

From Cognition to Consciousness: a discussion about learning, reality representation and decision making.

David Guez
Faculty of Applied Science
University of Canberra ACT 2615
Australia
david.guez@mac.com

Received 10 June 2010/Accepted August 8 2010

Abstract

The scientific understanding of cognition and consciousness is currently hampered by the lack of rigorous and universally accepted definitions that permit comparative studies. This paper proposes new functional and unambiguous definitions for cognition and consciousness in order to provide clearly defined boundaries within which general theories of cognition and consciousness may be developed. The proposed definitions are built upon the construction and manipulation of reality representation, decision making and learning and are scoped in terms of an underlying logical structure. It is argued that the presentation of reality also necessitates the concept of absence and the capacity to perform transitive inference. Explicit predictions relating to these new definitions, along with possible ways to test them, are also described and discussed.

Keyword: animal cognition; decision making; Umwelt; representation of reality; associative learning; transitive inference; Open Program; stereotypic behaviour; logic; rules

Publisher version: Guez, D. (2010). From Cognition to Consciousness: A Discussion About Learning, Reality Representation, and Decision Making. *Biological Theory*, 5(2), 136141. doi:10.1162/BIOT_a.00028

1 Introduction

The scientific study of cognition and consciousness is currently hampered by the lack of straightforward and universally accepted definitions. Without rigorously defined boundaries, in-depth, comparative research studies into these phenomena and their evolution cannot be achieved. Heyes (2000) for example, states a range of existing definitions for cognition but describes them as being “rough”. Some use the term “cognition” to mean goal directed behaviour that explicitly excludes associative learning (*e.g.* Dickinson and Balleine, 2000), whilst others encompass more or less the whole of animal behaviour and envision “cognition as information processing” (Shettleworth, 2000). Others, such as Zentall (1993), suggests that it is easier to specify what cognition is not rather than what it actually is.

It has been often proposed that cognition is what is left after “simple associative process and stimulus generalisation” have been ruled out (*e.g.* Tomasello and Call, 1997). The main issue with such an approach is to dismiss the possibility that the brain and its relation to behaviour can be understood within a unique framework. Too often it is postulated that “the simple associative process” explanation is more parsimonious. However, this is certainly not the case if, in order to accommodate the body of work that is not in accordance with the predictions made by classical theory of learning, *ad hoc* assumptions or additions to the theory are made (*e.g.* Haselgrove, 2010; Durlach and Rescorla, 1980). Obviously, modification of theories in light of new evidence is necessary as long as these modifications are applied consistently and not used solely to explain single results. Furthermore, having specific theories for different subsets of brain function and its relation to behaviour is certainly not parsimonious.

Theories are a valuable tool to the advancement of science and as Popper (1963) has pointed out it is the fact that they are falsifiable that makes them distinctively scientific. One could argue therefore that the rescuing of theories by *ad hoc* assumptions under the guise of parsimony is hardly scientific since it forbids falsification. In the process of science, theories should therefore be developed, and critically tested. Once falsified the theories field of application should be reduced (if possible) or the theory should be abandoned altogether. The Ockham’s Razor principle or parsimony principle should be used within a theory to choose the simplest possible principle able to explain the largest body of knowledge possible.

In order to develop theories one needs to have a definition of the problem that is attempting to be solved. Without such a definition theories are useless. For this reason I propose new and functional definitions of cognition and consciousness as described below. In order to be useful and avoid misunderstanding these definitions need to be devoid of ambiguity, and this is why I will explain the key concepts that constitute these definitions and propose ways in which to test them.

Cognition - a definition:

One possible functional definition of Cognition could be:

The capacity to construct and possibly manipulate one's predictive representation of reality and to act upon it.

In order to use this definition of cognition one needs to fully explain each part of the definition and this is what I will apply myself to in the following.

2 Representation of reality

A representation of reality of a given organism is dependent on its subjective world or Umwelt (Von Uexküll, 1965). It is constructed from an open program (Mayr, 1974) by learning and from developmental pre-wiring of the brain, and the interaction of the two. That is, that learning may change the pre-wiring of the brain. These two components and their interaction constitute the animal's representation of reality.

To be useful, a representation needs to be predictive in nature and evolve with experience. The possibility of evolution of the representation may be variable and this possibility is dependant on the amount of open program available to the given organism. Therefore, the representation of reality of a given organism is constituted of predictive templates of reality obtained by learning, which are superimposed on a pre-existing structure generated during development in a given species. It is therefore dependant on the perceptive world of the animal.

As the name 'representation' suggests it implies an underlying logical structure, and to be predictive this logical structure needs to be conclusive and complete; this representation is the knowledge of an organism. A structure that is logically complete and conclusive can be obtained using a single logical operator such as incompatibility (\uparrow) from which all logical operators can be constructed (including the negation or concept of absence, which will be discussed later), or from logical operators such as AND (\times), OR ($\dot{+}$), or the Conditional Proposition (\longrightarrow) when used in conjunction with the logical negation that will be equated here with the notion of absence (see Figure 1 for the "Truth Table" of these operators). Here it is important to understand that this notion of absence exists only through the notion of presence; that is, that we cannot conceptualise the absence of something if we do not know that it can exist.

For this reason a logical structure based on the operator AND or OR (with the negation) is improbable because translating a predictive relationship such as if A then B (where A and B are perceived events or features of the subjective world) in terms of OR or AND comes to:

$$A \longrightarrow B \iff \bar{A} \dot{+} B \iff \overline{A\bar{B}} \quad (1)$$

This is a representation of a present object in terms of its absence. It is therefore most likely that the underlying logic of an animal's representation of reality is based upon the

1. ‘+’, **inclusive OR**, $A + B$ is read as **A OR B**. The Truth Table for $A + B$ (**A OR B**) is given below, by convention true=1 and false=0.

A	B	$A + B$
0	0	0
0	1	1
1	0	1
1	1	1

2. ‘ \times ’, **exclusive AND**, $A \times B$ or AB is read as **A AND B**. The Truth Table for AB (**A AND B**) is given below.

A	B	AB
0	0	0
0	1	0
1	0	0
1	1	1

3. \bar{A} is the negation of A , if A is true then \bar{A} is false, if A is false \bar{A} is true. $\bar{A} = 1 - A$

4. ‘ \rightarrow ’ is the conditional proposition, $A \rightarrow B$ is read **if A then B**. The Truth Table for $A \rightarrow B$ (**if A then B**) is given below.

A	B	$A \rightarrow B$	$\bar{A} + B$
0	0	1	1
0	1	1	1
1	0	0	0
1	1	1	1

Note that $A \rightarrow B$ is equivalent to $\bar{A} + B$.

5. ‘|’ is the incompatibility or NAND (not AND), $A|B$ is read **A is incompatible with B**. The Truth Table for $A|B$ is given below.

A	B	$A B$
0	0	1
0	1	1
1	0	1
1	1	0

6. ‘ \perp ’ is the operator NOR (not OR), $A \perp B$ is read **A not OR B**. The Truth Table for $A \perp B$ is given below.

A	B	$A \perp B$
0	0	1
0	1	0
1	0	0
1	1	0

7. ‘ \iff ’ is the Logical Equivalence. It is a tautology and therefore always true. It is a rule of inference. $A \iff B$ is read **A is equivalent to B**.

The Logical Equivalence must not be confused with the logical operator biconditional, sometime called inclusive AND. (For information the biconditional Truth Table is given in 8 below.)

8. ‘ \iff ’ is the operator biconditional or inclusive AND. The Truth Table for $A \iff B$ is given below.

A	B	$A \iff B$
0	0	1
0	1	0
1	0	0
1	1	1

Figure 1: Conventions: The Truth Table and the notation used herein for each of the logical operators, OR, AND, NOT, Conditional Proposition, Incompatibility (NAND), NOR, and Biconditional are given in 1-6 and 8 respectively. The notation for the rule of inference “equivalence” is given in 8.

incompatibility operator ($|$ or NAND operator but also the NOR operator \perp), or upon the conditional proposition and the concept of absence, or upon the conditional proposition and the incompatibility operator. This operator allows the establishment of the concept of absence. In any case, the concept of absence is a necessary feature of the underlying logical structure. If we were to represent equation 1 using exclusively the incompatibility or the NOR operators we have:

$$\begin{aligned} A \longrightarrow B &\iff A|(B|B) \\ &\iff ((A \perp A) \perp B) \perp ((A \perp A) \perp B) \end{aligned} \quad (2)$$

We can see here that each event A and B are represented also by their presence. Nevertheless, a representation of reality needs to be directional, and at a minimum, needs to be able to encode for notions such as before, during and after. This directionality can be obtained by encoding the relation between events of the world as “conditional propositions”. This idea is further reinforced if we notice that the connection of two neurons undergoing long term potentiation is a biological equivalent of the conditional proposition. We are compelled to admit that most probably the basic logical operator underlying an animal representation of the world must be the conditional proposition (Guez, 2009).

However, we cannot exclude the possibility that the principle of absence is built upon the incompatibility operator, but it may be as simply admitted that it is based on the concept of non contradiction; that is, that something cannot be present at the same time as being absent. It is important to note that the principle of absence can only be built on the knowledge of the possibility of presence (for a logical structure of knowledge based on the conditional proposition see Guez (2009)).

In consequence, because the concept of absence is a necessary feature of a representation of reality, it should be able to be demonstrated in any organism equipped with a neuronal network. A demonstration of the notion of absence can be obtained when using a novelty detection test. If an animal explores a site more after the removal of an object it implies that it knows of its absence. Furthermore, a complete logical structure implies that we should be able to observe in any animal the demonstration of transitive inference¹, in particular if the structure of knowledge is represented in the form of conditional propositions.

It is probably important here to specify what is transitive inference. Transitive inference is performed on an ordered set, where, if a relation \mathfrak{R} exist between two consecutive elements of the ordered set (E_n and E_{n+1}) the same relation \mathfrak{R} exist between any E_n and E_{n+x} elements (with $x > 0$)². In consequence, if \mathfrak{R} is a transitive relation and we know that $E_1\mathfrak{R}E_2$ and $E_2\mathfrak{R}E_3$ we can conclude that $E_1\mathfrak{R}E_3$. This conclusion is said to be the product of a transitive inference, an inference based on transitivity, on the fact that \mathfrak{R} is a transitive relationship. So for transitive inference to be possible and experimentally testable we need at minima the following:

¹At least in the case of an “If... Then...” relationship.

²These specifications are at minima. If we consider a reciprocally valid relation \mathfrak{R}' such as “Equal to” where both $E_n\mathfrak{R}'E_{n+x}$ and $E_{n+x}\mathfrak{R}'E_n$ (with $x > 0$) are valid for the given ordered set, the relation \mathfrak{R}' is also transitive.

1. A transitive relationship that is a relation, which is valid between two consecutive element from an ordered set, is also valid between any E_n and any E_{n+x} elements of this set (with $x > 0$).
2. If we want to experimentally create the ordered set:
 - An element E_n of the ordered set always has the same identity.
 - An element E_n of the ordered set always has the same attributes.

Common transitive relationship are “Greater than”, “Bigger than”, “Smaller than”, “Equal to” or “If...Then”. But importantly, relationship such as “Better than” are only transitive if the criterion for “Better” is explicit. For example, let us consider the relationship \mathfrak{R} “Better than”. In the case where $A\mathfrak{R}B$ because A is Yellow and B is Green (the criteria here is that more yellow is better) and $B\mathfrak{R}C$ because B is made of Plastic and C is made of Metal (here, the more artificial the better), we do not have a transitive relationship since the criteria are different for each comparison. Whilst it can still be called “Better than” in each case, we can not conclude that $A\mathfrak{R}C$, *i.e.* that is A is better than C . Thus, it is important when testing transitive inference capacity to make explicit the criteria for comparison, particularly as non verbal task are the norm for non human animal research. Furthermore, the criteria for comparison needs to be perceived by the animal under consideration and therefore part of its Umwelt. This is exemplified for example in the work of Grosenick et al. (2007) and Paz-y Mino et al. (2004).

2.1 Biological Consequences

In a recent paper (Guez, 2009) I put forward a learning theory based on the acquisition of rules in the form of conditional propositions, where the most basic biological expression of this conditional proposition is the link between two neurons after undergoing long term potentiation. This means 3 important things: (1) Learning is the formation of this functional link, (2) the very same links are the engram and (3) these links are part of the animal’s representation of the world.

In fact, we need neuronal connections that are capable of changing quickly to rapidly acquire new information. But if these connections are capable of changing ‘weight’ extremely rapidly we would lose the information acquired when new events are perceived. In consequence, we could envision a layered system. The first layer would be an extremely plastic neuronal network where all synaptic connections can vary in both directions in function of what is observed in the world. With only this layer in place the extinction learned behaviour may be conceived as forgetting. The higher layers are constituted by progressively less plastic networks that learn from previous layers - this stage constitutes the consolidation period. The lesser plastic network represents the mid to long term memory engram.

3 Manipulate

In the definition above, I use the word “manipulate”. Let us first consider the definition without this word. In this case cognition could be purely an automatic process in which an environmental or internal cue would trigger the activation of predictions following the cascading activation of conditional proposition links. The prediction satisfying the current motivational state would be acted upon by the organism (this prediction would be encompassed by the current representation of reality of the organism considered). Although this would not be problematic *per se* it would exclude predictions based on a hypothesis from the field of cognition (see below).

Let us admit that our model organism has learned and therefore integrated into its representation of reality that if an event *A* (e.g. a sound) occurs then an event *B* will occur. The organism needs *B* to occur (e.g. food or water). Suppose that this organism knows how to cause a variety of sounds by performing a nose poke, and that each specific hole is associated with one particular sound but none of them are *A*. When presented with new holes in addition to the previous ones, if the animal preferentially begins to nose poke holes that have not previously been used it would be an illustration of hypothesis testing and therefore manipulation of its reality representation. It is important to note that the hypothesis can only be a construct that uses what is known and cannot make use of what is not part of the reality representation considered. (For a discussion of what is the known, the unknown and the not known see D. Guez (2009) and R. Guez (2000).)

4 Decision making

The last important concept to address in relation to this definition of cognition is decision making (to act upon it). I believe that a simple yet powerful way to model decision making is to imagine a hierarchical system of threshold driven rules. This would allow one to explain how goal directed behaviour arises. Goal directed behaviour is the consequence of a rule passing its threshold. It may then trigger a stereotypic behaviour that has been selected for its ability to maximise the fulfilment of the rule and therefore the organism’s survival and reproductive success.

In this representation of decision making it is proposed that if more than one rule passes its threshold it is the one of higher weight that gains priority, regardless of the state of the lower one. Furthermore, such a hierarchical system of threshold driven rules can only be prewired during development, at least for the highest weight ones because they govern decision making, and are necessary for the survival of an organism. For the same reason they are less likely to be modified by open program (*i.e.* learning) at least for the highest ranking rules. It is therefore most probable that only the lower part of the rule hierarchy is modifiable by open programs. Furthermore, the higher the rule in this hierarchy the more likely the resulting behavioural pattern is to be stereotypic, in order to maximise the chances of survival and reproductive success of the organism considered.

5 Consciousness

So what is consciousness, given the previous definition of cognition? Consciousness would be characterised by the capacity to manipulate the representation of reality by posing hypotheses and using such hypotheses to generate new predictions. This leads to the following definition of consciousness:

*The capacity to construct and **manipulate** one's predictive representation of reality and act upon it.*

It is most likely that the demonstration of manipulation of a representation can only be obtained within the open program domain of an organism. Interestingly, this definition is inclusive of the concept of theory of mind and therefore self consciousness and is akin to the 'access consciousness' of Block (1995), but excludes the 'phenomenal consciousness' that is unfalsifiable and therefore outside the realms of science. Although it is quite easy to demonstrate whether a given organism perceives a physical dimension or not, it is impossible to scientifically investigate "what it is like" for this organism to perceive this dimension. Even if we consider only our own species there is no way to ascertain if the red a person perceive is the same as that perceived by another; all things being equal, it may be what I perceive as green or purple and there is no experimental design that will be able to test this with complete certainty. Thus 'phenomenal consciousness' is an unfalsifiable concept that therefore should be excluded from scientific enquiry (see Popper, 1963, for a complete argument).

6 Conclusion

The definitions of cognition and consciousness provided in this paper hinge upon an organisms capacity to construct and manipulate a predictive representation of reality and to act upon it. The proposed way by which an organism constructs a representation of reality in this paper and developed further in Guez (2009) implies that an organism with a nervous system should possess the concept of absence (which is built upon the concept of presence) and be able to preform transitive inference due to the underlying logical structure of their representation of reality. These operational definitions of cognition and consciousness provide an integrated framework upon which further comparative research on cognition and consciousness can be conducted which is not provided by the prevailing *status quo*. It is suggested that proposing and utilising such rigorous definitions will greatly improve our current abilities to falsify hypothesis and propose global theories of cognition and consciousness, which in turn will assist in our ability to further these areas of research.

References

- Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18:227–287.
- Dickinson, A. and Balleine, B. W. (2000). Causal cognition and goal-directed action. In Heyes, C. and Huber, L., editors, *The Evolution of Cognition*, pages 185–204. The MIT Press, Cambridge, Massachusetts.
- Durlach, P. D. and Rescorla, R. A. (1980). Potentiation rather than overshadowing in flavor aversion learning: An analysis in terms of within-compound associations. *Journal of experimental psychology: Animal Behavior Processes*, 6:175–187.
- Grosenick, L., Clement, T. S., and Fernald, R. D. (2007). Fish can infer social rank by observation alone. *Nature*, 445:429–432.
- Guez, D. (2009). A bio-logical theory of animal learning. *Biological Theory*, 4(2):148–158.
- Guez, R. (2000). Les traces et l’art en question. In Berthet, D., editor, *Approche du rôle des traces au niveau plastique dans l’oeuvre visuelle*, pages 57–79. L’harmattan.
- Haselgrove, M. (2010). Reasoning rats or associative animals? a common-element analysis of the effects of additive and sub-additive pre-training on blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 36:[In Press].
- Heyes, C. (2000). Evolutionary psychology in the round. In Heyes, C. and Huber, L., editors, *The Evolution of Cognition*, pages 3–22. The MIT Press, Cambridge, Massachusetts.
- Mayr, E. (1974). Behavior programs and evolutionary strategies. *American scientist*, 62(6):650–659.
- Paz-y Mino, C. G., Bond, A. B., Kamil, A. C., and Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature*, 430:778–791.
- Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge and Keagan Paul.
- Shettleworth, S. (2000). Modularity and evolution of cognition. In Heyes, C. and Huber, L., editors, *Evolution of Cognition*, pages 43–60. The MIT Press, Cambridge, Massachusetts.
- Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford University Press, New York.
- Von Uexküll, J. (1965). *Mondes animaux et monde humain*. Gonthier.

Zentall, R. T. (1993). *Animal cognition: An approach to the study of animal behavior*.
In Zentall, R. T., editor, *Animal Cognition: A Tribute to Donald A. Riley*. Lawrence
Erlbaum Associates, Hillsdale, New Jersey.

Acknowledgement: I would like to thank Aung Si, C. Conway, S. Forêt and R. Guez for
our useful discussions during the elaboration of this paper.