Computing the motor-sensor map

Oswald Wiener and Thomas Raab

Kunstakademie Düsseldorf, Eiskellerstrasse 1, D-40213 Düsseldorf, Germany. **postmaster@kunstakademie-duesseldorf.de** **tr@sil.at**

Abstract: "Articulate models" subservient to formal intelligence are imagined to be heterarchies of automata capable of performing the "symbolic (quasi-spatial) syntheses" of Luria (1973), where "quasi-spatial" points to the abstract core of spatiality: the symbol productions, combinations, and substitutions of algebraic reckoning. The alleged cognitive role of internal "topographic images" and of "efference copies" is confronted with this background and denied.

An organism receives stimuli of a particular character and emits regular responses that are in general of an entirely different character. It can be regarded as computing a mapping from the afferent domain to the efferent range. Vice versa, the organism's environment "computes" a mapping from the organism's motor (etc.) output to a range of physical configurations; a subset thereof feeds back to the organism as updated stimuli. It seems useful to distinguish between a category of organisms that can muster simulations of the environment's computations in order to aid their own transformations of sensory input into motor commands, and a broader category of organisms that compute their responses more directly.

Let us for historical reasons call a mechanism capable of computing a particular mapping of the latter kind, a "sensorimotor schema." A particular sensorimotor schema does not have parts that co-operate in analogy to co-operating parts of the environment. It is a spatiotemporal mould for a particular class of stimuli. It cannot properly be regarded as a representation of that class (otherwise a hand, for example, would have to be termed a representation of the class of objects that it handles). Organisms relying exclusively on sensorimotor schemata exhibit various kinds of action readiness; they cannot correctly be said to entertain expectations. They remain on the level of what Grush has termed "sensation."

Organisms of the first kind, however, internalise relevant regularities of their environment by creating a realm of "formal schemata" that sits, as it were, on top of the realm of sensorimotor mapping and dominates it. Formal schemata work online by taking prerogative over input to the sensorimotor realm in order to establish perception proper, or by controlling sensorimotor control of efference. They can work offline to predict the outcome of external action and of state transitions of other schemata. They are necessary for the construction of new formal schemata qua solutions to problems, or can be used to supervise the building of new sensorimotor schemata to mitigate demand on the resources necessary for running formal schemata.

Furthermore, formal schemata can sometimes be used to infer generically what a particular external setup would *look* like, which amounts to problem-solving in the visual domain (the solution is a set of operations that, when executed, produce a picture in the environment). Because perception is the control of models by stimuli originating in external objects, the operations of the inference will produce some of the concomitants of seeing by priming sensorimotor schemata that would react to the relevant visual signals, should any of their kind arrive, and by pointing to formal schemata that could be activated in order to include additional information. This setting up of pointers is not a measurement (unless one wishes to treat "effect" and measurement as synonyms).

Whereas a sensorimotor schema can be conceived of as an "analogue" machine, a formal schema may didactically be compared to a string-processing device – a sequential, discrete-state, discrete-symbol automaton epitomised in the notion of the Turing machine (Wiener 1998; with a view to neuropsychology, a formal schema should rather be construed as a layered Petri net; with a view to introspection and "mental images," as a formation of Moore automata). Embedded in suitable running environments (heterarchies of other formal as well as of sensorimotor schemata that provide qualified control and parameter signals mediating, where applicable, connections with the external environment), any particular formal schema in an organism is indeed what seems to be intended by the term representation (utilisable *as such* for the organism): a *model* (Wiener 1988), a *structure* (Wiener 2002) of (the set of stimulus sequences that for the organism constitute) any object belonging to a particular set.

The simulations of the behaviour of external objects or even of the organism's own sensorimotor schemata (Piaget 1947; Rozin 1976), as accomplished by formal schemata, rest on functional equivalences of analogue devices and suitable symbol-processing automata. In their interactions the models generate and accept symbols that stand in bare *functional analogy* to stimuli on the one hand, and, on the other, to the organism's own efferent signals. An activated model's offline behaviour is thus determined by its own spatiotemporal regularities, as parameterised by the actual running environment and by other models' output functionally equivalent to the end products of sensory processing or to motor commands. It seems useful to assume the same format for all model output – no modal buffers, no efference copies effective on the model level. In the last analysis, mental images are pointers (to operations, viz., to models) that for their part can be operated on. Readers interested in the notions given are referred to Wiener (1996; 2000, and forthcoming).

Look again at Figure 7 of the target article. Because of some motor command, the effector changes the environment. The outcome of this event is measured, and the result of the visual measurement is a retinal image (sect. 4.4). The same motor command changes the state of the image emulator, which, as an emulator of the sensor sheet (sect. 4.4), now carries another fresh quasi-retinal image. At the same time, the motor command changes the state of the organism/environment model, and a modal measurement of the result projects a third image onto the inner eye's retina. The three images are somehow compared, and the resulting sensory residual – perhaps a pixel array having all the pixels active that did not partake in the formation of every single picture? – is used to correct the picture in the image emulator – perhaps by being added algebraically? – and to rectify the model state by virtue of a mapping of the sensory residual to, presumably, differential motor commands. Feedback from the model to the controller (the "motor centers") can now proceed: that is, information about the objects and states in its egocentric environment (sect. 5.1). The controller will somehow compare this information to its goal, itself specified in terms of objects and states in the environment (sect. 5.1), in order to update its motor commands.

Grush embraces the idea that perception is contingent upon matching the shapes of "percepts" and of internally generated images (Kosslyn & Sussman 1995). The principal component of his image emulator is a Kosslynian "visual buffer," implemented somehow in the manner of Mel (see Mel 1991). But Mel's dealing in changeable pictures, with their pixel transformations governed by neighbourhood relations, is successful – so far as it goes – only because there is just one moving complex (the "arm") that is accordingly identified without trouble. At the same time, this appears to be the most complicated case solvable by devices of that kind (we will not speak of Mel's run-of-the-mill backtracking mechanism that benefits from "representation without intelligence"). If such two-dimensional arrays of connectionist pixels were to carry geometric projections from spaces of higher dimensionality, and if the objects in those spaces were flexible and moving around independently, correct pictorial transformations and object recognition and tracking (even the "recognition" performed by contemporary computers) would be impossible. This consideration renders Grush's image emulator and the corresponding part of the target article's Figure 7 gratuitous.

Similar problems derive from Grush's adaptation of von Holst's and Mittelstaedt's notion of "efference copy" (cf. von Holst 1954). With the original authors the concept works because, again, the conditions they assume are quite clear-cut: The efference is oculomotor only, and the task concerns the entirety of the visual stimulus complex at the time of a single saccade. But to compute a

https://doi.org/10.1017/S0140525X04500092 Published online by Cambridge University Press

mapping from motor commands to pictures, a device would need input unique to each picture, a condition certainly not met by motor "plans" for human effectors. Of course, the lay would be different if the device were a model of the pictured object, already activated by other signals and thus all by itself constituting a specific interpretation of the accepted signals. We acknowledge that Grush points to this possibility, but in the case now at hand the input could be anything versatile enough to exploit the capabilities of the respective model.

Another problem of the same kind arises concerning the control of the controller. If Grush's sensory residual is questionable because the idea of picture subtraction does not make sense cognitively (while the idea of structure tests does), then the comparison of "environmentally specified" goal signals to feedback in the same format will work even less – unless the controller is able to use the model realm or, rather, reduces to a servomechanism of the latter. That would render the idea of model control by efference copies gratuitous.

Our arguments imply, or so we believe, that the idea of measuring forward models and Kalman filtering the results, if at all applicable to intelligent organisms, applies only to functions well beneath and possibly up to the level of sensorimotor schemata.

Motoric emulation may contribute to perceiving imitable stimuli

Margaret Wilson

Department of Psychology, University of California at Santa Cruz, Santa Cruz CA 95064. **mlwilson@cats.ucsc.edu**
http://psych.ucsc.edu/Faculty/mWilson.shtml

Abstract: First, I note three questions that need further exploration: how fast the emulator operates, compared to the real-time events it models; what exactly perceptual emulation, with no motor component, consists of; and whether images are equivalent to raw sensations. Next, I propose that Grush's framework can explain the role of motor activation in processing "imitable" stimuli.

A few questions . . . I shall begin by noting a few problems, puzzles, and gaps in the model. These are not intended to detract from the considerable accomplishment of Grush's proposal, but rather to suggest lines along which this set of ideas could be pushed forward.

(1) The timing problem. How closely in sync with the plant (the system being modeled) is the emulator? On the one hand, an emulator is supposed to be fast, bypassing delays in sensory feedback and allowing corrections in motor control to be made in a timely fashion. This requires an emulator that runs its simulation almost simultaneously with (perhaps even slightly ahead of?) the events in the external world. On the other hand, however, an emulator is supposed to correct for measurement error, by running measurement processes in reverse and employing Kalman filters. This is an architecture that adds several processing stages beyond simply sending sensory feedback to the controller, hence actually exacerbating the first problem that the emulator is asked to solve. Can both these functions be served by the same emulator, or is this architecture being asked to do too much?

(2) The perceptual simulation problem. Grush acknowledges that much of human cognition resides within the boxes in his Figure 7, rather than in the functional relations between them. In some respects, this is completely justified. For example, the box modestly labeled "measurement inverse" actually represents all of perceptual processing (to wit, reconstructing the distal stimulus from the proximal stimulus); but providing an account of this box is not the purpose of Grush's project. More problematic, though, is the box labeled "organism/environment model." It is here that, in the "degenerate case" of perceptual imagery without a motoric component, the entire business of imagery happens. And explaining imagery is a central part of Grush's purpose. Somehow, the "organism/environment model" has internalized principles and regularities that allow it to run simulations of the external world. How such a feat of simulation is possible seems to be the real heart of the imagery question. (Some possible beginnings of an answer to this are suggested below.)

(3) The images-aren't-pictures problem. Grush takes at face value the idea that modality-specific imagery is equivalent to unprocessed sensory input. In fact, this idea has been vigorously challenged (e.g., Reisberg & Chambers 1991; Reisberg et al. 1989). If a modality-specific image at all resembles a percept, as opposed to a raw sensory input, then the model in Figure 7 would need to be changed. It would not be possible to generate this kind of imagery by going all the way to an amodal three-dimensional representation (the organism/environment model) and then simply "looking at it" (the second measurement) to recreate the raw sensory input.

. . . and a proposal. Having issued these challenges, though, I would now like to turn to a different area in which Grush's model may offer considerable explanatory power: the perception and representation of stimuli that can be imitated with one's own body. There is growing evidence that such stimuli (which consist primarily of other humans' postures and movements) have a special status in the human cognitive system. Specifically, imitable stimuli appear to have privileged connections to motor representations of performing similar postures and movements with one's own body. Evidence for this comes from the literature on mirror neurons and on stimulus response compatibility, and from a variety of other sources (see Wilson 2001 for review). What is the purpose, though, of activating imitative motor programs in response to perceptual input, when there is no intention to overtly imitate? One possibility is that motor representations may actually play a role in perceptual processing.

As Grush suggests, an emulator which can capture predictable properties of perceptual events and simulate them in real time may serve both to assist perceptual processing and to recreate perceptual events in their absence. Indeed, the operation of just such a system appears to be behind the phenomenon of representational momentum, in which perceptual events are immediately misremembered as having continued beyond the point at which they ceased to be visible.

In most cases, representational momentum occurs for movement trajectories that are predictable based on simple geometric principles, such as rotation about an axis (Freyd & Finke 1984), circular paths (Hubbard 1996), spiral paths (Freyd & Jones 1994), oscillatory motion (Verfaillie & d'Ydewalle 1991), and predictable changes in direction such as bouncing off a wall (Hubbard & Bharucha 1988). These are the kinds of predictions one could build into a perceptual emulator without much difficulty (see Question 2, above). However, representational momentum also occurs for complex human motion (Verfaillie & Daems 2002; Verfaillie et al. 1994).

Based on what principles is the emulator able to predict the trajectories of complex biological motion? A likely answer is: based on principles of body biomechanics. That is, the emulator may have internalized the constraints of human body movement – the range of motion of joints, acceleration properties of muscles, and so on. In Grush's terms, the emulator may be an articulated model, isomorphic in its parameters to the relevant parameters of the human body.

From here it is only a small step to suggest that activation of motor brain areas in response to perception of conspecifics may be a functional part of an emulator, assisting in the perceptual prediction of those events. That is, just as in the case of emulating motor events and the case of emulating perceptual events as they are altered by self-motion, emulation of imitable stimuli may receive part of its drive from the controller. In each of these cases, the controller issues motor "commands" in the service of representation rather than (or in addition to) in the service of overt movement. The perception of conspecifics, then, may form an additional class of cognitive events that the emulation theory can help to explain.

https://doi.org/10.1017/S0140525X04500092 Published online by Cambridge University Press