Essay 2

The Causal and Symbolic Explanatory Duality as a Framework for Understanding Vision

Noel Murphy

Dublin City University,
Dublin, Ireland.

Abstract

The conventional approach to interpreting biological vision systems and experimenting with computer vision systems has been overwhelmingly dominated by a representational view of information. Even more recent connectionist approaches, though embodying a substantial change in viewpoint, have only involved a change of the *type* of representation, to one of a distributed nature. An alternative view is the notion of information as being constructed and co-dependent rather than instructional and referential. This is an interpretation based on the more embracing viewpoint of the complementary causal descriptions and symbolic descriptions playing clearly defined interrelated and dual roles, rather than mutually exclusive, or even muddled roles. This paper examines this radical change in perspective and compares it with a causality framework and with a position on the nature of perception which is based on the idea of universals.

2.1 Introduction

This paper is intended as an invitation to discussion. Many of the ideas presented have appeared in one form or another in the work of a number of authors, particularly Varela, Rosen and Watanabe. The primary concern is with perception, and other topics only in so far as they are required for a proper understanding of perception. This paper represents an attempt to come to terms with an appropriate framework, within which to examine and discuss the nature of perception. It is an attempt to distil a consistent approach from some of the wide variety of viewpoints and philosophies that have, and can be applied to this problem, and subsequently to test both the usefulness and comprehension of the approach by introducing ideas which extend it in certain directions. Its function in the context of this workshop on Autopoiesis and Perception is more of a tutorial than an expositional nature.

Particular attention is paid here to the two key notions of information and observation. The term "information" in particular is much abused both in its everyday and in its technical usage and we try to establish an appropriate context for it. The related notion of observation receives less attention, though not much more clarity. It is argued here however, to be pivotal to the beginnings of an explanation of perception. Because the framework used here involves a philosophical position different from the dominant scientific tradition in the study of biological organisation and the nature of intelligence, care is taken to establish the context of our philosophical position and even to be clear about the process and role of explanation itself. The motivation for this work arose from a control engineering requirement to develop artificial vision systems which could be useful in relatively unconstrained robotic environments. Our interest in biological systems is in support of and subordinate to this aim.

2.2 Epistemological Background

The scientific effort to come to terms with the nature of intelligent perception, thought and behaviour is usually labelled Cognitive Science and according to Varela (1992) there have been four major stages in its development over the last forty years: cybernetics, cognitivism, connectionism and enaction. The original programme, which was a wide-ranging cross-disciplinary effort to create a "science of mind", was called cybernetics. It achieved many far-reaching results including the application of mathematical logic to the study of the brain, the invention of computers, systems theory, control theory and information theory, and the demonstration of possibilities for self-organisation.

2.2.1 Cognitivitism

The heir-apparent to these early advances goes under many general titles, including cognitivism, computationalism, Artificial Intelligence, GOFAI[1], and so on. It has in turn, many sub-areas like expert systems, robotics, computer vision, speech recognition, etc., specialising in particular types of problem domain or sensory modalities. The methodology is generally of a *top-down* nature, and is used in both analysis (cognitive psychology, computational neuroscience) and synthesis (artificial intelligence). The central ethos of this approach is that cognition is *defined* as rule-based manipulation (computation) on symbolic representations, where the meaning of each symbol is made to correspond to an external item in a restricted well-defined domain. Information in this context, is considered as an objective quantity associated with objects and properties in the world. It can be detected, processed, and used to build representations of the way the world *is*, external to the organism or system.

[1] *Good Old Fashioned Artificial Intelligence.*

2.2.2 Connectionism

While the connectionist or emergence approach has its origins in the early work on cybernetics, for various well-documented reasons it has only recently developed a level of adherence sufficient to allow it to challenge and complement the dominant cognitivist position. The methodology in this case is usually *bottom-up*, and is characterised by distributed processing using simple sub-symbolic components, and by self-organization leading to global system coherence. The self-organisation is typically realised in terms of adaptive connections (between nodes) which, affected by "experience", change the strength of these connections according to certain rules (e.g. the Hebb rule and its variants, or error back-propagation). In terms of synthesis its successes have been primarily with lower level cognitive capabilities which cause most difficulties for the cognitive approach, such as recognition, association, and memory. It may be possible to integrate the cognitivist and connectionist positions by embedding symbolic levels of description in an underlying distributed system, though only limited effort seems to have been put into this problem so far.

Much of the emphasis and success within the connectionist community to date has been on the distributed and bottom-up aspect of connectionist models associated with the so-called PDP ideas (Hinton 1985, Rummelhart & McClelland 1986). The basic epistemological position is still representational however, though the form and construction of the representations is quite different from the cognitivist approach (Boden 1988, p. 252). In this case it is a global state or performance of the system which is related to meaning in some chosen domain, rather than the value of a localized symbol. Nevertheless, there is still an observer external to both the system and its sphere of operation, and this observer provides the connection between performance and meaning. That is, there is always a teacher to supervise the learning phase of the network model, and the model comes to reflect more or less accurately and successfully some of the cognitive concepts of the teacher. Even the measurement of accuracy and success are dependent in the final analysis on the teacher.

The objectivist position implicit in both the GOFAI and PDP traditions centres around the commonsense idea that the world as we experience it is independent of the knower. The problem of perception is then to find algorithms or mechanisms which will allow this absolute reality to be captured. Knowing is the act of "duplicating" what is already there outside the knower, using the senses to convey information to construct the appropriate representations. What is represented is a correspondence between symbolic units in one structure (the representation), and symbolic units in another structure (our world or frame description). But, as Varela clearly points out, the problem with the representational approach is that there is no way within the system supposed to construct these representations, of ever obtaining the appropriate assignment of correspondence. There is no independent access to the supposed external reality. The primary reason for this problem seems to be as a result of confusion between different levels of explanation. It is the confusing of notions proper to the domain of an observer (or strictly an observer community) whose vantage includes both the system and its interactions with its environment, on the one hand, with notions proper to the operation of the system, on the other. These are different phenomenal levels. Links, if any, between these levels, can only be established by someone external to both the system and its environment.

More recently however, the connectionist approach has been instrumental in forcing a reinterpretation of the role of individual neurons in the analysis of biological neural systems. This reinterpretation moves away from the information processing and representation role exemplified by, for example, the theories of hierarchial visual processing of Hubel & Wiesel (1977), or Barlow's "grandmother cell" (Barlow 1972). In its new role, the neuron is seen as belonging to large transient ensembles of coherently active neurons, where no single neuron is responsible for, or even restricted to, a single aspect of perceptual experience (Grey *et al.* 1989, Eckhorn *et al.* 1988, Freeman 1991). What is more, the reafferent neural projections from higher cortical areas to the early sensory cortex, which far outnumber the afferent projections from the sensory organs to this sensory cortex, having being mostly ignored in theories of cortical information processing heretofore, are now being recognised for the role they play in the emergence of global cortical phenomena. This alternative approach to connectionism is very different in its philosophy from either the GOFAI or PDP concepts. It emphasises the *self*-organising properties of connectionist models rather than their representational possibilities. Adaptation—if it can be called that—takes place without the benefit of supervision: the activity of the system is determined by the structure of the system itself. This approach is already pointing in the direction of Varela's fourth category, enaction. However, the enactive perspective is concerned with broader issues than just the properties of particular models. It seeks to revise the very roots of our epistemological stance, not eliminating the notion of representation but making clear its restriction to well-defined situations, described *a priori* by an external agent or observer.

2.2.3 Enaction

The single most important assumption, both within AI, and more generally, of the dominant cognitive scientific tradition, is "that the world as we experience it is independent of the knower" (Varela 1992). The primary task of the perceptual part of a cognitive system is thus to capture an accurate representation of aspects of this world. Even though this tradition found inspiration from the classical Newtonian view of physics, it has remained despite the radical overthrow of the Newtonian world view and its conception of ontological reality. The opposite extreme to the view that the nervous system objectively maps *the* external world, is the notion of solipsism—that what is perceived depends solely on the structure of the organism itself. The fact that cognitive phenomena cannot be understood in terms of a world that "informs" us, because there is no mechanism that makes this informing process possible, makes the non-objective extreme no less unpalatable (Maturana & Varela 1987). Fortunately there is an acceptable view, intermediate between these two extremes, which has been articulated extensively in the work of Maturana and Varela. Basically this view is that knower and known arise in a process of mutual specification. Neither the structure of the world as it is perceived by an organism, nor the operation of the observing organism, are pre-given. They are co-determined by a history of cognitive interaction, neither logically preceding the other but still logically compatible. There are two important issues implicit in this stance: the type of system that can participate in this co-determination of a constructed "reality" and the methodology of explanation which maintains distinct the type of description appropriate to each phenomenal domain. We examine these in turn.

2.3 The Control/Autonomy Duality

The drawing of a distinction between a system and its environment is the most fundamental act of system theory (Varela 1979, p. 84). Where we, as observers, put the emphasis of this distinction largely depends on our perspective or our purpose in making the distinction. If we focus on the internal operation and organisation of the system we are putting its environment into the background and relegating interactions with the environment to the status of perturbations. We are also emphasising that the properties of the system arise from within its own structure (the interactions of its components) with the environmental perturbations possibly triggering but not specifying the ongoing operation of the machine. On the other hand, if we focus on the environment, the system is treated as a simple system with given properties and its interactions with its environment constitute a part of its definition. The natural problem domain arising from this latter view is the control of the *behaviour* of the system by utilising the constraints with its environment. This is essentially the subject matter of control theory. The former case where the system is emphasised is the domain of autonomous systems theory.

Allonomy, literally meaning external law, implies the regulation or control of a system from outside (Varela 1979, p. xi). Interactions between the system and its environment are "instructive" and constitute part of the system's organisation. Unsatisfactory results from these interactions are errors. The organisational paradigm is usually formulated in terms of *input-process-output* and is organisationally open. This view of a system is suitable for the domain of design where an observer specifies by its use, what the environment should be and how the system ought to use it. In other words it involves a representational viewpoint with the observer or designer specifying the appropriate semantic correspondences. *Autonomy* on the other hand literally means self-law, implying the internal determination and regulation of the system's operation. Interactions with the systems are seen as perturbations which are non-instructive and independent of the definition of the systems organisation. Varela uses the metaphor of conversation to describe our interactions with autonomous systems. Unsatisfactory results from these interactions are represented by mis-understanding. The organisational paradigm is one of *circularity* and the system is organisationally closed. Information is considered as constructed and co-dependent, where the outcome of perturbative inputs and outputs reflects structure attributed both to the environment and to the internal operation of the system arising over a history of continued operation of the system (and hence viability in its environment).

In addition to these complementary ways of making the fundamental distinction involved in systems theory, it is important to distinguish between the organisation of a system and its structure of realisation. The precise definition of a machine cannot be in terms of a list of its parts or its potential use or purpose—rather it must be by these, plus a description of the permitted inter-relations of the machine's components. A machine's *organisation* is the set of "relations that define the machine as a unity, and determine the dynamics of interactions and transformations it may undergo as a unity" (Varela 1979, p. 9). There is no connection with materiality in the definition of a machine's organisation: it does not specify properties of components that allow the realisation of a machine as a particular concrete system. This closely parallels the idea of *relational biology* described in the 1950's by Rashevsky (see e.g. Rosen 1985b, 1985a)[2]. A machine's *structure* on the other hand, is the set of actual relations that hold between the components that realise a particular instance of a machine in a given space, and is determined by the properties of these components. Finally, the *use* to which a machine is put is not a feature of the organisation or even directly the structure of the machine, but rather the domain in which the machine operates. That is, it belongs to our description of the machine in a context wider than the machine itself—the domain of observation (or design). This clarification leads us directly to the next topic.

[2]According to Rashevsky, as described in Rosen (1985b, p. 172), "... we are interested in the organizational features common to all living systems; and in their material structures only in so far as they support or manifest these features. Therefore we have heretofore approached organisms in precisely the wrong way; we have abstracted out, or thrown away, all those global organizational features in which we are really interested, leaving ourselves with a pure material system that we have studied by purely material methods, hoping ultimately to recapture the organization from our material studies ... Why do we not, in effect, *abstract away the physics and the chemistry*, leaving us with a pure organization which we can formalize and study in completely general abstract terms; and recapture the *physics* later through a process of *realization*." It is important to make this distinction between organisation and realisation because the physics (including the molecular biology) involved in the realisation of real organisms is logically compatible with the organisation of an organism or biological system but not logically prior to it.

2.4 Descriptions and Explanations

In his 1979 book *Principles of Biological Autonomy* Varela sets out to lay bare the relationships between "a system's *identity*, its performance in its *interactions* with what it is not, and how we *relate* to these two distinct domains" (Varela 1979, p. xii). Already, embedded in this statement of the issues of concern is a pervading circularity which is the cause of much of the confusion of levels implied in objectivism. This is the case, for implicit in our act of description of a system and its environment are the peculiarities and particularities of the nature of the relationship between ourselves and *our* environment. More explicitly:

> ...the study of autonomy and [a] system's descriptions in general cannot be distinguished from a study of the describer's properties ... the system and observer appear as an inseparable duo.
>
> Varela (1979, p. 63)

By expressing an interest in the nature of perception—often inappropriately considered as generating a description of one's environment—we are immediately embroiling ourselves in these issues.

In spite of the circularity, we have to start somewhere, so we will start with the notion of description, though in the particular role of explanation—our explanations within a scientific community. In this context Varela draws a distinction between symbolic (or communicative) explanations and operational (or causal) explanations. The difference lies in both their form and use. Operational explanations are assumed to be defined in terms proper to the domain in which the systems that generate the phenomena in question operate—for the purposes of prediction and manipulation. Symbolic explanations are assumed "to belong to a more encompassing context in which the observer provides links and nexuses not supposed to operate in the domain in which the system that generate the phenomena operate" (Varela 1979, p. 66)—for the purpose of communicating an understanding between members of our scientific community. The fundamental basis of operational explanations is nomic or law-like relationships—the fundamental basis of symbolic explanations is order or pattern, and it is the observer who establishes the connection. But it is not meant by this that the causes or laws, often so-called "laws of nature", are in some sense superior by being more remote from the observer, more objective. Both types of explanation are

> ... modes of description adopted by enquiring communities for some intensional purpose ... and they specify modes of agreement and thus coupling with the environment.
>
> Varela (1979, p. 77)

The basic argument of autopoiesis is that all biological phenomena can in principle be reduced to a particular type of network of nomic relationships in some material domain. In this operational description notions of purpose, message, information or code play no *causal* role. But this is not the whole story: it may not be desirable or practical or useful to reduce every aspect of biological phenomena to operational descriptions. It may be very useful for *our* purposes to abstract or parenthesize a number of steps in a causal chain, choosing to ignore the operational connections in favour of more convenient descriptions. This is what Varela claims is at the base of all symbolic descriptions: a process of abstraction rooted in the emergence of certain "coherent patterns of behaviour" to which we *choose* to pay attention:

> Information does not exist independent of a context or organisation that generates a cognitive domain, from which *an observer community* can describe certain elements as informational and symbolic. Information, *sensu strictu*, does not exist. (Nor, of course do the 'laws' of nature).
>
> Varela (1979, p. 78)

But using information in a causal or operational role, e.g. relating behavioral regularities (in the domain of interaction between a system and its environment) to structural change (within the system), is a confusion of levels. The behavioral regularities are only available to us as external observers with simultaneous access to the operation of the system and its interactions with its environment. They reflect our operations and they are not operational for the system. The system does not have independent access to the nature of the structure of the environment.

So, what *is* a valid symbolic explanation? Well according to Varela, symbols in natural systems are characterised by two main features: internal determination and composition. Internal determination refers to the claim that an object or event can be considered as playing the role of a symbol,

> ... only if it is a token for an abbreviated nomic chain that occurs *within the bounds of the system's operational closure* ... whenever the system's closure determines certain regularities in the face of internal or external interactions or perturbations, such regularities can be abbreviated as a symbol, usually the initial or terminal element in the nomic chain.
>
> Varela (1979, p. 80)

In addition, symbols syntactically composible to yield valid combinations seen to confer selective value on the organism to which they belong.

2.5 The Causality of Systems

In answer to the question of why an object or artifact is the way it is, Aristotle attributed four different and inequivalent causes—four different ways of saying "because". That is, if we consider an object as the "effect", then its *material cause* is the matter comprising the physical manifestation of the object; its *formal cause* is the shape (form), plan or blueprint for the object; its *efficient cause* is the act of construction or the processes which shaped the object to its present form; its final cause is the reason for, the goal fulfilled by, or the use of the object. However, in addition to their classical usage these causal ideas can be used as a useful framework for understanding, not only objects but also systems. Consider, for example, the following definitions (Bunge 1979): the *material cause* is the passive receptacle on which the remaining causes act[3]; the *formal cause* is the essence, idea or quality of the thing concerned; the *efficient cause* is the external compulsion that bodies have to obey; the *final cause*, for a machine, is its use, aim or purpose. With these more general definitions, it is possible to relate this causal framework to the operational/symbolic distinction made by Varela. The material, formal and efficient causes belong to the operational description of a system—they all involve categories or relations within the phenomenology of the operation of the system. The final cause on the other hand, which can be interpreted in terms of purpose or use, does not pertain to the machine's operation—it is not a feature of its organisation alone. Rather it belongs to our description of the machine in a context wider

[3] It is important to distinguish *materiality* (involving the properties of components that define them as physical entities) and *material cause* as defined, which has very little to do with matter.

than the machine itself—in other words, Varela's symbolic explanation.

Rosen (1985b) is even more explicit about these relationships and uses this causal framework to directly interpret the dynamics of systems. Consider for example the dynamical system description in terms of the rate equations:

$$\frac{dz}{dt} = \Psi_g(z, \beta(t))$$

where:

$z(t)$ is a state (or phenotype) vector

g is a system or species (genome) vector

$\beta(t)$ is a vector of environments (inputs, forcings or controls).

If the "effect" is the state $z(t)$ of the system at a given time t (cf. the notion of phenotype), the *material cause* is the initial state of the parameters in the state space $z(t_0)$; the *formal cause* is the type, form or identity of the system labelled by coordinates in a function space; the *efficient cause* is the operator that transforms the initial state to the current state which depends on the organisation of the system and its environmental inputs, i.e. the operator:

$$\int_{t_0}^{t} \Psi_g(\ldots, \beta(\tau))d\tau$$

The notion of a *final cause* plays no role in this Newtonian-type formulation of a system's dynamics. Rosen's claim is that the Newtonian paradigm only applies to those systems for which the categories of causation can be segregated into mathematically independent structures, and for which there is no category of final causation, as in the example above. This class of systems is referred to by Rosen as *simple systems* and not all systems can be reduced to this Newtonian form.

Consider, for example the rate equations for a general dynamic system, where the environmental controls and genomic labelling are temporarily omitted:

$$\frac{dx_i}{dt} = f_i(x_1, \ldots, x_n)$$

Now consider the quantities:

$$u_{ij}(x_1, \ldots, x_n) = \frac{\partial}{\partial x_j}\left(\frac{dx_i}{dt}\right)$$

If u_{ij} is positive then an increase in x_j implies an increase in the *rate of production* of x_i, i.e. x_j is an *activator* of x_i. Similarly, if u_{ij} is negative, x_j can be described as an *inhibitor* of x_i.

Now, there are many situations where this type of activation-inhibition description is more appropriate than a rate-equation description. However, if we have a description of a system in terms of u_{ij}'s we can only go to a rate- equation formulation if the differential form for each i,

$$\omega_i = \sum_j u_{ij}\, dx_j$$

is an exact differential. For $n > 2$ this differential form is exact only if

$$u_{ijk} = \frac{\partial}{\partial x_k}(u_{ij}) = \frac{\partial}{\partial x_j}(u_{ik})$$

If u_{ijk} is positive, then x_k enhances or potentiates the effect of x_j on x_i and we can call x_k an *agonist* of x_j. Similarly if u_{ijk} is negative we can call x_k an *antagonist* of x_j. For arbitrary systems there is no special reason why the condition for exactness of the differential form should hold. When a differential form cannot be integrated to give an equation involving the x_i's only, it is referred to as a *non-holonomic* constraint. Each such equation of constraint between the x_i's and the dx_i's can be used to eliminate one degree of freedom of velocity but not the corresponding configurational coordinate in the phase space. Because we cannot obtain a rate-equation formulation for systems involving such non-holonomic constraints, we cannot describe the system in terms of separate categories of causation as in the so-called *simple* Newtonian-type formulation described above. Components of the system may play more than one causal role at a given time, as is typical of systems with a circular organisation[4] (Rosen describes such systems as these as *complex* systems). In particular there is no such thing in this case as a set of states which are assignable to the system for once and for all. Also, these non-holonomic constraints are examples of the type of regularities that an external observer might describe as symbolic in Varela's terms. Perhaps this type of situation is characteristic of systems which display non-trivial metadynamical organisation (see, e.g. Bourgine & Varela 1992).

2.6 Universals in Perception

In the classical philosophical problem of the relationship between universals[5] and particulars there

[4]Note that by circular organisation we do not simply intend systems with feedback, as even the simplest systems which can be expressed in terms of rate equations include feedback.

[5]Objects around us share features with other objects. It is in the nature of most such features that they can characterise

were two principal positions (though many shades of opinion within each). In what must be an unlikely use of terminology, *realism* is the view that ideas, forms or universals are the only true reality, belonging to a world beyond matter and appearance. The world of appearance has only a temporary, illusory existence—the human mind can only apprehend the particular by virtue of it being able to apprehend universals (the notion of universals prior to the objects). On the other hand one of the common denominators of *anti-realist* views is that the human mind can directly apprehend the particular (Watanabe 1985).

Plato is generally acknowledged to have been the founder of the realist view. He treated universals as "objects" (forms, ideas) separate from their instances[6] and independent of human understanding. It is possible (Watanabe 1985, p. 93) that much of the subsequent criticism of Plato's Forms arises because of the later Aristotelian bias that forced the idea into the role of substance, which Plato did not intend:

> The Form is not a perfect object in the best of worlds but rather the essential nature or functional meaning of the objects covered by the same name.
>
> Watanabe (1985, p. 47)

From the point of view expounded by Plato, particular objects do not have a real existence, only a deceptive, temporary illusory one. The sensory world of experience has no reality, but the eternal world of form has reality. A particular object belongs to a class corresponding to a universal because it "partakes" in the archetype or form corresponding to that universal. Not surprisingly, in modern parlance the term Platonist is usually associated with the reality or truth of abstractions (particular mathematical ones) such as numbers, sets or propositions etc. Let us set aside though for the moment the status in terms of reality, truth or origin, of the forms or ideas playing the role of universal in Plato's theory. The really crucial notion as far as understanding perception is concerned, is the claim that the human mind can only apprehend the particular by virtue of it being able to apprehend universals. The aim of this paper is to examine this claim in the enactive context outlined above and to use it as a starting point in the development of a theory of perception.

Consider for example the question of how an object (particular) is identified, as described by Watanabe:

> It is identified by observation, just as a predicate is confirmed or denied by observation … a particular object can only be identified through testing applicability of some general concepts (universals), which in our context (pattern recognition), amounts to observation of some predicates. If we agree that an object can be identifiable only by a group of observations, the object-predicate relation is no more than a relation between two groups of observations.
>
> Watanabe (1985, p. 92)

If realism is the view that universals have real existence, then the diametrically opposed view is referred to as *nominalism*. This is the notion that the universal is a name (or word) without any real existence. *Conceptualism*, holding the middle ground between these two extremes, is the view that the universal does not exist in the real world, but has a real existence as an idea in our mind. In the nominalist position there are only general words like 'dog', and no universals in the sense of entities like 'doghood'. The logical (if extreme) conclusion of this viewpoint has to be that there is nothing in common between the particular objects covered by the same general name (Watanabe 1985, p. 52).

For conceptualism, universals are thoughts or ideas in, and constructed by, the mind. That is, universals are concepts in the mind of those who understand the general word whose meaning the universal is. These ideas came to the fore in the seventeenth and eighteenth centuries and are largely associated with empiricists such as Locke, Berkeley and Hume. Locke, writing in his "Essay Concerning Human Understanding" (quoted in Watanabe 1985, p. 52) claims:

> General and universal belong, not to the real existence of things, but are inventions of the understanding, made by it for its own use, and concern only signs, whether words or ideas … Words are general when used for signs of general ideas and so are applicable indifferently to many particular things; and ideas are general when they are

indefinitely many objects and because of this these features are called universals. The problem is to describe their status. See Lacy (1986).

[6] Aristotle is also credited with holding the realist viewpoint although he denied that universals are objects or separate from their instances, instead claiming that they are real things which exist just *by* being instantiated—the notion of universals in the objects.

> set up as the representatives of many particular things. But universality belongs, not to things themselves, which are all of them particular in their existence ... [when] we quit particulars the generals that rest are only creatures of our own making, their general nature being nothing but the capacity they are put into the understanding, of signifying or representing many particulars.

There are a number of points of interest to us here. The most straightforward one is that universals or concepts or generalisations are constructs of the mind which are found to be useful and can be represented as, or are equivalent to symbols (cf. Locke's use of the term 'signs'). Secondly, he pointed out the role of general idea as an abstraction, leaving out all the particular ideas of individual particular objects that are not common between the objects. It is interesting to note that this is exactly the same sentiment as underlies the modern notion of abstraction used in pattern recognition.

There is one point with which we would like to take issue i.e. the notion of the real existence of particular objects. This, we argue, is itself a construct of the mind, for an object is not perceived independently of observation. Perception is based solely on primitive observations or measurements of predicates and the relationships between these measurements, and as such, consists only of the satisfaction of generalisations (concepts or universals). Popper makes what is essentially the same point, claiming that association psychology—the psychology of Locke, Berkeley and Hume—was merely a translation of Aristotelian subject-predicate logic into psychological terms:

> Aristotelian logic deals with statements like 'Men are mortal'. Here are two 'terms' and a 'copula' which couples or associates them. Translate this into psychological terms and you will say that thinking consists in having the 'ideas' of man and mortality 'associated'.
>
> Popper (1976, p. 76)

The reality of particular objects is something that can only be inferred on the basis of relationships between observations and cannot be used as the basis for perception itself. So while some of what Locke is saying seems plausible we must disagree on the most fundamental point about the direct perception of the particular.

Most modern pattern recognition is based on some form of *similarity theory*—the commonsense view of classes as a collection of particular objects. That is,

- what really exist are particulars, not universals,
- the particular objects in a class are bound together by *similarity*.

Watanabe describes several objections to this position, related to the necessity to include universals and the arbitrary nature of the notion of similarity, but he also derives a result which in fact removes the entire foundation of similarity theory. He shows that by a logical and empirical extension of the similarity theory position (to what he refers to as a *radical nominalism*) all objects must be indistinguishable—equally similar and equally different. He refers to this result as the *Theorem of The Ugly Duckling*. The reason for this result is that in a logical treatment all predicates must be treated equally. The way to defeat it is to make some predicates more "important" in some sense than others.

> To be similar may be to share more of the important predicates. But, how can we evaluate the scale of importance? To answer this question, we have to reflect on the reason why we use similarity and classification in life. The answer is because it is useful, In other words, our scale of importance must be such that the resulting classifications carry utility or value. We can overcome the radical nominalism only by axiological considerations. I do not hereby mean any ultimate value, but various instrumental values towards more fundamental ends.
>
> Watanabe (1985, p. 84)

Watanabe considers that mathematically each predicate must be assigned a different weight (what he calls a preferential ponderation) that varies depending on the use that is going to be made of the resulting classification. He uses an "entropy"-type function to measure the distribution of these predicate weightings concluding that we can only see similarity, and hence grouping of objects if there is an "uneven" emphasis on the empirical data about objects:

> ... epistemology can subsist only through its interaction with axiology. It will be deprived of its major function, concept formation, if it relies only observational experience and logical manipulation. What

> makes cognition possible is the evaluative ponderation, whose origin is aesthetic and emotional in the broadest sense of the term.
>
> Watanabe (1985, p. 88)

In terms of the nature of perception Watanabe goes much further than the need just described, to evaluate predicates of objects in terms of their usefulness. The object-predicate table is a mathematical expression of the Aristotelian idea that the world consists of a discrete number of self-identical objects, subject to a discrete number of attributes. On the basis of the realist position, particularly that associated above with Plato, that the human mind can only apprehend particulars by virtue of it being able to apprehend universals, Watanabe suggests that the notion of the object-predicate table should be inverted. Recall the "it is identified by observation ..." quote above.[7] Consider now the following extract from C.S. Peirce:

> I have maintained since 1867 that there is one primary and fundamental logical relation, that is illation ... A proposition, for me, is but an argument divested of the assertiveness of its premise and conclusion. That makes every proposition a conditional proposition at bottom ... This is the very same relation that we express when we say that 'every man is mortal,' or 'men are exclusively mortal.' For this is to say, 'Take anything whatever, M, then if M is a man it follows necessarily that M is mortal.'
>
> Pierce (1960)[8]

A proposition $P(a)$ that an object a satisfies a predicate P means, according to the usual Aristotelian interpretation of pattern recognition, that object a is placed in the class corresponding to P, (the class of all elements for which P is true). The interpretation of $P(a)$ suggested by Peirce above is that if X satisfies A, which is the predicate or property of being a (i.e. A-ness), then X satisfies P. In other words the Aristotelian logical formula $P(a)$ becomes an implication between predicates: $A \to P$; i.e. the relation which underlies all logic is *implication*, and the basis which underlies all perception is the observation of predicates. Watanabe goes on to show that the logical formalism or algebra derivable from implication is non-distributive, in contrast to the distributive nature of ordinary Boolean logic and set theory.

In human pattern recognition there seldom is a definite (yes/no) implication. We can seldom say that an implication is definitely true or definitely false. Usually there is some sort of "a graduated evaluation of the veracity of an implication", i.e. there is a probability associated with whether or not the implication holds. This is of the form of a conditional probability: $p(P|A)$ which assigns a measure to the probability of P being true or applicable given that A is known to be true or applicable.

> In ordinary thinking, a vaguely conceived association between cause and effect with a graduated degree of certainty is generated first in mind, and in rare occasions it is crystallized as an infallible implicational law. The logical axioms can be considered as a formalization of such exceptional cases of singular associations.
>
> Watanabe (1985, p. 520)

Recall again that our starting point is the implication $A(x) \to P(x)$, and we want to introduce some way of dealing with "a graduated evaluation of the veracity" of the implication. According to Watanabe the most natural way of doing this is to allow a continuous range for the truth value of $A(x)$ or $P(x)$, which usually have one of the dichotomous values 0 or 1. To represent this he introduces a function $f(A, x)$ (for A say), such that $0 \leq f(A, x) \leq 1$, where as usual the value 1 means that the object x definitely satisfies the predicate A (is in class A), and the value 0 means that the object x definitely does not satisfy the predicate A (is outside class A). The class A can be understood as the extension of the predicate A (the set of all objects that satisfy A). When f is limited to the values 0 and 1, we get the usual Boolean logic out of the formalism. This is equivalent to the assumption that at any instant each predicate corresponds one-to-one to a well-defined, fixed set of objects that satisfy the predicate (i.e. a fixed extension). This is an assumption which Watanabe calls "the postulate of definite (or fixed) truth set" and which he attributes to Frege with the name "the Frege Principle" (Watanabe 1985, p. 521).

[7] Watanabe compares this subversion of substance implicit in the object-predicate inversion with the negation of substance (anatman) in Buddhism. A similar shift from substance to function is noted in the introduction of the quantum theory of elementary particles where the self-identity of elementary must be relinqished. Instead of the Aristotelian description: "a particle P is in quantum state Q" we must have "quantum state Q is occupied by a certain number N of particles of a certain kind" for it is quantum states and not individual particles that are distinguishable.

[8] As quoted in (Watanabe 1985, p. 510).

This introduction of $f(A,x)$ as a graduated measure of the certainty of an object x being a member of a class defined by the predicate A, arose out of the fundamental role assigned to implication by Pierce, and the extension of this to ideas about the vagueness of human thinking by Watanabe. Having defined the f-function there are two interpretations that can be assigned to it, depending on the interpretation of implication. In the first case f is called a 'propensity' function. Here we assume that we have an empirical method of determining whether or not the object x satisfies the predicate, with the f-function expressing the "degree of expectation of obtaining the positive empirical result in the A-ness test". But, the critical point is that after the observation is made, the result is either definitely true or definitely false. After the observation any uncertainty about the membership of x in the A-set or class is unambiguously removed. In the second interpretation of the formalism, f is called a 'fuzzy' function or 'membership' function. In this interpretation there is no empirical method or test that can affect the values of the f-function. "This function expresses a purely subjective evaluation of A-ness of object x" (Watanabe 1985, p. 521). We are not concerned further with this fuzzy set theory here. The propensity theory, however, is extremely interesting from the point of view of perception. Using an interpretation of the process of perception introduced by Wilson & Knutsson (1988) we are able to explain several interesting aspects of perception using the propensity theory.

Three assumptions underlie the propensity theory (Watanabe 1985, p. 521 *ff*):

(i) An observational method called an A-ness test or A-test can be defined to determine whether or not an object x satisfies predicate A.

(ii) The observer has a degree of expectation, $f(A,x)$ of getting an affirmative result in the A-test of x.

(iii) The result of two consecutive tests, an A-test and a B-test may depend on the order of the two observations.

The fact that the result of two tests depended on the order in which they were carried out could be because (a) the observed object is changed due to the observation or (b) the observer changes as a result of an observation on an object or (c) both observer and observed change as a result of an observation. The quantal nature of microscopic physical systems seems to arise from reason (a). The effect of the measuring apparatus on the physical system being measured (and therefore interacting with the measuring apparatus) has been discovered to be finite. This means that the effect of measurement on the measured system cannot be ignored, which was one of the basic assumptions of classical mechanics. Watanabe, discussing the fact that information loss is an inevitable consequence of observation or measurement, indicates that the term measurement is somewhat of a misnomer—the actual process is something more akin to preparation. Our knowledge about the system before the act of 'observation' is entirely probabilistic and random. Our knowledge about the system after the act of 'observation' is that the system is in a definite state which can be represented by the eigenvector (of the measurement operator) whose corresponding eigenvalue was the outcome of the act of 'observation'. A possible example of the need for a propensity theory on the basis of reason (b), is the psychology of medical diagnosis:

> . . . when A and B are very close or similar to each other, the ordinary human doctor will tend to classify a patient with a higher probability into A when A is considered before B than when B is considered before A.
>
> Watanabe (1985, p. 522)

2.7 Conclusions

The aim of this paper has been to explore possible frameworks for the understanding of perception. The enactive approach proposed by Varela and the closely related but less elaborated causal approach described by Rosen seem to be the correct overall framework to use for the understanding of biological systems and implicitly the phenomenon of perception. On the other hand, the ideas of Watanabe which are based on the realist position usually attributed to Plato seem to account correctly for the fundamental operations which are at the base of perceptual observations. It is not determined here if these ideas are in conflict or if they address different aspects of the same problems, but the issue is proposed as a basis for discussion.