Ethology and Functionalism

Behavioral Descriptions as the Link between Physical and Functional Descriptions

Matthias Scheutz
Department of Computer Science and Engineering
University of Notre Dame, Notre Dame, IN 46556, USA
Department of Philosophy of Science and Social Studies of Science
University of Vienna, Vienna, A-1090, Austria
mscheutz@cse.nd.edu

Abstract

In this paper, I introduce the notion of "behavioral state" as a means to bridge the gap between functional specifications of cognitive architectures and their physical implementations based on ethological methods of describing animal behavior. After briefly sketching some of the problems resulting from mere functional descriptions of cognitive architectures, I define the notion of behavioral state and discuss some of the properties of behavioral states that are relevant for describing and modeling minds. I show that behavioral states can serve as mediators between functional and physical descriptions of cognitive systems, arguing in particular that behavioral states can capture "distance in time", an essential aspect of real world cognition missing in mere functional descriptions.

1 Introduction

Ever since Descartes, philosophers interested in the mind have divided the world into a mental and a physical realm and consequently contemplated the relation between these two realms, a topic today widely known as the "mind-body" problem. While this problem is far from being resolved, today's most commonly held position on the "mind-body" relation in the philosophy of mind is *functionalism*, a view, which despite its appearance in many different forms is based on the central common claim that mental states *are* functional states. The general understanding is that mental states (i.e., states such as "believing that *p*" or "desiring *x*", or even psychological predicates such as "pain" or "pleasure") can be explained in terms of functional states and functional architectures.

Besides the fact that to my knowledge no one has ever attempted to specify concepts from folk psychology *in detail* using a functional architecture, the cognitive scientist who wants to understand and model cognitive systems will still face significant problems even if a complete functional specification of a given cognitive system could be provided: for one, the question of how functional states are related to physical states remains unanswered. Usually, philosophers assume that functional states "supervene" on physical states without paying particular attention to the question as to *how* (and consequently also *why*) they supervene.[1] In other words, what plays a secondary role (if at all), is of crucial importance to the cognitive scientist: (some) implementation details of the functional

---

[1] The questions of exactly how these states supervene on the physical and in what kinds of structures they are realized are rarely addressed in detail, let alone answered satisfactorily. This is most likely due to the fact that the notions of "realization" and "supervenience" are mostly used as unexplained "primitive" terms in the philosophical literature (which is quite surprising given the theoretical importance and practical consequences that hinge upon them). Although some have attempted more or less precise definitions of "realization"—e.g., Kim, Block, et al.—these definitions are not very helpful for those who, interested in building minds, are trying to understand the relation between architectures and their implementations.

architecture of these very abstract mental states. For example, it is not clear whether functional states can be realized as computational states (maybe only combined “computational-physical” states will realize functional states or maybe only physical states alone). And more generally, the question arises what the constraints are that a functional architecture imposes on systems implementing it: are functional descriptions besides being general enough to include all possible mental architectures specific enough to constrain the class of possible realizing systems in such a way as to suggest possible ways of implementing them?

It seems that relating functional states *directly* to physical states is very unlikely to succeed in the light of multiple realization arguments for functional architectures (the more complex the architecture gets, the less we will be able to see what kinds of possibly very diverse physical systems will share the functional specification).[2] The level of functional specification of the psychology of minds will be too high and abstract a level of description to suggest *possible implementations* of the functional states (not to mention all the problems connected with the involved notion of “implementation” or “realization” that seem to be largely ignored by the philosophical community).[3]

It is my conviction that functional specifications of psychologies are not sufficient to suggest ways of understanding and modeling minds. To be of any *practical importance* in modeling a mind at all, a level of description of a cognitive architecture has to incorporate at least *some* of the relevant physical properties of its possible implementations (which will constrain both possible implementations as well as functional architectures). In this paper, I will suggest such an intermediary level, which I call *the level of behavioral states*. This level of description is largely inspired by ethological studies of animal behavior (and to some extent by research in behavior-based robotics) and will therefore bear the insignia of its intellectual sources very visibly on its sleeves.

First, I will briefly point to one of the problems resulting from a mere functional description of a cognitive system (the “implementation problem”). Then I will introduce the notion of “behavioral state” and locate its place as mediator between functional and physical states, sketching briefly the role behavioral states could play in understanding, designing, and implementing (simple) cognitive architectures. Finally, I argue that behavioral states are sufficient to capture relevant aspects of cognition and, thus, provide an intermediary level of architectural specification located between functional and physical descriptions.

2 Functionalism

2.1 The Functionalist Picture

A functional specification of a cognitive architecture consists of a set of input states, a set of output states, and a set of “inner” or “functional” states together with a specification of how they are *causally* related. That way it is possible to determine what state a cognitive system will be in next, given the current state and all the input conditions.[4] While input and output conditions have to be tied to physical inputs and outputs, the functional states do not require a direct correspondence to their physical realizers as expressed in the phrase that “functional states supervene on physical states” (e.g., see Kim, 1997). This lack of a “direct” correspondence between functional and physical states is what gives functionalism its explanatory power, while keeping it metaphysically palatable: it com-

---

[2] Even with simple functionally specified objects this is problematic. Think of tables as functionally specified, for example, and consider all possible physical implementations of the specification “table” and what they could possibly have in common at a physical level.

[3] Note that this obviously does not hold for all functional specifications: a functional specification of an abstract finite state automaton, for example, can be easily related to physical states in a standard PC by “implementing” the automaton in a programming language.

[4] Of course, the behavior elicited by the organism realizing the cognitive system is specified as well.

bines advantages of behaviorist approaches to mind (i.e., considering solely the input-output behavior of an organism) with advantages of identity theories (i.e., mental state/event tokens are physical state/event tokens) leaving out the pitfalls of both such as the lack of being able to account for "inner states" in the former, and the requirement of type identities between mental and physical state/event types of the latter. Yet, this strength comes at a price: it is not clear what it means to *implement* or *realize* a functional architecture.

2.2 Implementation of a Functional Architecture

So what are the implementation conditions for a functional architecture? To say that a system implements a functionalist description is to require that in addition to the input and output mapping, it has to get the mapping of the inner states right. Usually, these "inner states" are assumed to be multiply realizable, i.e., many different, possibly very diverse physical systems will realize a given functional architecture. Therefore, the mapping between physical states and functional states has to be a many-to-one (very much in the spirit of Chalmers, 1997). Yet, inner states are viewed by functionalists as intrinsically relational states, being *mutually defined* by all states in the functional architecture (which is sometimes expressed by saying that they are defined by their "causal role" in the functional architecture).

To illustrate this interdependence, consider, for example, the following automaton, which has two inner states 'E' and 'O' standing for "even" and "odd". Depending on whether the number of '1's that the automaton has seen so far is even or odd, it outputs either 'a' or 'b', respectively.

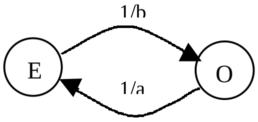

Figure 1 The even-odd transducer with two inner states.

A functionalist account (e.g., see Block, 1996) of what it means to be in state E would look like this:

> Being in E $=_{\text{def}}$ Being an x such that $\exists P\,\exists Q$ [x is in $P \wedge$ (if x is in P and receives input '1', then it goes into Q and outputs 'b') $\wedge$ (if x is in Q and gets input '1', then it goes into P and outputs 'a')].[5]

Since it is only claimed that there has to be an arrangement of physical states that corresponds to the functional states in a way that preserves inputs and outputs as well as transitions between states, it is possible for one physical state to serve as the instantiation of more than one functional state (and vice versa). Therefore, the correspondence between physical and functional states is not necessarily that of a mapping between physical types and functional types (let alone a 1-1 mapping), but rather that of a relation that preserves state transitions. "Implementation of a functional architecture", therefore, has to be viewed as some sort of "bisimilarity" between functional and physical architecture rather than some sort of *isomorphic* relation from a functionalist point of view.[6] As a conse-

---

[5] Note that the existential quantifiers could be viewed as ranging over properties or as picking out particular physical states of the system.

[6] The notion of "bisimilarity" is defined as follows: let I and O be two finite sets (e.g., the sets of input and output states, respectively) and let $M_1=\langle S_1,\rightarrow_1\rangle$ and $M_2=\langle S_2,\rightarrow_2\rangle$ be two structures with domains S_1 and S_2, respectively, where relation $\rightarrow_1$ is defined over $S_1\times I\times S_1\times O$ and relation $\rightarrow_2$ is defined over $S_2\times I\times S_2\times O$. These structures are then said to be *bisimilar* if there exists a non-empty relation R between S_1 to S_2 such that for all $s_1\in S_1$, $s_2\in S_2$, $i\in I$, and $o\in O$ the following two conditions hold: (1) if $R(s_1,s_2)$ and $(s_1,i)\rightarrow_1(t_1,o)$, then $(s_2,i)\rightarrow_2(t_2,o)$ and $R(t_1,t_2)$, and (2) if $R(s_1,s_2)$ and

quence, not every functional state might have a unique correspondence in the physical system, i.e., functional difference might not amount to physical difference, as it is possible that two different functional states are realized by the very same physical state (e.g., think of virtual memory systems in computers), a possibility that can complicate the search for a physical correlate of functional states (in section 4 I will address another essential difficulty of merely "causal" descriptions, namely their failure to capture "distance in time").

3 Behavioral States

3.1 An Ethological Perspective

To overcome the difficulties of tying functional specifications to physical implementations, I suggest to consider work done in animal behavior research as a venture point. According to animal behaviorists (e.g., McFarland, 1981), animal behavior can be categorized in terms of

(1) reflexes (i.e., rapid, involuntary responses to environmental stimuli)
(2) taxes (i.e., responses orienting the animal towards or away from a stimulus)
(3) fixed-action patterns (i.e., time-extended sequences of simple responses)

While (1) and (2) are solely connected to external stimulation, (3) can have a contributing "internal" component as well (fixed action patterns can be "motivated"; take, for example, the "egg-retrieving" behavior of the greyling goose, see Lorenz, 1981, or Lorenz and Leyhausen, 1973). All three kinds of behaviors can be combined in complex ways to form hierarchies of behaviors (see figure 2).

In these behavioral structures, behaviors form "competitive clusters", in which behaviors are mutually exclusive (e.g., in figure 2 the "fighting behavior" is such a competitive cluster comprising the mutually exclusive behaviors "chasing", "biting", and "display").

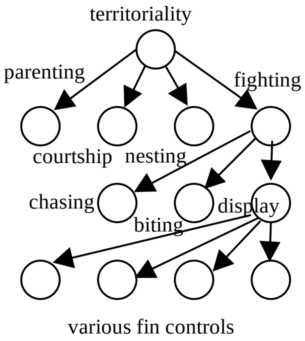

Figure 2 A part of a behavioral hierarchy for the male stickleback fish (see Lorenz, 1981). The various fin controls can be divided further into rays of each fin, the muscle fibers for each ray, and the motor neurons for each fiber.

To make these ideas of behavioral hierarchies more concrete, I will introduce the notion of *behavioral state*, which roughly corresponds to what is indicated by a "circle" in figure 2. Putting it crudely, a behavioral state is a state an individual is in if it performs a particular behavior (e.g., such as "food handling" or "looking out for prey").[7] "Behavior" is meant be understood in a *wide sense*

$(s_2,i)\rightarrow_2(t_2,o)$, then $(s_1,i)\rightarrow_1(t_1,o)$ and $R(t_1,t_2)$. For a detailed elaboration of the role of bisimulation in a theory of implementation and functional realization, see Scheutz (2000a).

[7] A note of terminology: while it is common usage to use "mental states" and "functional states" to refer to states of an individual's mind, the notion of state is not exclusively used to describe "static" entities, but often times serves the role of a general term that subsumes *states* as well as *events*, i.e., processes. In a sense, the term "behavioral state"

to include behaviors that are not necessarily observable from the outside alone (such as "memory recall" or "thinking", in general). Hence behavioral states are not simply combined input-output states, but rather they are some sort of "inner states" of an organism, states in which the organism is if it performs a particular kind of behavior. Note, however, that nothing is implied or claimed about a particular physical correlate of a behavioral state—it might or might not exist (I will return to this issue later).

Behavioral states are not restricted to "motor actions", but include sensory actions as well as more abstract proprioceptive and reflective actions (such as monitoring inner physiological states, generating images, producing plans, recalling poems, analyzing pictures, making logical derivations, etc.). The latter ones are more "abstract behaviors", which are mostly (if not completely) internalized and often involve solely parts of the cognitive architecture; in fact, they might not result in any externally observable change at all (a mathematician contemplating abstract objects and manipulating their representations in her mind, for example, might not need any stimulation from the outside world in performing this task, nor might any motor action result from it—this "brain in a vat"-idea with sustained cognitive activity whilst lacking external interaction seems to be at least conceivable in principle).

Memory and reflective processes, for example, are then viewed as special kinds of behavioral processes that lead to actions performed directly on the cognitive architecture, as opposed to the effectors of the individual which act on the environment.

In general, an individual will be in many behavioral states at the same time reflecting the fact that (1) some behaviors are contained in or shared among others (for example, searching for food as well as searching for a mate will both involve locomotion, despite the fact that the kind of search might be different), and (2) that many behaviors are performed in parallel (such as monitoring my hand as I move it to pick up an object).

3.2 Behavioral Architectures

In a sense, the classical ethological picture outlined above is mainly concerned with the relation between various behaviors, it only depicts (some of the) causal relations between behaviors, and is, therefore, really a functional specification of the behavioral architecture. Yet, partly *implicit* in and partly *external* to this picture is information about the time constraints as well as the strength of interactions and influences among behaviors (as studied and gathered by animal behaviorists). In other words, the picture is *incomplete* in so far as it leaves out essential implementation details that cannot be retrieved from a picture like figure 2 alone. Without these implementation details, however, some behaviors would not be the kinds of behaviors they are, since what distinguishes them from other behaviors might just be constraints on timing and strength of response (take, for example, a retraction reflex caused by touching a hot plate with your finger as opposed to the same movement being performed very slowly). Furthermore, the strength and configuration of interactions between behaviors is an integral part of their defining characteristics, which cannot be captured by a causal structure alone: suppose behavior A *causes* behavior B. Then this can happen in many different behavioral arrangements, for example, by A enforcing B directly or A suppressing C, which in turn inhibits B, or by A enforcing D, which enforces C, etc. Implicit in A (as defined by an animal behaviorist, say) is already information, which of these possible arrangements are realized in the animal. Hence, the causal structure might get restricted by the behavioral structure if (some of) the information implicit in the definition of behaviors is made explicit. In the following, I will briefly sketch how

---

should have been avoided in favor of "behavioral processes", as the latter emphasizes the dynamic character of the activity taking place in the individual. Following established terminology, however, I will continue using the term "behavioral state", even if (systematic) dynamic changes in the individual are being referred to.

behavioral states can be defined to explicitly incorporate some of the otherwise implicit aspects of behaviors.

3.3 The Structure of Behavioral States and Networks

First and foremost, each behavioral state has an *activation level* and a *behavior* associated with it. This activation may depend on any of the following factors (and additional factors could be considered):

(1) its own activation level
(2) the activation level of other states
(3) inputs from extereoceptive and proprioceptive sensors
(4) energy constraints (of the organism)
(5) decay over time

The behavior associated with a behavioral states can be simple (such as reflexes and taxes), or a more complex fixed behavior (such as fixed action patterns), or an even more complex adaptive behavior (which results from the interplay of fixed action patterns, reflexes, and taxes). The term "adaptive" indicates that the latter kinds of behaviors can change over time, i.e., they can be learned, altered, etc. (utilizing the dynamic interplay of behavioral states).

Behavioral states are connected via inhibitory and excitatory links to other behavioral states and possibly to sensors (via "information channels", i.e., filtering mechanisms that select parts of one or more sensory inputs and combine them in particular task-specific ways). Connections between behavioral states have a distance associated with them (expressed in terms of a time-lag), reflecting the "distance in space" that a signal has to travel from one locus of action to interact with another, allowing temporal as well as spatial integration of incoming signals.

Groups of behavioral states that are connected via mutually inhibitory links form so-called "competitive clusters". They inhibit each other to various degrees, while usually entertaining excitatory connections to lower and upper level states (and possibly to some behavioral states of other clusters at the same level as well). In such a cluster the behavior associated with the highest activated state will become activate and all behaviors of the other states are suppressed.[8] This way hierarchical structures similar to the one in figure 2 can be defined which reflect the relationship between behaviors and in part also the complexity of each behavior associated with the various states (the lowest levels corresponding to simple reflex-like, reactive behaviors—this level has been explored in great detail in behavior-based robotics, e.g., see Arkin, 1992, or Brooks, 1986).

With respect to the spread of activation, networks of behavioral states are very similar to I(interactive) A(ctivation) and C(ompetition) networks (e.g., see Rumelhart and McClelland, 1986). Therefore, results from connectionist research about effects such as "blocking", "settling", "oscillation", "hysteresis", and others (often) apply *mutatis mutandis* to behavioral networks as well. The essential difference between IAC networks and behavioral networks is that the behavior associated with a behavioral state could affect the activation level of the very state itself as well as the activations of other states *via environmental feedback*. For example, a behavioral node representing the "search for black objects in visual field"-behavior might initiate ocular motor commands that lead to the detection of a small black object by another node, which in turn inhibits the search node, thus

---

[8] There is evidence that similar mechanisms are at work in animals that inhibit all behaviors with lower activation values, e.g. see Lorenz (1981).

decreasing its activation, which in a mere IAC network (lacking environmental feedback) would have otherwise not decreased.[9]

As already mentioned, not all behaviors will involve physical effectors; in fact, only low level behaviors will directly exert influence on them (these are behaviors that would normally be localized in what roboticists refer to as "reactive layer"). Higher level behavioral states will mostly operate on structures internal to the cognitive system (these states would be situated in the "deliberative layer"). For example, a "retrieve image of mother" node (assuming for a moment there is such a node), might initiate a search in long-term memory (possibly involving other behavioral states) for a particular image that is associated with the individual's mother. Or a "project-hand-move-forward" node might initiate a "simulated" hand movement in an emulator circuit, which is used to plan motions, resulting in a change in the circuit and as a consequence in other behavioral nodes (such as "collision detectors" in the emulator circuit, etc.).[10] A behavioral network divided into a layered structure consisting of a reactive and a deliberative layer is schematically depicted in figure 3 below.

There are special cases of behavioral states that do not have any behavior directly associated with them. Instead of initiating an action directly, they contribute to behaviors indirectly by influencing other behavioral states, and can, therefore, assume the role of affective states. A state corresponding to "hunger", for example, might receive inputs from proprioceptive sensors (i.e., a sensor monitoring the blood sugar or, more generally, the energy level) and exert positive influence on other states such as "search-for-food" (e.g., see Scheutz, 2000b). That way it is possible to entertain states that do not directly and immediately "cause" the individual to act in a particular way, but might have indirect, long-term effects on the individual (e.g., depression, memory loss, etc.).[11]

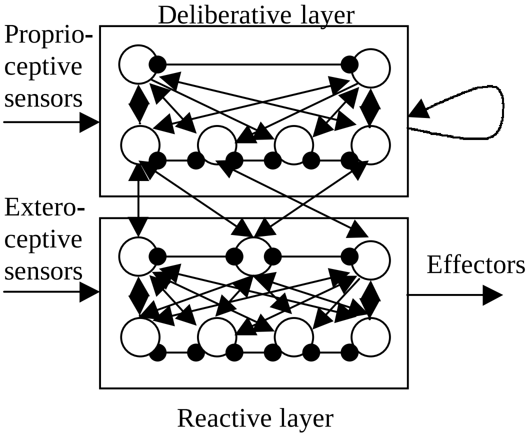

Figure 3 A hierarchy of behavioral states viewed as a two-layered architecture consisting of a deliberative and a reactive layer. Links with arrows indicate excitatory connections; links with circles inhibitory ones. While the behavioral units in the reactive layer operate on effectors (performing behaviors such as navigating through the environment, avoiding obstacles, etc.), behavioral nodes in the deliberative layer do not operate on effectors directly, but rather perform internal operations (such as memory lookups, symbolic combinations, etc.).

---

[9] While environmental feedback can obviously be simulated with neural networks, the neural architectures that incorporate such feedback will be different from behavioral architectures that perform the same function because of their intrinsic embeddedness in the real world. Since it is one of the design principles of behavioral architectures that they can rely on environmental feedback resulting from the behaviors of activated behavioral states, this property has to be taken into account in modeling cognitive architectures.

[10] I am currently investigating various possibilities of implementing simple emulator circuits using behavioral states.

[11] Compare this to standard philosophical talk about "pain *causing* wincing and groaning, etc.", where it is never clear whether pain always causes all the behaviors, exactly when the effects surface, whether showing the effects is necessary and/or sufficient for the individual to have pain, etc.

4 The Case for an Intermediate Level

4.1 The Relations between Physical, Functional and Behavioral States

So far, I have not explicated how physical and functional states relate to behavioral states as defined above. From an implementation perspective, behavioral states can be realized in many ways in different physical substrates. In brains, for example, they could correspond to a single neuron or to a group of neurons. They could be realized solely neuronally or maybe by involving other systems (such as the hormonal system) as well. Another physical medium, in which behavioral states can be realized, is the silicone of computers: computers can implement behavioral states by virtue of computational processes.

Some behavioral states might be (directly) "implemented" in the system in the sense that there exists a corresponding physical state or a set/sequence of physical states that are in *type correspondence* with the behavioral state. Other behavioral states might "supervene" on physical states in that there does not exist such a type correspondence—note that programs running on modern operating systems with virtual memory architectures exhibit such supervenience relations: when a program does not entirely fit into physical memory, it is loaded in parts on an "as-needed" basis, where different virtual memory locations get mapped onto the same physical memory location.

Another possibility for behavioral states to have no *fixed correlate* at all is to be only *partially* implemented (see Sloman, 1998) or to depend on environmental conditions (e.g., in terms of other behavioral states and/or environmental states—an example might be my performing the multiplication algorithm using paper and pencil: I am in a behavioral state which is implemented by a number of other states such as states of the paper and pencil, several visual routines, rule-retrieving memory processes and rule-following routines, etc.).

Behavioral states implemented in (sequences of) physical states are tightly coupled to their physical realizers (still allowing for multiple realizations), while behavioral states supervening on physical states do not exhibit such a coupling at all. They are realized by some physical states, but they might not show any systematic correlation to their realizers. For example, consider two networks of behavioral states, which are *functionally* identical except for the fact that the first explicitly implements a higher level behavioral state called "avoid-obstacle", which is active if the agent is engaged in obstacle avoidance behavior. The second one does not have such as state, but can still control the same obstacle-avoidance behavior. In this case, the behavioral state "obstacle-avoidance" has a physical correlate in the former and no fixed physical correlate in the latter (what corresponds physically to the "obstacle-avoidance" state in the latter is a complex sequence of patterns that might, under different circumstances, not correspond to this state at all, e.g., if the agent follows another agent, which is avoiding obstacles, and thus is a "follow other agent" state, which by pure chance causes it to go through the same sequence of physical states... see also Pfeiffer and Scheier, 1999, ch. 12 for another example).[12]

This aspect of behavioral states seems very similar to the kinds of functional states about which philosophers tend to worry, and maybe most of the "high-level" functional states such as "belief states", etc. are not directly (i.e., physically) implemented in the system (often the term "emergent" is used in this context). Even so, these kinds of rather abstract behavioral states still retain one aspect lost in the mere "causation talk" of functional architectures, and that is *time*!

---

[12] Note that it should be possible to derive, beyond the causal propertied, the temporal properties of the "obstacle-avoidance" state from the interaction of the (physically) implemented states.

4.2 Causation and Time

It has been pointed out by philosophers (e.g., see Chalmers, 1997) that there is an essential difference between functional descriptions of physical systems like clocks, combustion engines, CD players, etc. and the functionalist descriptions of minds: in the former case some aspects of the physical structure matter, they are essential to any system realizing the functional architecture. Thus, these physical aspects are (if not explicitly, so then implicitly) retained in the functional architecture, thereby constraining the set of possible realizers. In the latter case, however, it is the very functional structure itself—so it is claimed—that matters, that is, the patterns of causal organization regardless of the underlying physical structure. Therefore, only causal organization, or put differently, "the flow of causation" is retained in functionalist abstractions from the physical as *the essential aspect* of minds. But is this really true?

Real minds are intrinsically tied to their environments and thus affected by the temporal structures imposed on them. Timing plays a crucial role in every aspect of a cognitive architecture pertaining to the proper functioning and survival of the organism. Many recent studies in cognitive science emphasize the importance of time as opposed to "mere temporal order" (see, for example, Port and van Gelder, 1995).

What distinguishes *time* from mere (temporal) *order* (as implicitly provided by the notion of causality) is that in addition to order a *metric* is defined (on the set of time points), that is, a notion of *distance* in time. This notion of distance in time allows one to differentiate functions according to their temporal behavior that would otherwise be indistinguishable. Take, for example, two microprocessors that work at different clock speeds—functionally they are identical, yet there is an essential difference between them, which is usually also reflected by any price tag put on them: their speed (another example of a function, where time is the distinctive factor, would be vowel production and recognition).

Is it problematic that causation alone does not suffice to capture the temporal structure of cognitive architectures? I would claim: Yes. Imagine two different physical systems that share the same functional specification of a human mind, one a regular human, another the People's Republic of China "implementing the human brain" at a much, much slower pace (to use Block's example). A human body controlled by the People's Republic of China would fail terribly in the real world, because it could not react to its environment in due time.[13] Well, one might say, it would do just fine if everything surrounding it, that is, its environment had been "slowed down" appropriately. This objection, however, strikes me as severely flawed, since it would entail *a completely new physics* (as in our physical universe certain processes have to happen at a certain speed otherwise they would not be the kinds of processes they are). Whether a "slowed down version" of a human mind could control a "slowed down version" in such a "slowed down universe" (with possibly completely different physical properties) seems too speculative a question to be taken seriously. What seems to be a productive approach, however, is to ask whether it is possible to understand a certain architecture (that evolved or was designed to meet the temporal constraints of its environment) at a mere causal level? I suspect that the answer would be *no* for systems that are sufficiently complex (like brains of vertebrates or VLSI microchips, for that matter).[14]

---

[13] Many parts of our cognitive system have especially developed to meet time constraints of the environment. There is evidence for neural as well as chemical internal clocks (that work at certain clock rates), oscillator circuits that adapt to external cycles, etc. None of this would work if the system ran at 1/10000th of its regular speed. The same is true for digital circuits that have been designed to work at certain clock rates.

[14] It is easy to imagine that nature came up with all kinds of "hacks" to solve timing problems which could and would have otherwise be implementated very differently. To give an example from computing, imagine a video conferencing system used to transmit video information accross the internet. Because of current traffic on the net and

If, on the other hand, causal structure were augmented by temporal constraints (i.e., information about distance in time between causally connected states), then this would in theory suffice to capture an essential aspect of possible physical implementations of the functional architecture. It would, for example, allow us to model the functional architecture computationally, i.e., to implement a virtual machine that abides to the temporal constraints (as many computational descriptions can handle temporal metrics, just take programming languages for real-time systems).

Behavioral states, therefore, seem to be an abstraction, which can be implemented computationally, and thus realized physically on computational systems. At the same time, behavioral states are abstract enough to capture aspects of minds that seem to be intrinsically connected to their causal structure and not to their physical realization ("organizational invariants" as Chalmers, 1997, puts it), thereby connecting them to functional descriptions of cognitive architectures.

5 Conclusion

The level of description of behavioral states is *intermediate* and *intermediary*, because it specifies states that could be realized in many different physical ways (in neural architectures, but possibly also in digital ones, and others), yet retains at least one crucial physical and causal aspect not retained in mere functional descriptions: (distance in) time! By explicitly incorporating time, behavioral states make it possible to model the temporally extended interactions among different parts of a cognitive system as well as interactions of the cognitive system with its environment. The level of description of behavioral states might, therefore, not only prove useful for constructing systems that exhibit complex causal interactions (such as minds), but also for explaining *how* functional states are related to physical states by viewing them as (not necessarily disjoint) collections of behavioral states.

References

Arkin, R. C. (1989) "Motor Schema-Based Mobile Robot Navigation". *International Journal of Robotic Research*, Vol. 8, No. 4, 92-112.

Block, N. (1996) "What is Functionalism?" *The Encyclopedia of Philosophy Supplement*, Macmillan.

Brooks, R. A (1986) "Robust Layered Control System for a Mobile Robot". *IEEE Journal of Robotics and Automation*, Vol. RA-2, 1, 14-23.

Chalmers, D. J. (1997) "A computational Foundation for the Study of Cognition", (published on the internet)

Kim, J. (1996) *Philosophy of mind*, Westview.

Lorenz, K. (1981) *The foundations of ethology*, New York, Springer Verlag.

Lorenz, K. and Leyhausen, P. (1973) *Motivation and Animal Behavior: An Ethological View*, Van Nostrand Co., New York.

McFarland, D. (1981) *The Oxford Companion to Animal Behavior*, Oxford University Press.

Port, R. and van Gelder, T. (1995) *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press, Cambridge.

McClelland, J. L. and Rumelhart, D. E. (1986) *Parallel Distributed Processing*, Vol. 1 and 2, MIT Press, Cambridge.

Pfeiffer, R. and Scheier, Ch. (1999) *Understanding Intelligence*, MIT Press, Cambridge.

---

to meet real time constraints it only sends partial information of each image, which has to be reconstructed as much as possible from previous images on the other side. It seems that it would be very difficult (if not impossible) to judge out what the system does from the program code alone.

Scheutz, M. (2000a) “Computational vs. Causal Complexity” (forthcoming in *Minds and Machines*).

Scheutz, M. (2000b) “Surviving in a Hostile Multi-Agent Environment: How Simple Affective States Can Aid in the Competition for Resources”, *Proceedings of the Thirteenth Canadian Conference on Artificial Intelligence*, Springer Verlag.

Sloman, A. (1998) “Supervenience and Implementation: Virtual and Physical Machines” (published on the internet)