Science*, vol. 5, no. 2, pp. 69–76, 1994.

[78] ——, "Principles of perceptual learning and development." 1969.

[79] D. K. James, "Fetal learning: A critical review," *Infant and child development*, vol. 19, pp. 45–54, 2010.

[80] A. Pitti, Y. Kuniyoshi, M. Quoy, and P. Gaussier, "Modeling the minimal newborn's intersubjective mind: the visuotopic-somatotopic alignment hypothesis in the superior colliculus," *PloS one*, vol. 8, no. 7, p. e69474, 2013.

[81] J. Piaget and B. Inhelder, *The Psychology of the Child*. New York, USA: Basic Books, 1966.

[82] F. Guerin, N. Kruger, and D. Kraft, "A survey of the ontogeny of tool use: from sensorimotor experience to planning," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 1, 2013.

[83] J. W. Santrock, "Motor, sensory, and perceptual development," *A topical approach to life-span development. McGraw-Hill Higher Education, Boston*, pp. 172–205, 2008.

[84] C. Von Hofsten, "Eye–hand coordination in the newborn." *Developmental psychology*, vol. 18, no. 3, p. 450, 1982.

[85] D. A. Rosenbaum, *Human Motor Control*. London, UK: Academic Press, 1991.

[86] P. Rochat and P. Rochat, *The infant's world*. Harvard University Press, 2009.

[87] J. G. Bremner, *Infancy*. Malden, MA, USA: Blackwell Publishing, 1988, 2nd edition, 1994.

[88] I. Lézine, "The transition from sensorimotor to earliest symbolic function in early development." *Research publications-Association for Research in Nervous and Mental Disease*, vol. 51, pp. 221–232, 1972.

[89] M. R. Fiorentino, *A basis for sensorimotor development-normal and abnormal: the influence of primitive, postural reflexes on the development and distribution of tone*. Charles C Thomas Pub Limited, 1981.

[90] M. W. Casby, "The development of play in infants, toddlers, and young children," *Communication Disorders Quarterly*, vol. 24, no. 4, pp. 163–174, 2003.

[91] D. Rosenblatt, "Developmental trends in infant play," *Biology of play*, pp. 33–44, 1977.

[92] E. J. Gibson, "Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge," *Annual review of psychology*, vol. 39, no. 1, pp. 1–42, 1988.

[93] J. Provasi, C. D. Dubon, and H. Bloch, "Do 9- and 12-month-olds learn means-ends relation by observing?" *Infant Behavior and Development*, vol. 24, no. 2, pp. 195–213, Feb. 2001.

[94] M. Carpenter, K. Nagell, and M. Tomasello, "Social cognition, joint attention, and communicative competence from 9 to 15 months of age," *Monographs of the Society for Research in Child Development*, vol. 63, no. 4, 1998.

[95] B. Elsner and G. Aschersleben, "Do I get what you get? Learning about the effects of self-performed and observed actions in infancy," *Consciousness and Cognition*, vol. 12, pp. 732–751, 2003.

[96] S. C. Want and P. L. Harris, "How do children ape? Applying concepts from the study of non-human primates to the developmental study of imitation in children," *Developmental Science*, vol. 5, pp. 1–13, 2002.

[97] R. W. White, "Motivation reconsidered: the concept of competence." *Psychological review*, vol. 66, no. 5, p. 297, 1959.

[98] D. E. Berlyne, "Conflict, arousal, and curiosity." 1960.

[99] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *Evolutionary Computation, IEEE Transactions on*, vol. 11, no. 2, pp. 265–286, 2007.

[100] B. Elsner and B. Hommel, "Effect anticipation and action control," *Journal of Experimental Psychology*, vol. 27, pp. 229–240, 2003.

[101] J. Piaget, *The Origins of Intelligence in Children*. New York, USA: International University Press, 1952.

[102] M. Tomasello, *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press, 1999.

[103] P. Willatts, "Development of means-end behavior in young infants: Pulling a support to retrieve a distant object," *Developmental Psychology*, vol. 35, pp. 651–666, 1999.

[104] ——, "The stage IV infant's solution of problems requiring the use of supports," *Infant Behavior and Development*, vol. 7, pp. 125–134, 1984.

[105] M. Tomasello, "Do apes ape," *Social learning in animals: The roots of culture*, pp. 319–346, 1996.

[106] C. Huang and T. Charman, "Gradations of emulation learning in infants: imitation of actions on objects," *Journal of Experimental Psychology*, vol. 92, no. 3, pp. 276–302, 2005.

[107] M. Paulus, S. Hunnius, M. Vissers, and H. Bekkering, "Imitation in infancy: Rational or motor resonance?" *Child development*, vol. 82, no. 4, pp. 1047–1057, 2011.

[108] M. van Elk, H. T. van Schie, S. Hunnius, C. Vesper, and H. Bekkering, "You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy," *Neuroimage*, vol. 43, no. 4, pp. 808–814, 2008.

[109] R. Brand, D. Baldwin, and L. Ashburn, "Evidence for 'motionese': Modifications in mothers infant-directed action," *Developmental Science*, vol. 5, pp. 72–83, 2002.

[110] Y. Nagai, "From Bottom-Up Visual Attention to Robot Action Learning," in *Proc. of the 8th IEEE International Conference on Development and Learning*, 2009.

[111] J. A. Ungerer, P. R. Zelazo, R. B. Kearsley, and K. O'Leary, "Developmental changes in the representation of objects in symbolic play from 18 to 34 months of age," *Child Development*, pp. 186–195, 1981.

[112] H. Takeshita, "Development of combinatory manipulation in chimpanzee infants (pan troglodytes)," *Animal cognition*, vol. 4, no. 3-4, pp. 335–345, 2001.

[113] M. Ikuzawa, "Developmental diagnostic tests for children," *Nakanishiya, Kyoto*, 2000.

[114] A. M. Damast, C. S. Tamis-LeMonda, and M. H. Bornstein, "Mother-child play: Sequential interactions and the relation between maternal beliefs and behaviors," *Child Development*, vol. 67, no. 4, pp. 1752–1766, 1996.

[115] E. Bates, *The emergence of symbols: Cognition and communication in infancy*. Academic Press, 2014.

[116] J. J. Lockman, "A perception–action perspective on tool use development," *Child Dev.*, vol. 71, no. 1, pp. 137–44, 2000.

[117] F. J. Varela, E. Thompson, and E. Rosch, *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press, 1991.

[118] J. Fagard, L. Rat-Fischer, and J. K. O'Reagan, "The emergence of use of a rake-like tool: a longitudinal study in human infants," *Front. Psychol.*, vol. 5, no. 491, 2014.

[119] W. Kohler, *The Mentality of Apes*. New York, NY: Harcourt, Brace & Company Inc., 1925.

[120] V. Argiro and K. Pastra, "A multimodal dataset of spontaneous speech and movement production on object affordances," *Scientific Data*, vol. 3, 2016.

[121] L. G. Ungerleider and M. Mishkin, "Two cortical visual systems," in *Analysis of visual behavior*. Cambridge MA: MIT Press, 1982, pp. 549–586.

[122] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action," *Trends Neurosci*, vol. 15, pp. 20–25, 1992.

[123] J. C. Culham and K. F. Valyear, "Human parietal cortex in action," *Current opinion in neurobiology*, vol. 16, no. 2, pp. 205–212, 2006.

[124] J. Norman, "Two visual systems and two theories," *Behavioral and Brain Sciences*, vol. 25, pp. 73–144, 2002.

[125] A. D. Milner and M. A. Goodale, *The visual brain in action*. England, 1995, vol. 27.

[126] A. D. Milner, D. I. Perrett, R. S. Johnson, P. J. Benson, T. R. Jordan, D. W. Heeley, D. Bettuci, F. Mortara, R. Mutani, E. Terazzi, and D. L. W. Davidson, "Perception and action in 'visual form agnosia'," *Brain*, vol. 114, pp. 405–428, 1991.

[127] L. G. Nowak and J. Bullier, *Extrastriate Cortex in Primates*. Boston, MA: Springer US, 1997, ch. The Timing of Information Transfer in the Visual System, pp. 205–241.

[128] E. Oztop, H. Imamizu, G. Cheng, and M. Kawato, "A computational model of anterior intraparietal (AIP) neurons," *Neurocomputing*, vol. 69, no. 10, pp. 1354–1361, 2006.

[129] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez, and L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?" *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1847–1871, 2013.

[130] J. Norman, "Ecological psychology and the two visual systems: Not to worry!" *Ecological Psychology*, vol. 13, no. 2, pp. 135–145, 2001.

[131] G. Rizzolatti and M. Matelli, "Two different streams form the dorsal visual system: anatomy and functions," *Exp Brain Res*, vol. 153(2), pp. 146–57, 2003.

[132] L. Fogassi, V. Gallese, L. Fadiga, G. Luppino, M. Matelli, and G. Rizzolatti, "Coding of peripersonal space in inferior premotor cortex (area f4)," *J Neurophysiol*, vol. 76(1), pp. 141–57, 1996.

[133] G. Rizzolatti, L. Fadiga, L. Fogassi, and V. Gallese, "The space around us," *Science*, vol. 277, pp. 190–191, 1997.

[134] G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli, "Functional organization of inferior area 6 in the macaque monkey. ii. area f5 and the control of distal movements," *Exp Brain Res*, vol. 71(3), pp. 491–507, 1988.

[135] M. A. Arbib, "Schemas for the temporal organization of behaviour," *Hum Neurobiol*, vol. 4, pp. 63–72, 1981.

[136] N. Bernstein, *The co-ordination and regulation of movements*. Oxford: Pergamon Press, 1967.

[137] A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, and G. Rizzolatti, "Object representation in the ventral premotor cortex (area f5) of the monkey," *J Neurophysiol*, vol. 78(4), pp. 2226–30, 1997.

[138] G. Rizzolatti and G. Luppino, "The cortical motor system," *Neuron*, vol. 31(6), pp. 889–901, 2001.

[139] R. P. Dum and P. Strick, "Frontal lobe inputs to the digit representations or the motor areas on the lateral surface of the hemisphere," *J Neurosci*, vol. 25(6), pp. 1375–86, 2005.

[140] T. Brochier and M. A. Umiltá, "Cortical control of grasp in non-human primates," *Curr Opin Neurobiol*, vol. 17(6), pp. 637–43, 2007.

[141] G. Rizzolatti and L. Fadiga, "Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area f5)," *Novartis Found Symp*, vol. 218, pp. 81–95, discussion 95–103, 1998.

[142] G. D. Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, "Understanding motor events: a neurophysiological study," *Exp Brain Res*, vol. 91(1), pp. 176–80, 1992.

[143] V. Raos, M. A. Umiltá, A. Murata, L. Fogassi, and V. Gallese, "Functional properties of grasping-related neurons in the ventral premotor area f5 of macaque monkey," *J Neurophysiol*, vol. 95(2), pp. 709–29, 2006.

[144] A. Murata, V. Gallese, G. Luppino, M. Kaseda, and H. Sakata, "Selectivity for the shape, size, and orientation of object for grasping in neurons of monkey parietal area aip," *J Neurophysiol*, vol. 83(5), pp. 2580–601, 2000.

[145] S. Rozzi, P. F. Ferrari, L. Bonini, G. Rizzolatti, and L. Fogassi, "Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas," *Eur J Neurosci*, vol. 28(8), pp. 1569–88, 2008.

[146] J. Grezes, M. Tucker, J. Armony, R. Ellis, and R. E. Passingham, "Objects automatically potentiate action: an fmri study of implicit processing," *Eur. J. Neurosci.*, vol. 17, no. 12, pp. 2735–2740, 2003.

[147] G. Buccino, M. Sato, L. Cattaneo, F. Roda, and L. Riggio, "Broken affordances, broken objects: A {TMS} study," *Neuropsychologia*, vol. 47, no. 14, pp. 3074–3078, 2009.

[148] M. Franca, L. Turella, R. Canto, N. Brunelli, L. Allione, N. G. Andreasi, M. Desantis, D. Marzoli, and L. Fadiga, "Corticospinal facilitation during observation of graspable objects: a transcranial magnetic stimulation study," *PLoS One*, vol. 7, no. 11, 2012.

[149] E. Bartoli, L. Maffongelli, M. Jacono, and A. D'Ausilio, "Representing tools as hand movements: early and somatotopic visuomotor transformations," *Neuropsychologia*, vol. 61, pp. 335–44, 2014.

[150] P. Cardellicchio, C. Sinigaglia, and M. Costantini, "The space of affordances: a tms study," *Neuropsychologia*, vol. 49 (5), pp. 1369–1372, 2011.

[151] ——, "Grasping affordances with the others hand: A tms study," *Social Cognitive and Affective Neuroscience*, vol. 8, no. 4, pp. 455–459, 2013.

[152] A. E. Patla and M. A. Goodale, "Obstacle avoidance during locomotion is unaffected in a patient with visual form agnosia," *NeuroReport*, vol. 8, no. 1, pp. 165–168, 1996.

[153] G. Vingerhoets, K. Vandamme, and A. Vercammen, "Conceptual and physical object qualities contribute differently to motor affordances," *Brain and Cognition*, vol. 69, no. 3, pp. 481–489, 2009.

[154] M. A. Goodale, "Action without perception in human vision," *Cognitive Neuropsychology*, vol. 25, no. 7-8, pp. 891–919, 2008.

[155] G. Humphreys, "Objects, affordances ... action !!!" *The Psychologist*, vol. 14, no. 8, p. 5, 8 2001.

[156] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti, "Action recognition in the premotor cortex," *Brain*, vol. 119, pp. 593–609, 1996.

[157] S. Treue, "Neural correlates of attention in primate visual cortex," *Trends in Neuroscience*, vol. 24(5), pp. 295–300, 2001.

[158] P. Cisek and J. F. Kalaska, "Neural mechanisms for interacting with a world full of action choices," *Annu. Rev. Neurosci.*, vol. 33, pp. 269–98, 2010.

[159] A. Clark, *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press, 1997.

[160] V. Gallese, "A neuroscientific grasp of concepts: from control to representation," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 358, no. 1435, pp. 1231–1240, 2003.

[161] R. A. Brooks, "Elephants don't play chess," *Robotics and Autonomous Systems*, vol. 6, no. 1&2, pp. 3–15, June 1990.

[162] R. C. Arkin, *Behavior-based Robotics*. Cambridge, MA, USA: MIT Press, 1998, iSBN:0262011654.

[163] E. Sahin, M. Çakmak, M. Dogar, E. Ugur, and G. Ucoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.

[164] R. R. Murphy, "Case studies of applying gibson's ecological approach to mobile robots," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 29, no. 1, pp. 105–111, 1999.

[165] A. P. Duchon, W. H. Warren, and L. P. Kaelbling, "Ecological robotics," *Adaptive Behavior*, vol. 6, no. 3, pp. 473–507, 1998.

[166] E. Krotkov, "Perception of material properties by robotic probing: Preliminary investigations," in *International Joint Conferences on Artificial Intelligence (IJCAI)*, 1995, pp. 88–94.

[167] G. Metta and P. Fitzpatrick, "Better vision through manipulation," *Adaptive Behavior*, vol. 11, no. 2, pp. 109–128, 2003.

[168] P. Fitzpatrick and G. Metta, "Grounding vision through experimental manipulation," *Phil. Trans. R. Soc. A: Mathematical, Physical and Engineering Sciences*, pp. 2165–2185, 2003.

[169] P. Fitzpatrick, G. Metta, L. Natale, A. Rao, and G. Sandini, "Learning about objects through action -initial steps towards artificial cognition," in *Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA)*, 2003, pp. 3140–3145.

[170] A. Stoytchev, "Toward learning the binding affordances of objects: A behavior-grounded approach," in *Proceedings of AAAI Symposium on Developmental Robotics*, March 2005, pp. 21–23.

[171] K. F. MacDorman, "Responding to affordances: Learning and projecting a sensorimotor mapping," in *Proceedings of 2000 IEEE International Conference on Robotics and Automation (ICRA)*, San Fransisco, California, USA, 2000, pp. 3253–3259.

[172] I. Cos-Aguilera, L. Canamero, and G. M. Hayes, "Motivation-driven learning of object affordances: First experiments using a simulated Khepera robot," in *Proceedings of the 9th International Conference in Cognitive Modelling (ICCM)*, Bamberg, Germany, 4 2003.

[173] ——, "Using a SOFM to learn object affordances," in *In Proceedings of the 5th Workshop of Physical Agents*, Girona, Catalonia, Spain, March 2004.

[174] G. Fritz, L. Paletta, M. Kumar, G. Dorffner, R. Breithaupt, and R. Erich, "Visual learning of affordance based cues," in *From animals to animats 9: Proceedings of the Ninth International Conference on Simulation of Adaptive Behaviour (SAB)*, ser. LNAI. Volume 4095., S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, and D. Parisi, Eds. Roma, Italy: Springer-Verlag, Berlin, 25–29 September 2006, pp. 52–64.

[175] S. May, M. Klodt, E. Rome, and R. Breithaupt, "Gpu-accelerated affordance cueing based on visual attention," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 2007, pp. 3385–3390.

[176] E. Ugur and E. Sahin, "Traversability: A case study for learning and perceiving affordances in robots," *Adaptive Behavior*, vol. 18, no. 3-4, 2010.

[177] E. Erdemir, C. B. Frankel, K. Kawamura, S. M. Gordon, S. Thornton, and B. Ulutas, "Towards a cognitive robot that uses internal rehearsal to learn affordance relations," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2016–2021, Sep. 2008.

[178] D. Kim, J. Sun, S. M. Oh, J. M. Rehg, and A. F. Bobick, "Traversability classification using unsupervised on-line visual learning for outdoor robot navigation," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 518–525.

[179] Y. Demiris and A. Dearden, "From motor babbling to hierarchical learning by imitation: a robot developmental pathway," in *Fifth International Workshop on Epigenetic Robotics*. Lund University, 2005, pp. 31–37.

[180] S. Hart, R. Grupen, and D. Jensen, "A relational representation for procedural task knowledge," in *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, 2005, pp. 1280–1285.

[181] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Affordances, development and imitation," in *IEEE International Conference on Development and Learning (ICDL)*, 2007.

[182] ——, "Modeling affordances using bayesian networks," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[183] M. Lopes, F. Melo, and L. Montesano, "Affordance-based imitation learning in robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1015–1021.

[184] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory–motor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[185] P. Osorio, A. Bernardino, R. Martinez-Cantin, and J. Santos-Victor, "Gaussian mixture models for affordance learning using bayesian networks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 4432–4437.

[186] J. Sun, J. L. Moore, A. Bobick, and J. M. Rehg, "Learning visual object categories for robot affordance prediction," *The International Journal of Robotics Research*, vol. 29, pp. 174–197, 2009.

[187] M. Kopicki, S. Zurek, R. Stolkin, T. Mörwald, and J. Wyatt, "Learning to predict how rigid objects behave under simple manipulation," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5722–5729.

[188] O. Kroemer, E. Ugur, E. Oztop, and J. Peters, "A kernel-based approach to direct action perception," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2605–2610.

[189] T. Hermans, F. Li, J. M. Rehg, and A. F. Bobick, "Learning contact locations for pushing and orienting unknown objects," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*. IEEE, 2013, pp. 435–442.

[190] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[191] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning probabilistic relational models," in *International Conference on Artificial Intelligence (IJCAI)*, 1999.

[192] J. Neville and D. Jensen, "Dependency networks for relational data," in *International Conference on Data Mining*, 2004.

[193] J. Neville, D. Jensen, L. Friedland, and M. Hay, "Learning relational probability trees," in *International Conference on Knowledge Discovery and Data Mining*, 2003.

[194] S. Griffith, J. Sinapov, M. Miller, and A. Stoytchev, "Toward interactive learning of object categories by a robot: A case study with container and non-container objects," in *Proc. of the 8th IEEE Intl. Conf. on Development and Learning (ICDL)*. Shanghai, China: IEEE, Jun. 2009, pp. 1–6.

[195] S. Griffith, J. Sinapov, V. Sukhoy, and A. Stoytchev, "A behavior-grounded approach to forming object categories: Separating containers from noncontainers," *IEEE Transactions on Autonomous Mental Development*, vol. 4, no. 1, pp. 54–69, 2012.

[196] D. Pelleg and A. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *International Conference on Machine Learning*, 2000, pp. 727–734.

[197] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *NIPS*, 2007, pp. 801–888.

[198] B. Ridge, D. Skočaj, and A. Leonardis, "Self-supervised cross-modal online learning of basic object affordances for developmental robotic systems," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 5047–5054.

[199] B. Ridge, E. Ugur, and A. Ude, "Comparison of action-grounded and non-action-grounded 3-d shape features for object affordance classification," in *Advanced Robotics (ICAR), 2015 International Conference on*. IEEE, 2015, pp. 635–641.

[200] J. Mugan and B. Kuipers, "Autonomous learning of high-level states and actions in continuous environments," *Autonomous Mental Development, IEEE Transactions on*, vol. 4, no. 1, pp. 70–86, 2012.

[201] H. M. Pasula, L. S. Zettlemoyer, and L. P. Kaelbling, "Learning symbolic models of stochastic domains," *Journal of Artificial Intelligence Research*, pp. 309–352, 2007.

[202] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2000.

[203] D. Prattichizzo and J. C. Trinkle, *Grasping*. Springer handbook of robotics, 2008.

[204] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis – a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[205] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *Robotics & Automation Magazine, IEEE*, vol. 11, no. 4, pp. 110–122, 2004.

[206] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, vol. 2. IEEE, 2003, pp. 1824–1829.

[207] R. Detry, E. Başeski, M. Popović, Y. Touati, N. Krüger, O. Kroemer, J. Peters, and J. Piater, "Learning object-specific grasp affordance densities," in *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on*. IEEE, 2009, pp. 1–7.

[208] R. Detry, D. Kraft, A. G. Buch, N. Krüger, and J. Piater, "Refining grasp affordance models by experience," in *Robotics and automation (icra), 2010 ieee international conference on*. IEEE, 2010, pp. 2287–2293.

[209] R. de Souza, S. El-Khoury, J. Santos-Victor, and A. Billard, "Recognizing the grasp intention from human demonstration," *Robotics and Autonomous Systems*, vol. 74, pp. 108–121, 2015.

[210] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.

[211] L. Montesano and M. Lopes, "Active learning of visual descriptors for grasping using non-parametric smoothed beta distributions," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 452–462, 2012.

[212] D. Kraft, N. Pugeault, E. BAŞESKI, M. POPOVIĆ, D. KRAGIĆ, S. Kalkan, F. Wörgötter, and N. Krüger, "Birth of the object: detection of objectness and extraction of object shape through object–action complexes," *International Journal of Humanoid Robotics*, vol. 5, no. 02, pp. 247–265, 2008.

[213] A. Morales, P. J. Sanz, A. P. Del Pobil, and A. H. Fagg, "Vision-based three-finger grasp synthesis constrained by hand geometry," *Robotics and Autonomous Systems*, vol. 54, no. 6, pp. 496–512, 2006.

[214] A. Wood, T. Horton, and R. S. Amant, "Effective tool use in a habile agent," in *IEEE Systems and Information Engineering Design Symposium*, 2005, pp. 75–81.

[215] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Proceedings of 2005 IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, April 2005, pp. 18–22.

[216] J. Sinapov and A. Stoytchev, "learning and generalization of behavior-grounded tool affordances," in *IEEE ICDL*, 2007.

[217] A. Stoytchev, *Learning the Affordances of Tools using a Behavior-Grounded Approach*, affordance-based robot control ed. Springer Lecture Notes in Artificial Intelligence (LNAI), 2008.

[218] V. Tikhanoff, U. Pattacini, L. Natale, and G. Metta, "Exploring affordances and tool use on the icub," in *IEEE Humanoids*, 2013.

[219] T. Mar, V. Tikhanoff, G. Metta, and L. Natale, "Self-supervised learning of grasp dependent tool affordances on the icub humanoid robot," in *IEEE ICRA*, 2015.

[220] R. Jain and T. Inamura, "Bayesian learning of tool affordances based on generalization of functional feature to estimate effects of unseen tools," *Artificial Life and Robotics*, vol. 18, no. 1-2, pp. 95–103, 2013.

[221] A. Gonçalves, G. Saponaro, L. Jamone, and A. Bernardino, "Learning visual affordances of objects and tools through autonomous robot exploration," in *IEEE ICARSC*, 2014.

[222] A. Gonçalves, J. Abrantes, G. Saponaro, L. Jamone, and A. Bernardino, "Learning intermediate object affordances: Towards the development of a tool concept," in *IEEE ICDL-Epirob*, 2014.

[223] A. Dehban, L. Jamone, A. Kampff, and J. Santos-Victor, "Denoising auto-encoders for learning of objects and tools affordances in continuous space," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[224] B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, and L. De Raedt, "Learning relational affordance models for robots in multi-object manipulation tasks," in *IEEE ICRA*, 2012.

[225] B. Moldovan, P. Moreno, and M. van Otterlo, "On the use of probabilistic relational affordance models for sequential manipulation tasks in robotics," in *IEEE ICRA*, 2013.

[226] J. Pisokas and U. Nehmzow, *Experiments in subsymbolic action planning with mobile robots, in: Adaptive Agents and Multi-Agent Systems II, Lecture Notes in Artificial Intelligence, 2005*. Springer, 2005, pp. 80–87.

[227] J. Modayil and B. Kuipers, "The Initial Development of Object Knowledge by a Learning Robot." *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 879–890, Nov. 2008.

[228] A. Maye and A. K. Engel, "Using sensorimotor contingencies for prediction and action planning," in *From Animals to Animats 12*. Springer, 2012, pp. 106–116.

[229] ——, "Extending sensorimotor contingency theory: prediction, planning, and action generation," *Adaptive Behavior*, vol. 21, no. 6, pp. 423–436, 2013.

[230] M. Hoffmann, N. M. Schmidt, R. Pfeifer, A. K. Engel, and A. Maye, "Using sensorimotor contingencies for terrain discrimination and adaptive walking behavior in the quadruped robot puppy," in *From Animals to Animats 12*. Springer, 2012, pp. 54–64.

[231] M. Hoffmann, K. Štěpánová, and M. Reinstein, "The effect of motor action and different sensory modalities on terrain classification in a quadruped robot running with multiple gaits," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1790–1798, 2014.

[232] V. Hogman, M. Bjorkman, and D. Kragic, "Interactive object classification using sensorimotor contingencies," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 2799–2805.

[233] H. Arie, T. Endo, T. Arakaki, S. Sugano, and J. Tani, "Creating novel goal-directed actions at criticality: A neuro-robotic experiment," *New Mathematics and Natural Computation*, vol. 5, no. 1, pp. 307–334, 2008.

[234] S. Murata, Y. Yamashita, H. Arie, T. Ogata, S. Sugano, and J. Tani, "Learning to perceive the world as probabilistic or deterministic via interaction with others: A neuro-robotics experiment," *Transactions on Neural Networks and Learning Systems (TNNLS)*, 2015.

[235] C. Geib, K. Mourão, R. Petrick, N. Pugeault, M. Steedman, N. Krüger, and F. Wörgötter, "Object action complexes as an interface for planning and robot control," in *IEEE Humanoids-06 Workshop: Towards Cognitive Humanoid Robots*, 2006.

[236] R. Petrick, D. Kraft, K. Mourão, N. Pugeault, N. Krüger, and M. Steedman, "Representation and integration: Combining robot control, high-level planning, and action learning," in *Proceedings of the 6th International Cognitive Robotics Workshop, 2008*, P. Doherty, G. Lakemeyer, and A. Pobil, Eds., July 2008, pp. 32–41.

[237] F. Wörgötter, A. Agostini, N. Krüger, N. Shylo, and B. Porr, "Cognitive agents – a procedural perspective relying on the predictability of object-action-complexes (oacs)," *Robotics and Autonomous Systems*, vol. 57, no. 4, pp. 420–432, 2009.

[238] N. Krüger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omrcen, A. Agostini, and R. Dillmann, "Object-action complexes: Grounded abstractions of sensorymotor processes," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740–757, 2011.

[239] P. Kaiser, M. Grotz, E. E. Aksoy, M. Do, N. Vahrenkamp, and T. Asfour, "Validation of whole-body loco-manipulation affordances for pushability and liftability," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE, 2015, pp. 920–927.

[240] E. Ugur and J. Piater, "Bottom-up learning of object categories, action effects and logical rules: From continuous manipulative exploration to symbolic planning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2627–2633.

[241] ——, "Refining discovered symbols with multi-step interaction experience," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE, 2015, pp. 1007–1012.

[242] A. Antunes, L. Jamone, G. Saponaro, A. Bernardino, and R. Ventura, "From human instructions to robot actions: Formulation of goals, affordances and probabilistic planning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[243] T. Lang and M. Toussaint, "Planning with noisy probabilistic relational rules," *Journal of Artificial Intelligence Research*, vol. 39, pp. 1–49, 2010.

[244] G. Salvi, L. Montesano, A. Bernardino, and J. Santos-Victor, "Language bootstrapping: Learning word meanings from perceptionaction association," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, pp. 660–671, 2012.

[245] O. Yuruten, E. Sahin, and S. Kalkan, "The learning of adjectives and nouns from affordance and appearance features," *Adaptive Behavior*, vol. 21, no. 6, pp. 437–451, 2013.

[246] S. Kalkan, N. Dag, O. Yuruten, A. M. Borghi, and E. Sahin, "Verb concepts from affordances," *Interaction Studies*, vol. 15, no. 1, pp. 1–37, 2014.

[247] H. Celikkanat, G. Orhan, and S. Kalkan, "A probabilistic concept web on a humanoid robot," *IEEE Transactions on Autonomous Mental Development*, 2014.

[248] H. Kjellstrom, J. Romero, and D. Kragic, "Visual object-action recognition: Inferring object affordances from human demonstration," *Computer Vision and Image Understanding*, vol. 115, no. 1, pp. 81–90, 2011.

[249] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *International Journal of Robotics Research (IJRR)*, vol. 32, no. 8, pp. 951–970, 2013.

[250] K. Varadarajan and M. Vincze, "Afrob: The affordance network ontology for robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 1343–1350.

[251] ——, "Afnet: The affordance network," in *Computer Vision ACCV 2012*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7724, pp. 512–523.

[252] Y. Zhu, A. Fathi, and L. Fei-Fei, "Reasoning about object affordances in a knowledge base representation," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 408–424.

[253] A. Myers, C. Teo, C. Fermuller, and Y. Aloimonos, "Affordance detection of tool parts from geometric features," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1374–1381.

[254] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *International Conference on Machine Learning*, 2001.

[255] C. Sutton, K. Rohanimanesh, and A. McCallum, "Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data," in *International Conference on Machine Learning*, 2004.

[256] B. Taskar, V. Chatalbashev, and D. Koller, "Learning associative markov networks," in *International Conference on Machine Learning*, 2004.

[257] M. Richardson and P. Domingos, "Markov logic networks," *Machine learning*, vol. 62, no. 1-2, pp. 107–136, 2006.

[258] P. Kontschieder, S. R. Bulo, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *International Conference on Computer Vision*, 2011, pp. 2190–2197.

[259] L. F. G. Rizzolatti and V. Gallese, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nat Rev Neurosci*, vol. 2(9), pp. 661–670, 2001.

[260] G. Hickok, "Eight problems for the mirror neuron theory of action understanding in monkeys and humans," *Journal of cognitive neuroscience*, vol. 21(7), pp. 1229–1243, 2009.

[261] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori *et al.*, "Integration of action and language knowledge: A roadmap for developmental robotics," *Autonomous Mental Development, IEEE Transactions on*, vol. 2, no. 3, pp. 167–195, 2010.

[262] A. Cangelosi and M. Schlesinger, *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press, 2015.

[263] S. Hart and R. Grupen, "Learning generalizable control programs," *Autonomous Mental Development, IEEE Transactions on*, vol. 3, no. 3, pp. 216–231, 2011.

[264] S. Fichtl, D. Kraft, N. Kruger, and F. Guerin, "Bootstrapping the learning of means-end action preconditions," *IEEE Transactions on Autonomous Mental Development (TAMD)*, 2015, submitted.

[265] E. Ugur, S. Szedmak, and J. Piater, "Bootstrapping paired-object affordance learning with learned single-affordance features," in *IEEE ICDL-Epirob*, 2014.

[266] E. Ugur and J. Piater, "Emergent structuring of interdependent affordance learning tasks," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*. IEEE, 2014, pp. 489–494.

[267] C. Wang, K. V. Hindriks, and R. Babuska, "Effective transfer learning of affordances for household robots," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*. IEEE, 2014, pp. 469–475.

[268] F. Worgotter, C. Geib, M. Tamosiunaite, E. Aksoy, J. Piater, H. Xiong, A. Ude, B. Nemec, D. Kraft, N. Kruger *et al.*, "Structural bootstrapping-a novel, generative mechanism for the faster and more efficient acquisition of action-knowledge," vol. 7, no. 2, pp. 140–155.

[269] S. Ivaldi, N. Lyubova, D. Gerardeaux-Viret, A. Droniou, S. Anzalone, M. Chetouani, D. Filliat, and O. Sigaud, "Perception and human interaction for developmental learning of objects and affordances," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2012, pp. 248–254.

[270] S. M. Nguyen and P.-Y. Oudeyer, "Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation," *Paladyn Journal of Behavioural Robotics*, vol. 3, no. 3, pp. 136–146, 2012.

[271] D. Caligiore, A. M. Borghi, D. Parisi, and G. Baldassarre, "Tropicals: A computational embodied neuroscience model of compatibility effects," *Psychological Review*, vol. 117, p. 11881228, 2010.

[272] D. Caligiore, A. M. Borghi, D. Parisi, R. Ellis, A. Cangelosi, and G. Baldassarre, "How affordances associated with a distractor object affect compatibility effects: A study with the computational model tropicals," *Psychological Research*, vol. 77, no. 1, pp. 7–19, 2012.

[273] S. Franklin, T. Madl, S. D'Mello, and J. Snaider, "Lida: A systems-level architecture for cognition, emotion, and learning," *Autonomous Mental Development, IEEE Transactions on*, vol. 6, no. 1, pp. 19–41, 2014.

[274] R. Sun, "The importance of cognitive architectures: an analysis based on clarion," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, no. 2, pp. 159–193, 2007.

[275] N. Krüger, J. Piater, F. Wörgötter, C. Geib, R. Petrick, M. Steedman, A. Ude, T. Asfour, D. Kraft, D. Omrcen, B. Hommel, A. Agostini, D. Kragic, J.-O. Eklundh, V. Krüger, C. Torras, and R. Dillmann, "A formal definition of object-action complexes and examples at different levels of the processing hierarchy," Tech. Rep. PACO-PLUS, 2009. [Online]. Available: www.paco-plus.org

[276] D. Vernon, C. Von Hofsten, and L. Fadiga, *A roadmap for cognitive development in humanoid robots*. Springer Science & Business Media, 2011, vol. 11.

[277] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, "The icub humanoid robot: An open-systems platform for research in cognitive development," *Neural Networks*, vol. 23, no. 89, pp. 1125–1134, 2010.

[278] K. Pastra and Y. Aloimonos, "The minimalist grammar of action," *Philosophical Transactions of the Royal Society B*, vol. 367, pp. 103–117, 2012.

[279] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 333–356, 1988.

[280] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.

Angelo Cangelosi received a PhD from Genoa University. He is Professor of Artificial Intelligence and Cognition and the Director of the Centre for Robotics and Neural Systems at Plymouth University (UK). Cangelosi's main research expertise is on language grounding and embodiment in humanoid robots, developmental robotics, human-robot interaction, and on the application of neuromorphic systems for robot learning. He has produced more than 250 scientific publications.

Luciano Fadiga received his M.D. from the University of Bologna, Italy, in 1989 and his Ph.D. in Neuroscience from the University of Parma, Italy, in 1994. Actually he is Professor of Physiology at the University of Ferrara, Italy and Senior Scientist at the Italian Institute of Technology. He is one of the discoverers of mirror neurons in macaque monkeys and has also provided empirical evidence by TMS and neuroimaging for the existence of a mirror-neuron system in humans.

Alexandre Bernardino received the Ph.D. degree in electrical and computer engineering from Instituto Superior Técnico (IST), Lisbon, Portugal, in 2004. He is currently an Associate Professor at IST and a researcher at the Institute for Systems and Robotics (ISR-Lisboa) in the Computer and Robot Vision Laboratory (VisLab). He participated in several national and international research projects in robotics, cognitive systems and computer vision. He published more than 180 articles in international journals and conferences (h-index 21).

Lorenzo Jamone received his MS in Computer Engineering from the University of Genoa in 2006 (with honors), and his PhD in Humanoid Technologies from the University of Genoa and the IIT in 2010. He was Associate Researcher at the Takanishi Laboratory in Waseda University from 2010 to 2012, and since January 2013 he is an Associate Researcher at VisLab (Instituto Superior Técnico, Lisbon, Portugal). His research interests include cognitive humanoid robots, motor learning and control, force and tactile sensing.

Justus Piater received his Ph.D. degree at the University of Massachusetts Amherst. After eight years as a professor at the University of Liege, Belgium, including one year as a visiting scientist at the Max Planck Institute for Biological Cybernetics in Tubingen, Germany, moved to Innsbruck in 2010. There he founded the IIS group at the Institute of Computer Science, which counts 6 postdoctoral researchers and 8 doctoral students. He has written about 150 publications, and has been a principal investigator in 1 EU-FP6 and 6 EU-FP7 projects.

Emre Ugur is an Assistant Professor in Bogazici University, Istanbul, Turkey. He received his Ph.D. degree in Computer Engineering from Middle East Technical University (METU, Turkey) in 2010. He was a research assistant in METU between 2003-2009; worked as a researcher in ATR, Japan between 2009-2013; visited Osaka University as a specially appointed Assistant Professor in 2015; and worked as a senior researcher in University of Innsbruck between 2013-2016. He is interested in developmental and cognitive robotics.

José Santos-Victor received the Ph.D. in Electrical and Computer Engineering from Instituto Superior Tcnico (IST), Lisbon, Portugal, in 1995. He is currently Full Professor at IST, Director of the Institute for Systems and Robotics (ISR-Lisboa) and head of the Computer and Robot Vision Laboratory (VisLab). He is the scientific responsible for IST in many European and national research projects in cognitive and bio-inspired computer vision and robotics. He published more than 300 articles in international journals and conferences (h-index 37).