Reply to comment

The scope and limits of action semantics

Reply to comments on ‘Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge’

Michiel van Elk a,*, Hein van Schie b, Harold Bekkering c

a University of Amsterdam, Department of Social Psychology, The Netherlands
b Behavioural Science Institute, Radboud University Nijmegen, The Netherlands
c Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

Received 25 March 2014; accepted 26 March 2014
Available online 2 April 2014
Communicated by L. Perlovsky

In our target paper [1], we have provided an integrative review of the current state of affairs regarding research on the use of conceptual knowledge for action. In short, we argue that humans have developed declarative and procedural knowledge about objects, i.e. action semantics that enable them to use objects in a purposeful and effective manner. We were impressed by both the quantity and the quality of the commentaries on our position paper. We would like to thank the commentators for their valuable contributions, pointing out supportive evidence for our framework, indicating additional and complementary strands of research in this field and raising important conceptual and theoretical challenges for our proposed model. Here we take the opportunity to briefly indicate how additional lines of research could be integrated in our framework and how potentially conflicting findings could be reconciled.

Joachim Hermsdörfer suggests that the overlap between the kinematics of pantomimed and object-directed actions points towards the existence of a common motor program, while at the same time task context also modulates the low-level features of action execution [2]. The suggestion that the context may exert both an indirect effect (i.e. through a top-down selection of motor programs) and a direct effect on movement execution (e.g. by providing participants with stronger task constraints and sensory feedback through a direct interaction with an object) provides an excellent illustration of the role of context, as proposed in our model of action semantics. A similar role of context on the activation of motor programs is proposed by Anna Borghi [3], who showed that the context (i.e. provided by the task, the inclusion of other objects, semantic information or the presence of other persons) can have a strong influence on the activation of object affordances.

DOI of original article: http://dx.doi.org/10.1016/j.plrev.2013.11.005.
* Corresponding author. Tel.: +31205256891.
E-mail address: m.vanelk@uva.nl (M. van Elk).

http://dx.doi.org/10.1016/j.plrev.2014.03.009
1571-0645/© 2014 Elsevier B.V. All rights reserved.
Borghi makes a valuable and important suggestion to implement our framework in a computational model and to relate it to current developments in robotics [4]. A similar suggestion was made by Michael Masson, who notes that the framework of action semantics does not lead to testable predictions but rather to general suggestions for future experiments [5]. We are aware of this issue and we would like to point out we ourselves and others have proposed computational models to implement the notion of action hierarchies and to model the execution of goal-directed actions [6–8]. At the same time, a big challenge in the field of robotics is to bridge the gap between embodied systems that take the sensorimotor experiences in the real world as the starting point for learning [9,10] and symbolic systems that represent learned knowledge in a symbolic connectionist network [11,12]. The notion of ‘action semantics’ refers precisely to the interplay between different levels of knowledge representation that enables us to act in the real world. The challenge of implementing such a conceptual model in a computational model is non-trivial and we hope that our framework may guide future researchers working in the field of robotics and artificial intelligence to implement our framework in a model that allows generating testable predictions.

Guy Vingerhoets rightly points out that there is a hemispheric asymmetry between the use of action semantics, which is lateralized to the left hemisphere and is effector-independent, and the online control of movements, which is associated with a bilateral activation of the fronto-parietal action network and is effector-dependent [13]. The asymmetrical neural organization of different components of object-directed actions fits well with the proposed distinction between action goals, which are effector-independent (i.e. the spatial goal location associated with using an object) and action means that are effector-dependent (i.e. the specific way in which an object is grasped; cf. [1]). However, at the same time we would like to note that the neural organization of goals and grips remains a matter of ongoing debate and that the representation of both goal- and grip-related information has been localized to the posterior parietal cortex and the left inferior parietal lobe [14–17]. The observation that the lateralization of action semantics mirrors the left hemisphere specialization for language processing and production has actually resulted in several theoretical proposals regarding the evolutionary origin of these functions [e.g. language could have evolved out of gestural communication; cf. [18]].

A comparable evolutionary argument is put forward by Scott Glover [19], who makes the interesting suggestion that the expansion of the left IPL in humans compared to primates may have supported the co-evolution of semantic and tool use systems and that the need for a highly specialized effector for object manipulation may have driven the development of right-handedness [20]. This observation places the framework of action semantics in an evolutionary perspective and fits well with the finding that the dexterity of human tool use marks a discontinuity between humans and non-human primates [for a recent review of tool use in humans and primates, see [21]]. We concur with the criticism that the distinction between dorsal and ventral streams in our model represents an oversimplification and we are strongly in favor of a more nuanced view on this topic, stressing the interconnectedness of many brain areas that are typically considered part of the dorsal or the ventral stream [see for instance [22]]. We note that the dorsal–ventral distinction is not central to our model (see Fig. 1) and that these labels were merely used to indicate that higher order visual areas that are typically considered ‘core regions’ of the ventral stream, play an important role in action semantics as well [23,24].

In our review we extensively discuss developmental studies focusing on the acquisition and learning of action semantics. Arthur Glenberg and Tamir Soliman propose a complementary hypothesis, according to which the learning of tool use involves the incorporation of the object in the body representation [25]. This perspective is supported by single-cell studies in monkeys [26,27], studies with neuropsychological patients and behavioral studies in humans, using the cross-modal congruency task as a measure of the multisensory integration of objects in the body representation [28]. In fact, in several studies we have investigated how the use of manipulable objects facilitates body-object integration [29–32], resulting in what Soliman and Glenberg call a joint ‘Hand-Tool Body Schema’ (HTBS; [33]). For instance, we found that the presentation of manipulable objects (e.g. a cup) compared to non-manipulable objects (e.g. a car) resulted in the extension of peripersonal space around the body, likely related to the retrieval of the sensorimotor programs associated with using these objects [32]. Furthermore, we found that the interaction with a novel haptic robotic interface, enabling people to perform actions at remote locations (e.g. surgeons, space engineers), also resulted in an extension of peripersonal space to the distal location at which the robot was actually performing the movements generated by the participant [29,30]. This finding relates to the comment made by Soliman and Glenberg [33] who discuss the question whether internal forward models involved in tool use code for the proximal or distal effects of tool use. The results of the study with the robotic interface indicate that actions are coded primarily in terms of the distal outcomes of the tool use action – even when there is no direct physical connection between the user and
Fig. 1. Extended conceptual framework of action semantics (upper panel) and neural dynamics of action semantics (lower panel). Action semantics consists of multimodal object representations that are associated with modality-specific sub-systems involving functional knowledge, manipulation knowledge and representations of the proprioceptive, somatosensory, interoceptive visual, auditory and olfactory consequences of object use. Action semantics are hierarchically organized and the selection of action outcomes and the relevant sub-systems is driven by action intentions, which are in turn determined by the context in which the action takes place. Different modality-specific brain areas are associated with a multimodal association area, involved in the retrieval of action semantics. Frontal areas modulate activation in brain areas involved in action semantics and determine the relative strength of activation of the different modality-specific systems. (FPC = Frontopolar Cortex; DLPFC = Dorsolateral Prefrontal Cortex; SMA = Supplementary Motor Area; PMC = Premotor Cortex; left SMG = left Supramarginal Gyrus; PPC = Posterior Parietal Cortex; left IPL = left Inferior Parietal Lobe; IT = InferoTemporal Cortex; MTG = Middle Temporal Gyrus; left pMTG = left posterior Middle Temporal Gyrus.)

the tool [29,30]. These findings are in line with the proposed model of action semantics, according to which tool use actions are organized and represented in a hierarchical fashion, around the final outcome or goal of the action [1]. Further support for this notion is discussed by Michael Masson [5], who describes a study indicating that neurons in the premotor cortex coded actions in terms of the outcome rather than the specific movements involved.

We agree with Jessica Witt that the costs and benefits associated with actions play an important role in the way in which we perceive and interact with the surrounding world [34]. Indeed, many of the studies in the target article describe actions that come at relatively little costs, whereas in daily life actions are associated with specific costs or benefits. Also from an evolutionary point of view, actions need to be evaluated primarily in terms of their adaptive value and their potential for successful survival. We speculate that the costs and benefits of actions may be considered an integral part of the way in which actions are represented (i.e. in terms of their outcomes or effects). It has been found for instance that the learned value of objects exerts a strong influence on the kinematics of object-directed grasping [35]. Similarly, in a previous study we found that associating novel objects with specific sensory consequences (e.g. a sound or a smell, which can be considered an implicit reward) modulates the acquisition of representations about the end-location associated with the object [36]. The suggestion that the interoceptive consequences (i.e. related to one’s bodily states) of an action should be incorporated in the model fit nicely with recent studies, underlining the
central role of interoceptive signals for supporting a sense of the bodily self [37] and for the integration of external objects in the body representation [38]. In the extended model of action semantics, we have included the interoceptive consequence of one’s action as well (see Fig. 1) and we propose the insula as the core region for processing interoceptive signals [39].

In their comment, François Osiurak and Mathieu Lesourd [40] propose an alternative account of the available neuropsychological data. They argue that many of the deficits observed in apraxia patients can be accounted for in terms of impairments with mechanical problem solving rather than a loss of stored motor programs for object use. A similar proposal has been made by Georg Goldenberg [e.g. [41]], who argues that tool use deficits in apraxia are primarily related to impairments in perceiving the spatial relation between one’s body and the objects involved. An important argument for this account can be found in studies with apraxic patients, indicating that impairments in the use of novel tools are often strongly correlated with an impaired use of well-known objects as well [42,43].

This view provides an important challenge to the conceptual framework of action semantics, as the notion of an action hierarchy allowing the flexible use of objects could be replaced with a more parsimonious account in terms of technical problem solving. However, we would like to point out that a correlation between impairments in the use of novel and well-known tools does not preclude the possibility that this is related to a general loss of stored hand-postures for object use. In fact, it is likely that when using novel objects we typically rely on comparable motor programs as involved in the use of well-known objects. For instance, when presented with a new piece of electronic equipment, we may initially try to manipulate it in a similar fashion as our mobile phone or I-Pad. Thus, people may apply learned motor primitives to the interaction with novel objects as well [e.g. [44]], and accordingly impairments in stored hand postures should affect the use of novel and well-known objects equally well. Furthermore, we would like to point out that we rely on more strands of evidence than neuropsychology alone to support the idea of stored manipulation knowledge guiding the interaction with objects. We discuss developmental, behavioral and neuroimaging studies as well, and we argue that these different findings provide converging evidence for the notion of an embodied representation of ‘action semantics’ and the idea of an action hierarchy finds further support in recently proposed predictive coding models of the brain [45,46]. Thus, although mechanical problem solving may contribute to object use in specific cases (e.g. trying to figure out how to use a coin as a screwdriver), we suggest that this is not the default way of interacting with objects and that our tool use ability relies on the use of stored motor programs instead.

Alex Martin, Kyle Simmons, Michael Beauchamp and Stephen Gotts provide an important challenge for our claim that multimodal object information converges in a semantic hub that is localized in the anterior temporal lobes (ATL; [47]). Their argument is based on a critical examination of the available evidence from studies with neuropsychological patients with semantic dementia, neuroimaging studies and studies on functional and anatomical connectivity in the human brain [for a similar discussion, see [48,49]]. We believe that in this discussion it may be important to distinguish between the different claims entailed by our framework of action semantics, namely (1) the notion that action semantics involves multimodal object representations (as represented in Fig. 1 of the review paper), (2) the suggestion that these representations are instantiated in a semantic hub and (3) the claim that this hub is localized in the ATL (as represented in Fig. 2 of the review paper; [1]). The first claim dates back to Piaget, who argued that throughout development children increasingly abstract away from concrete sensorimotor experiences to acquire concepts about the world [50]. An important theoretical argument for postulating multimodal semantic representations, is that in order to recognize and categorize object categories, one needs to abstract away from the idiosyncratic features that make up individual exemplars of objects [12]. Furthermore, multimodal representations can function as a proxy to provide direct access to a concept from a specific category, without having first to re-enact the modality-specific representations associated to that concept [51]. In language research, similar considerations have resulted in postulating a ‘hybrid model’ that involves both modality-specific representations and multimodal or ‘amodal’ symbolic representations of concepts [52]. Regarding the second claim, as Martin et al. note [47], recent studies using network analysis techniques have supported the existence of ‘anatomical hubs’ in the human brain that are involved in coupling information from different networks [53–55]. In addition, studies with neuropsychological patients indicate that damage of neural tissue may result in impairments in selective aspects of object use (e.g. as observed in apraxia) or in a more general loss of semantic knowledge (e.g. as observed in semantic dementia), also suggesting the existence of a ‘general semantic hub’ [56]. The notion of a semantic hub is further supported by studies in the field of artificial intelligence and robotics, also indicating the importance of generalized semantic representations for a successful interaction with the world [e.g. for communication or language understanding; [10,57]]. Thus, we believe there is strong and converging evidence.
to postulate the existence of a level of ‘multimodal object representations’ as in our model of action semantics [1]. Accordingly, what is at stake here is the third claim that the semantic hub is instantiated in the ATL. Taking into consideration the arguments by Martin et al. and their expertise in this field [47], we concur with the suggestion that a more likely candidate region acting as a hub for representing information about tools can be found in the left posterior Middle Temporal Gyrus (pMTG) and we have included this region in our extended model of action semantics (see Fig. 1).

Bernhard Hommel raises a concern about the scope of our theory of action semantics [58], indicating that the proposed framework selectively accounts for object-directed actions, whereas an alternative theoretical framework (i.e. the Theory of Event Coding; TEC; cf. [59]) accounts for action planning and control in general. When framing a scientific theory there is always a need for scope (i.e. the explanatory potential) and specificity (i.e. the degree to which a theory gives an explanation for a specific instance, cf. [60]). We defined action semantics as the ‘declarative and procedural knowledge that enables the use of objects in a purposeful and effective manner’ and as such the framework meets up to its expectations. A potential problem with a more general theory of action control is that it fails to account for specific subtleties that can be observed when taking a closer look at a specific phenomenon. For instance, it is unclear how the TEC would account for the selective deficits observed in neuropsychological patients, reflected in a dissociation between functional and manipulation knowledge [61]. Also, it is difficult to specify the brain networks underlying a general framework like TEC. Regarding Hommel’s criticism on the notion of an action hierarchy, we would like to point out that a hierarchical organization is not only required to control one’s actions, but to learn and monitor one’s actions as well. In fact, according to the predictive coding model of the brain [45, 46], the notion of hierarchically organized layers with increasing levels of complexity is an essential prerequisite for motor learning and adjusting one’s actions online. Prediction error signals between lower and higher levels in the hierarchy support learning, either by changing the internal predictive model of the world, or by changing the sensory input through one’s actions in the world. Finally, we would like to note that the distinction between sensory and proprioceptive consequences of one’s actions corresponded to the respective origin of the signals (i.e. external to the body or internally generated). To clarify our definitions, in the extended model of action semantics, we now use the following terminology: proprioceptive consequences (i.e. related to the relative position of one’s body parts), somatosensory consequences (i.e. related to touch), interoceptive consequences (i.e. related to the internal state of one’s body), visual, auditory and olfactory consequences (see Fig. 1).

In sum, we believe that the different perspectives in the commentaries have further strengthened our framework by providing corroborating evidence in support of our model and by providing excellent suggestions for further improvement of conceptual and neurological details that we incorporated. We thank the commentators for contributing their insights in this shared domain of scientific study. It is our hope that this target paper may be used as a comprehensive review and introduction to the field of research on the use of tools and objects and may guide future research to further elucidate the functional and neural basis of action semantics.

References


[22] Schenk T, McIntosh RD. Do we have independent visual streams for perception and action? J Cogn Neurosci 2010;1:52.


