Do somatic markers need to be somatic? Analogies from evolution and from hardware interlocks

Colin G. Johnson

*Computing Laboratory
University of Kent
Canterbury, Kent
CT2 7NF, England
C.G.Johnson@kent.ac.uk

Abstract

This paper considers Damasio's concept of the *somatic marker* from two new perspectives. The first of these considers them from the point of view of Dawkins's concept of the *extended phenotype*. This is used to develop the idea of the *extended somatic marker*, *viz.* a marker which uses some non-somatic feature of the external world in a similar fashion to the somatic marker. Secondly an analogy is developed with the concept of *hardware interlocks* in safety-critical systems. This is used to suggest why it is important that somatic markers are bodily states and not just mental markers.

1 Somatic Markers

Damasio Damasio (1994) has introduced the notion of the *somatic marker*—a bodily state which plays a role in cognition, in particular the direction of attention. More specifically, a somatic marker is some bodily state which is generated as the consequence of some mental process. This state is then *reperceived* by the mind, and as a consequence the mental state changed. An example of such a marker is the rapid onset of nausea upon witnessing an act of violence. This bodily state does not have any immediate relevance to the mental state which has generated it, in contrast, say, to a feeling of nausea generated by viewing a plate of rotting food. Some such states might be explained away as side-effects. For example a rapid change of hormone levels upon witnessing violence in preparation for running from the danger might also trigger nausea.

However the somatic marker hypothesis suggests that such reactions are not mere side-effects. Instead they are a way of generating a rapid shift of attention, using the body state in an arbitrary fashion to draw mental attention to the current situation. The presence of the marker in the body draws the mind's attention towards it, and as a consequence the mind if focused on the meaning of that marker. It is plausible that such phenomena are exaptations Gould and Lewontin (1979) from unwanted physical reactions to change in body state as discussed above.

This can be seen as an aspect of mind which is realised away from the usual mental substrate. The somatic response is being used as a way of carrying out a process (bringing the attention of many mental processes together to focus on a single danger point) which cannot be carried out within the computational model implemented on the substrate.

The aim of this paper is to consider why the markers in question need to be *somatic* as such. Two aspects of this question are considered. Firstly, would it be possible for markers to extend beyond the body? This is explored with reference to Dawkins's concept of the *extended phenotype*. Secondly, why is it important that such markers be in the body, instead of being more simply realised by mental markers? This is explored with regard to the idea of *hardware interlocks* in engineering design.

2 Could "somatic" markers extend beyond the body?

Why do markers need to be *internal* body states. Is there anything which is special to the body which means that the markers could not instead be realised elsewhere in the world, external to the body? Might some of our actions in the world act as triggers to affect, perceived directly through the usual perceptive system rather than by bodily self-awareness?

One approach to this draws on ideas from Dawkins's book *The Extended Phenotype* Dawkins (1982). In biology, the *phenotype* is the expression of a gene or set of genes in the world. This encompasses both the aspects concerned with the physical structure of the creature and through the ways in which genes have influences on behaviour. For example we can talk about the "blue-eyed" phenotype versus the "brown-eyed phenotype" of some animal. This is distinguished from the "genotype", i.e. the set of genes of interest. Sometimes more than one genotype can give rise to the same phenotype (e.g. where there are regressive traits).

The difficulty starts when we want to say where the boundary of the phenotype lies. Clearly certain things are in the phenotype for certain. A clear example of this is the sequence of proteins associated with a particular expression of a particular gene. A standard definition would extend this to the whole body; genes influence the growth, development, and activity of the body (alongside other influences).

Dawkins's argument is that it is naive to simply say that everything inside the body should be considered to be phenotype, whereas everything outside should not. For example consider an imaginary species of bird in which the male has a gene which predisposes itself to mate with females which have blue feathers. It could be argued that this gene is also a gene for blue feathers in the female, as as a result of the presence of the gene blue feathers will spread through the female population. To abstract this, the genotype in the male bird is having a phenotypic effect in the female bird. Why should we regard the gene's effect on the feathers of the female bird in any different way to another gene which causes the male bird to have red eyes?

A similar kind of argument can be made about the somatic marker hypothesis. Damasio argues for a body-minded brain in which we create emotions via "somatic markers". These work by parts of the brain recognizing an emotionally charged stimulus, and then rather than creating a direct link to an action on that stimulus, the "marker" consisting of a bodily reaction is created. This is then re-perceived by the brain as is the basis for action or for rapid alteration of emotional state.

Why do these markers have to be physically internal to the body? It would seem that the same reasoning could be applied to markers which I leave in the external world when I have an emotion. For example if I am anxious then I might scribble on the pad of paper in front of me, without attending to this scribbling. This could then become a marker, in this case perceived via the eyes rather than through internal perception of bodily state. Why should it matter whether I use a bodily state or an external state as the substrate for the marker?

It may be that there are reasons why somatic markers need be somatic. One could be that the speed of reaction required is just too quick to be capable of being carried out by the external perceptive system. Another more convincing explanation is that the reason we use somatic markers is to communicate with multiple brain regions in a simultaneous and co-ordinated way, and therefore we need something which can be perceived in a direct way by different parts of the brain.

This might be a continuum effect. An example of a thing which might be seen as either an external or somatic marker is biting nails when anxious. This is in many ways an external physical process, nonetheless we can perceive the nail state internally via soreness of fingers. There must be other similar examples. Perhaps nail-chewing is "causing" the anxiety (in the sense of being part of the causal chain between subconscious perception of an anxiety-producing stimulus and the affective response) rather than being an epiphenomenon of the emotional state.

3 Why do markers need to be confined to the body?

So far we have considered why it is that the somatic marker need be constrained to the body, and is it important to make a body/non-body distinction. Now we address the opposite question: why is it not sufficient for the marker to be a mental marker? Why not just make a "mental note"? Whilst there are circumstances where a truly somatic marker can get transformed into a mental process in the limbic system Damasio (1994), this is not always the case; markers are not always transferred in this fashion. It is interesting to consider whether there might be reasons why the evolution of the mind might have led to the markers being body-centred rather than mind-centered.

One reason may be for safety. In the design of complex systems involving computer-controlled mechanical and electrical devices it is common for there to be conservative safety devices included in the system known as *hardware interlocks* Leveson and Turner (1993); Leveson (1995). A hardware interlock is a device which is independent of the main control system, and which is designed to monitor just on small aspect of the system, typically by using its own sensor system. So for example in a radiotherapy device, an interlock might exist which monitors the output of radiation, and if more than a certain amount is let out in one minute, the interlock shuts down the device completely.

Hardware interlocks are designed to be parts of the overall system which do not depend on the abstraction offered by the overall control system. For example they do not take information from the main system sensors, nor do they use the main control system e.g. for timing, and they do not sit upon the operating system abstraction used by the controlling structure. To do this would compromise their role as a safety-critical component; they provide a reassurance of safety because they are separate, they are independent from the main abstraction. If the main sensors go wrong, or the builder of the controller has misunderstood the relationship between the abstraction offered by the operating system and the real hardware and software, it does not matter.

One important role in the body-mind system is to react quickly and reliably to dangerous phenomena. There would seem to be a *prima facie* case for thinking that if engineers consider the use of such hardware interlocks as an important way of responding to danger in computer-controlled systems, evolution may have created such interlock systems for dangers to animals.

It may be that our body-grounded response to danger is a response of this kind. Instead of making a mind-centered judgement about the danger of a situation, we instead make a rapid decision based on a few simple cues. One characteristic of hardware interlocks is that they typically work on a small number of basic sensors which facilitate a conservative approximation to safety. The same may be true of interlocks in the mind-body system: our sensory system perceives a small number of simple "danger signals" (such as a rapid movement) and triggers an action within the body immediately. This "massive synchronization" acts as a counterpart to the more commonly-discussed "massive parallelism" of the neural-network-based mind.

Typically the fact that the brain is a unified system with all aspects connected and mutually-accessible is seen to be to its advantage. Similarly the unity found in a complex software system is often seen as being to its advantage; instead of having to connect individual components together as needed (as might be the case in an electronic system) all information is passed to a central repository and accessed as needed. However in some situations it is necessary both with computers and with minds for the complete attention of the system to be directed towards one thing. Hardware interlocks provide a way for such responses to "leap out" of the complexity of the control software for certain emergency situations. This non-decomposability, and the consequent need for a powerful way of leaping out of the complex interactions, would seem to be particularly strong for neural-network-based systems where the system is highly non-decomposable.

Acknowlegdements

Many thanks to the reviewers for interesting suggestions about the ideas contained in this paper.

References

A.R. Damasio. *Descartes' Error : Emotion, Reason and the Human Brain*. Gosset/Putnam Press, 1994.

Richard Dawkins. *The Extended Phenotype*. Oxford University Press, 1982.

S.J. Gould and R.C. Lewontin. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proceedings of the Royal Society of London, Series B*, 205(1161):581–598, 1979.

N. Leveson and C. Turner. An investigation of the Therac-25 accidents. *IEEE Computer*, 26(7):18–41, 1993.

Nancy Leveson. *Safeware: System Safety and Computers*. Addison-Wesley, 1995.