On the dynamic interplay between perception and action - a connectionist perspective

Haazebroek, P.

Author: Haazebroek, Pascal
Title: On the dynamic interplay between perception and action : a connectionist perspective
Issue Date: 2013-12-11

Chapter 1
Introduction

How do we interact with our environment? We effortlessly turn door handles, reach for a cup of coffee, and use various kinds of tools and electronic equipment. But how do we coordinate our actions in response to these environmental demands? Intuitively, we first perceive an object, then we think for a very brief moment, and, finally, we perform actions on it. So, somehow in 'the thinking' our perception and action systems must 'connect'. The nature of this connection has been a central topic within the field of Cognitive Psychology (Ward, 2002). Indeed, actions that are not guided by perception would not only be inefficient but might also be rather dangerous. Moreover, coordinating perception and action is potentially very complex as natural environments offer an overwhelming number of perceivable objects and natural bodies allow for a virtually unlimited number of different responses. As the human cognitive system usually seems to cope quite well with this complexity, understanding its perception and action connection could be beneficial for developing artificial embodied cognitive systems (i.e., robots) that need to cope with similar challenges.

In this thesis I argue that perception and action planning do not represent separable stages of a unidirectional processing sequence, but rather emerging properties of highly interactive mental processes. In other words, information processing is the result of a (context modulated) dynamic interplay between perception and action.

Traditional views of human information processing

Traditionally, very much in line with the intuitive reasoning described above, responding to stimuli in our environment has theoretically been conceived as a *sequence of separable stages of processing* (e.g., Donders, 1868; Neisser, 1967; Sternberg, 1969; see Figure 1). The separation of information processing into a sequence of steps has a strong history in various theories of human information processing. Moreover, in many models and cognitive systems different steps are often realized by different modules.

For example, in their seminal work, Card, Moran and Newell (1983) describe the Model Human Processor being composed of three main *modules*: perception, cognition and action modules. Information processing is defined as a cyclic, sequential process from stimulus perception to cognitive problem solving to response execution. The perceptual system is considered to contain sensors and is responsible for coding the sensory input into symbolic representations. The cognitive system combines this symbolic input with long term memory and determines how to respond. Finally, the motor system is assumed to carry out the specified response.

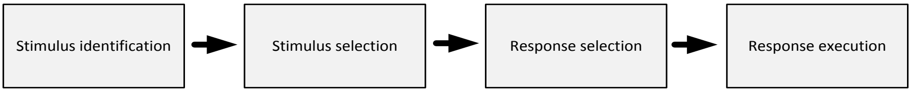

Figure 1. The perceive-think-act sequence is the basis of various theories of human information processing

In similar vein, the Seven stages of Action model (Norman, 1988) — a conceptual model of human task performance popular in the field of Human Computer Interaction — decomposes the interaction between people and their environment into the following seven *stages*: people (1) perceive the state of the world, (2) interpret their perception, (3) form evaluations based on these interpretations, (4) match these evaluations against their goals, (5) form an intention to act, (6) translate this intention into a sequence of actions and (7) execute this action sequence. Executing an action sequence subsequently results in a change in the world state which can again be perceived in the first stage.

More recently, cognitive architectures have been developed (e.g., ACT-R, Anderson, 1993; SOAR, Newell, 1990; EPIC, Kieras & Meyer, 1997) to address the challenge of computationally characterizing human information processing. Crucially, these architectures also separate processing in stages and mostly focus on the middle, cognitive steps of the perceive-think-act processing sequence. It is assumed that the first steps, perceiving and interpreting the world state, are performed relatively easily. The main focus is on comparing the world state with a goal state and deciding upon which action to take next in order to achieve the goal state. It is further assumed that once an action is chosen, its execution is easy, leading to a predictable new world state. The core mechanism used by most cognitive architectures is a production rule system (Byrne, 2003). A production rule defines the translation of a pre-condition into an action that is known to produce a desired post-condition. This can be interpreted as "IF (x) THEN (y)" rules. By specifying a set of production rules, a cognitive architecture can be given some prior knowledge resulting in response tendencies to choose those actions that eventually realize certain goals. When putting a cognitive architecture, endowed with a set of production rules, in interaction with an environment, however, conflicts between rules or unexpected conditions may present themselves. Moreover, by assuming a set of production rules, a cognitive architecture also assumes a set of action alternatives. However, when someone is interacting with a physical or virtual environment, it is often unclear which actions can be performed. Also, in certain contexts, people may not readily detect all action opportunities and action alternatives may differ in their availability, leading to variance in behavior (Kirlik, 2007). This is hard to capture in a cognitive architecture that assumes a predefined set of (re)actions.

Artificial Intelligence and Robotics

Responding to environmental demands in the environment has also been a major challenge in the fields of Artificial Intelligence and Robotics. In the 1960s – 1970s these fields started out with top down approaches focusing on robots that could reason about the world, create internal maps and figure out with hard computation how to navigate through the world. A well-known example was Shakey, a robot built in the late 1960s (Nilsson, 1984). Shakey was essentially a box on wheels with a camera. It was accompanied with an off-board computer that was programmed to make plans of 'what to do next'. The interaction with the environment started with a perception stage in which camera input was analyzed and a world model was computed in the off-board computer. Then, during the 'think' stage, the computer would go through all alternatives of what to do next, an algorithm taking minutes to compute. Finally, during the 'action' stage, essentially with eyes shut, Shakey would move a couple of feet, hoping that the world would remain stable. Then, in a new cycle, Shakey opened up its eyes again, looked at the environment, built a new world model and continued its journey. As demonstrated by rather hilarious scenes where culprits would come in and alter the environment precisely when Shakey was in its 'blind' action stage resulting in inaccurate internal models and inappropriate actions, this perceive-think-act architecture seemed to pose a problem for real world robotics: robots constructed like Shakey are limited by the need for all information from the sensors to pass through the modeling and planning modules before having any effect on the robot's actions (Brooks, 1991). As a result, Shakey could only cope with highly impoverished, static environments. Natural, dynamic environments would require too much time to construct a plan in response to ongoing, unexpected events.

In the decades that followed, some AI researchers took stronger notice of nature and observed that rather simple organisms such as bugs and insects are quite able to cope with environments that are too challenging for Shakey. Brooks (1986) proposed an activity-based decomposition of information processing. He reasoned that perception, cognition and action should be considered intertwined and suggested that a system might rather be decomposed in different behavior-producing subsystems and that *each* of these subsystems in itself forms a *complete perception, cognition, action pathway*. As these pathways may inhibit or suppress each other, such a system is able to exhibit a wide variety of complex behaviors. This approach resulted in a decade of developing insect-like robots that demonstrated much better performance in dealing with real environments than the earlier robots based on the top down approach, like Shakey. However, linking their behavior and internal representations to higher level cognitive activities such as planning, reasoning about and communicating with other robots or humans proved to be rather hard (Shanahan, 1998).

The issue of modularity is still a central topic in modern day robotics. Robots are complex systems and functional decomposition into hardware and/or software modules makes sense from an engineering point of view. In the last decade we have witnessed the dawn of highly advanced robot vision systems that recognize complex objects instantaneously (e.g., Detry & Piater, 2011) and reconstruct entire 3D scenes in internal world models (e.g., Baseki et al., 2010) . Moreover, video clips of robots showing immensely impressive behavioral repertoires (e.g., drumming, dancing, walking stairs) appear in the media weekly. How to architect and interconnect perceptual, cognitive and action systems, however, remains a matter of debate and an issue to be explicitly addressed by roboticists (e.g., the three-level architecture described in Kraft et al., 2008).

Information processing in the brain

Where traditional views on human information processing focus on the 'software' processing steps irrespective of the 'hardware' (i.e., the brain) that is assumed to perform these steps, connectionist theories stress that the structural and functional properties of the brain may have strong influences on human information processing. Indeed, the human brain does not contain a single complex central processor that does all the computations; it rather consists of billions of *simple computing units* (neurons) that are *interconnected* by trillions of connections and primarily engage in local interactions (i.e., with their directly connected 'neighbors').

Given the complexity of the brain early work on network models of cognitive performance was not aimed at modeling brain activity in complete detail. Researchers rather set out to model cognitive phenomena in systems that exhibited some of the same basic properties as networks of neurons in the brains. McCulloch and Pitts (1943) laid the foundation with networks composed of binary units and demonstrated (Pitts & McCulloch, 1947) that these networks could be used to perform pattern recognition tasks. Later approaches (e.g., Rosenblatt 1961) explored similar networks of units with connections of varying weights. In addition, following Hebb's (1949) suggestion that when two neurons in the brain were jointly active, the strength of their connection might increase, procedures came to be that allowed these networks to learn and demonstrate associative memory abilities (e.g., Taylor, 1956). The success of these early network models was, however, rather short-lived as there were strong limitations (e.g., demonstrated by Minskey & Papert, 1969) to what this type of networks could compute and serious learning algorithms were lacking. These limitations turned the focus of AI research towards symbolic models of information processing.

In the mid-1980s, however, important limitations of rule-based symbolic systems were identified (e.g., inflexibility, difficulty in learning from experience, inadequate generalization) and network-inspired approaches came back in vogue. Rumelhart, Hinton, and McClelland (1986) published their very influential *Parallel Distributed Processing* (PDP) work that essentially defined the connectionism. In the connectionist approach there is a network of elementary units, each of which has a certain degree of activation. The network is considered to be a *dynamical* system which, once provided with initial input, spreads activation among its units for a set period of *time* or until a stable state is achieved. Such a connectionist system is considered to 'perform' a cognitive task by interpreting the inputs as a problem and the resulting stable configuration of the system as the solution to that problem. Compared to the symbolic approach, that involves transformation of symbols according to specific rules, the connectionist approach focuses on causal processes by which the units spread activation to each other. Hence, information processing in connectionist networks is *distributed.*

As connectionism became increasingly popular in the late 1980s, some researchers (e.g., Fodor, 1983; Pinker, 1997) argued that connectionism actually constituted a reversion toward behaviorism (e.g., Watson, 1913) by focusing on mere input-output associations rather than addressing mental processes in terms of explicit logical algorithms. In their view, mental activity is computational; that is, performing operations on symbols. Indeed one could argue that connectionist-like hardware (i.e., the brain) may actually only implement the symbolic-like software algorithms. Hence, the mind could still very well be decomposed in separate (e.g., perception, cognition and action) subsystems. Indeed, to what extent the mind can be considered a *modular* system is still a matter of lively debate among cognitive scientists (Prinz, 2006).

Direct interaction between perception and action

Now, interestingly, empirical findings in psychology have demonstrated that parts of human information processing do not seem to involve conscious cognitive decision making. Features of perceived objects (such as location, orientation, and size) can influence actions *directly* and beyond (tight) cognitive control, as illustrated by stimulus–response compatibility phenomena, such as the Simon effect (Simon & Rudell, 1967). In the typical Simon task, stimuli vary on a spatial dimension (e.g., randomly appearing on the left or right) and on a non-spatial dimension (e.g., having different colors). Participants have to respond to the non-spatial stimulus feature by performing a spatially defined response (e.g., pressing a left or right key). Although the location of the stimulus is irrelevant for the response choice, it nevertheless influences response time and accuracy: participants respond faster (and more accurately) when the stimulus location is congruent with the response location than when the stimulus location is incongruent with the response location. This finding suggests that there is a direct interaction between stimulus perception and response planning. The Simon effect is a very robust finding, has been replicated numerous times and has been used frequently as a methodological tool to investigate perception, action, and cognitive control (for general overviews, see Hommel, 2011; Proctor, 2010).

To account for both controlled and automatic processing, various dual route process accounts have been proposed (e.g., Kornblum, Hasbroucq, & Osman, 1990; Zorzi & Umilta, 1995). These accounts propose that there is a second, direct route from perception to action that can bypass cognition, as explicitly modeled in various computational models of the Simon effect (e.g., Zorzi & Umilta, 1995; see Chapter 4 for a more elaborate discussion). Essentially, dual route accounts consider the observed direct stimulus-response interaction as an exception requiring an additional route next to the 'normal' one that does involve cognition. Moreover, they typically do not address the reason *why* some stimulus features directly influence action and others do not.

Representing perception and action using common codes

An alternative view that gives much more weight to this direct interaction between perception and action is the Theory of Event Coding (TEC, Hommel, Müsseler, Aschersleben & Prinz, 2001; illustrated in Figure 2). TEC is a general theoretical framework that addresses how perceived events (i.e., stimuli) and produced events (i.e., actions) are cognitively represented and how their representations interact to generate both perceptions and action plans. TEC holds that stimuli and actions are represented in the same way and by using the same 'feature codes'. These codes refer to the *distal* features of objects and events in the environment, such as shape, size, distance, and location, rather than to proximal features of the sensations elicited by stimuli (e.g., retinal location or auditory intensity; see Heider, 1959; Hommel, 2009). For example, a haptic sensation on the left hand and a visual stimulus on the left both activate the same distal code representing 'left'.

Crucially, these feature codes can represent the properties of a stimulus in the environment just as well as the properties of a response — which, after all, is a perceivable

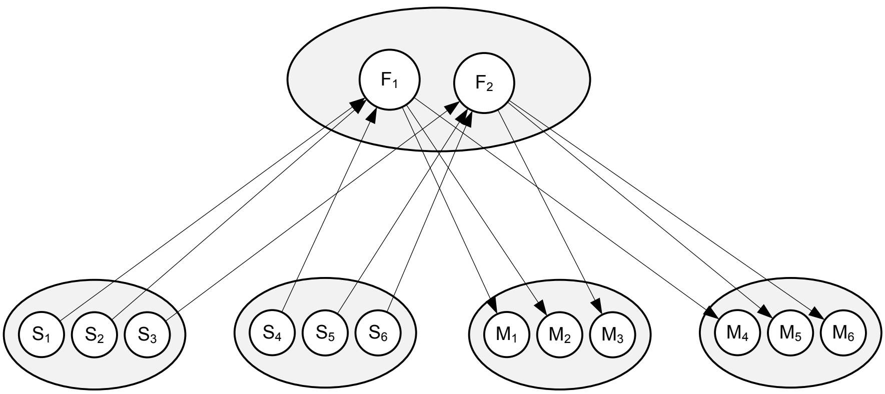

Figure 2. Sensory, feature and motor codes in TEC (adapted from Hommel et al., 2001). Multiple sensory codes can relate to the same feature codes (and vice versa). The same holds for motor codes and feature codes.

stimulus event itself. This theoretical assumption is derived from ideomotor theory (James, 1890; see Stock & Stock, 2004, for a historical overview), which presumes that actions are cognitively represented in terms of their perceivable effects. According to *the ideomotor principle*, when one executes a particular action, the motor pattern is automatically associated to the perceptual input representing the action's effects (action–effect learning; Elsner & Hommel, 2001). Based on these action-effect associations, people can subsequently plan and control (Hommel, 2009) a motor action by anticipating its perceptual effects, that is, (re-)activate a motor pattern by intentionally (re-)activating the associated feature codes. Thus, stimuli and actions are represented in a *common representational medium* (Prinz, 1990). Consequently, stimulus perception and action planning are considered to be similar processes: both involve activating[1] feature codes that represent external events.

Neuroscientific evidence for common codes at a distal feature level can be found in the response characteristics of mirror neurons in the premotor cortex (cf., Keysers & Perrett, 2004). In the macaque monkey, these neurons are active both when the monkey performs a particular action and when it perceives the same action carried out by another monkey or human, such as picking up food. Crucially, this overlap occurs at a distal representational level, that is, at the level where planned and perceived actions can be described as having the same goal or end state such as picking up an object (Rizzolati & Craighero, 2004). Also, various behavioral studies show that, in humans, action planning can actually influence object perception (e.g., Fagioli, Hommel & Schubotz, 2007; Stoet & Hommel, 2002; Wykowska, Schubö & Hommel, 2009), suggesting that perceptual processes and action processes overlap in time (see also Hommel, 1997) and influence each other.

[1] TEC also addresses how more complex cognitive codes ('event files') are created, an aspect that refers to the integra-tion of feature codes rather than their mere activation. This structure-building aspect will not be dealt with in this thesis but be left for future work.

Finally, TEC stresses the role of *task context* in stimulus and response coding. In particular, the responsiveness of feature codes to activation sources is considered to be modulated according to the task or goals at hand (the *intentional weighting principle*, Memelink & Hommel, 2013). For example, if the task is to grasp an object, feature codes representing features relevant for grasping (such as the object's shape, size, location and orientation) are assumed to be *enhanced*, while feature codes representing irrelevant features (such as the object's color or sound) appear to be attenuated (Hommel, 2010; Wykowska et al., 2009).

HiTEC connectionist model

In this thesis I aim to shed more light on the biological and computational plausibility of common representations underlying perception and action planning. To this end I have developed HiTEC, a connectionist model based on TEC. Our aim was to formulate a clear alternative to sequential models of perception and action and to develop a *minimal framework* for considering how perceptual and action processes may interact in the control of behavior. HiTEC extends and further specifies TEC's principles to account for a series of key experimental findings in a unitary theoretical framework and at a level of specificity that allows for computer simulation.

Outline of the thesis

The thesis is organized as follows. Chapter 2 presents HiTEC, the connectionist model developed to study the feasibility of common representations and interactive processing; in Chapters 3 to 5, various simulations of empirical phenomena are described. Here, the focus is on research questions that particularly challenge existing models of stimulus-response translation that assume separate modules or processing stages. Finally, general conclusions are described in Chapter 6. In this endeavor the following research questions are addressed in this thesis.

How do neuron-like representations realize stimulus-response translation?

This research question is addressed in Chapter 2. In this chapter, the HiTEC connectionist model is presented. In HiTEC, neuron-like representations are *distributed* over *multiple levels* and processing involves both feedforward and *feedback* interaction between lower and higher level representations. In addition, one of the HiTEC levels contains *common* representations; these representations are used both for stimulus perception and response planning. As a result, stimulus-response processing is fully interactive rather than in stages. The HiTEC model is used in all simulations discussed in this thesis.

How do situation-specific meanings of motor actions emerge?

In order to control its actions in response to demands in the environment the cognitive system needs to know what actions are possible and what these actions 'mean'. Various empirical findings suggest that for a cognitive system this 'meaning' is not a fixed fact; it rather depends on the (perceptual) effects within the task context. Consequently, in order to select and execute an appropriate response to a stimulus a plausible cognitive model must first learn (i.e., from experience) what the effects of its motor actions are and how to interpret these effects in the task context. How these situation-specific meanings of actions may emerge and how these meanings are used in action control is addressed in Chapter 3. Simulations in this chapter demonstrate that HiTEC allows for associating action effects with motor actions. Moreover, the strengths of these associations depend on the context allowing for the emergence of situation-specific meanings.

How and why do parts of stimulus–response translation occur automatically?
Some parts of the translation from stimulus to response are considered to occur automatically as demonstrated by stimulus–response compatibility (SRC) effects such as the Simon effect (Hommel, 2011; Simon & Rudell, 1967). How and why these effects may occur is addressed in Chapter 4. Simulations in this chapter demonstrate that HiTEC provides a parsimonious rationale for these effects, most notably in terms of the common representation level and the fact that task-relevance is considered to apply to both stimuli and responses.

How does the task context modulate stimulus-response translation?
How the task context may modulate stimulus-response translation is more explicitly addressed in Chapter 5, which includes both the simulation of an existing empirical study and a novel behavioral study and its simulation. The first simulation in this chapter demonstrates how the task context may modulate action control by means of (spatial) attention within the environment; the empirical study and the second simulation show how intentional weighting may also operate on a more abstract (distal) level.

Finally, in Chapter 6, these research questions and their interrelations are further discussed as not only perception and action are strongly interrelated, so are the different research questions addressed in this thesis.

Publications

Note that these chapters contain major parts of various articles published in the course of this research. Rather than presenting a collection of published and submitted articles divided over chapters, I have chosen to assist the reader with what I consider a more logical structure. Following this structure, one chapter is devoted to presenting the entire model, and the other chapters focus on various major aspects of the interaction between perception and action combining simulations from various articles. In my view, this structure better reflects the integrated character of the work and avoids unnecessary repetition of common or iteratively refined parts such as model implementations, theoretical background and simulation procedures.

The thesis is an integration of a number of articles I wrote in collaboration with co-authors. Note that this is reflected in the various chapters by the use of 'we' rather than 'I'. The interested reader is referred to these articles.

Haazebroek, P., & Hommel, B. (2009a). Anticipative control of voluntary action: Towards a computational model. *Lecture Notes in Artificial Intelligence, 5499*, 31-47.

Haazebroek, P., & Hommel, B. (2009b). Towards a computational model of perception and action in human computer interaction. *Lecture Notes in Computer Science, 5620*, 247-256.

Haazebroek, P., Raffone, A., & Hommel, B. *HiTEC: A Connectionist Model of the Interaction between Perception and Action Planning. Manuscript submitted for publication.*

Haazebroek, P., van Dantzig, S., & Hommel, B. (2009). Towards a computational account of context mediated affective stimulus-response translation. *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1012-1017). Austin, TX: Cognitive Science Society.

Haazebroek, P., van Dantzig, S., & Hommel, B. (2011a). A computational model of perception and action for cognitive robotics. *Cognitive Processing, 12*, 355-365

Haazebroek, P., van Dantzig, S., & Hommel, B. (2011b). Interaction between Task Oriented and Affective Information Processing in Cognitive Robotics. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 59*, 34-41.

Haazebroek, P., van Dantzig, S., & Hommel, B. (2013). How task goals mediate the interplay between perception and action. *Frontiers in Psychology*, *4:247.*

As my PhD project was embedded into an interdisciplinary robotics project I also got the chance to collaborate with scientists from other disciplines. Some of this collaborative work has not been integrated in this thesis, even though it contains some of the ideas captured therein; the interested reader is referred to following articles.

Broekens, J. & Haazebroek, P. (2007). Emotion and reinforcement: Affective facial expressions facilitate robot learning. In *Proceedings of the IJCAI Workshop on AI for Human Computing (AI4HC'07, Hyderabad, India)* (pp.47-54).

Lacroix, J. P. W., Postma, E., Hommel, B. & Haazebroek, P. (2006). NIM as a brain for a humanoid robot. In *Proceedings of the Toward Cognitive Humanoid Robots workshop at the IEEE-RAS International Conference on Humanoid Robots 2006*. Genoa, Italy.

Spiekman, M.E., Haazebroek, P., & Neerincx, M.A. (2011). Requirements and Platforms for Social Agents that Alarm and Support Elderly living Alone. *Lecture Notes in Computer Science 7072.*, 226-235.