Associative Learning and Behaviour:

An Algebraic Search for Psychological Symmetries

Eduardo Alonso* and Esther Mondragón†

***Department of Computing, City University, Northampton Square,**

London EC1V 0HB, United Kingdom

eduardo@soi.city.ac.uk, tel: +44 020 70404049, fax: +4402070400244

†Department of Psychology, University College London, Gower Street,

London WC1E 6BT, United Kingdom

e.mondragon@ucl.ac.uk, tel: +4402076795365

Abstract. We propose in this paper to use group theory (more generally, abstract algebra) to unify different models of theories of associative learning --as complementary to current psychological, mathematical and computational models of associative learning phenomena and data. The idea is to apply Felix Kein's *Erlangen Program* from Geometry to Psychology and to carry out a comparative study of recent researches in associative learning so as to identify the symmetries of behaviour. This approach, a common practice in Physics and Biology, would help us understand the structure of conditioning as opposed to the study of specific linguistic (either natural or formal) expressions that are inherently incomplete and often contradictory.

Keywords. Associative learning, behaviour, symmetries, groups, groupoids.

1. Introduction

The ability of animals to recognise and adapt to patterns of stimuli and to form causal links in dynamic environments is essential for their survival. Associative learning studies how animals *learn* (that is, how they acquire information) and *behave* (that is, how what has been learned is expressed in their behaviour) and is, therefore, of paramount importance in Psychology. Indeed, models of associative learning have proved to be relevant to human learning both theoretically (judgement of causality and categorization, *e.g.*, Shanks, 1995) and in practice (in such diverse areas as behavioural therapy, drug addiction rehabilitation, or anticipatory nausea in cancer treatment to name just a few).

Of course, associative learning is not the only type of learning. There are learning phenomena such as habituation or sensitization that are traditionally considered as non-associative. Others such as spatial learning, perceptual learning and some forms of social learning seem to admit an associative account but such an interpretation is debatable. Besides, behaviour – not even adaptive behaviour – cannot be reduced to learned behaviour. Some reflexes such sucking in babes or sexual patterns of behaviour are indeed adaptive but not learned (although this is also controversial, see, *e.g*., Dickinson & Balleine, 2002). Finally, it must be stressed the difference between learning, the hypothetical psychological and physical changes in the brain (memory), and performance, the manifestation of such change in behaviour (see, *e.g*., Bouton & Moody, 2004)[i].

All this taken into account, it is universally accepted that *associative learning is at the basis of most learning phenomena and behaviour*.

2. Psychological Models of Associative Learning

The study of associative learning in Psychology has specialised in two sub-fields: Classical (Pavlovian) conditioning focuses on how "mental" representations of stimuli are linked whereas instrumental conditioning deals (mainly) with response-outcome associations. It is agreed though that, at the most general level, their *associative structures* are the same (Hall, 2002). In both procedures, changes in behaviour[ii] are considered the result of an association between two concurrent events and explained in terms of operations of a (conceptual) system that consists of nodes among which links can be formed. Since research in associative learning has predominantly focused on classical conditioning[iii] we will use it as our leading example.

At the risk of over-simplification[iv], we can identify the main trends in classical conditioning according to two dimensions, namely, the mechanisms of the learning process and the way in which the stimuli are represented by the learning system. The former fuels the debate between stimulus-processing theories *vs* "connectionist"[v] models, exemplified in the competitive model of (Rescorla & Wagner, 1972) and the Standard Operating Procedures (SOP) theory (Wagner, 1981) respectively; the latter illustrates the distinction between elemental models (for instance, both Rescorla and Wagner's and SOP) and configural approaches (*e.g.*, Pearce, 1987).

Rescorla and Wagner's model rests on a sum error term. The very idea that all stimuli present in a trial compete for associative strength is at the heart of the model. It is precisely this characterizing feature that differentiates it from earlier models such as (Hull, 1943). This assumption allows the model to explain phenomena such as blocking and conditioned inhibition, that is, phenomena that result from the interaction among different stimuli. Other assumptions of the model are path-independence (*i.e.*, that the associative strength of a stimulus does not depend on its previous learning history), monotonicity (*i.e.*, that learning and behaviour are one and the same thing), that acquisition and extinction are opposite processes, and that the associability of the CS is fixed.

It has been argued, quite rightly, that Rescorla and Wagner made such assumptions not to reflect strong psychological principles but, rather, to express their main discovery (competitiveness among stimuli) in a general, abstract model. It should not come as a surprise, therefore, that many phenomena cannot be accounted for by their model (latent inhibition being, perhaps, the most paradigmatic) and that myriads of extensions and truly innovative variants regarding the underlying psychological processes involved have been proposed (*e.g.*, attentional approaches like Mackintosh, 1975 and Pearce & Hall, 1980). It remains the case however, that Rescorla and Wagner's model is still the most influential theory of associative learning.

SOP, on the other hand, is a broader theoretical framework of stimulus processing and memory. Unlike Rescorla and Wagner's model, SOP is not based on familiar theories of conditioning (although stochastic approaches used in SOP can be traced back to Estes, 1950) but instead borrows ideas from both information-processing theories and connectionism. It is beyond this proposal to give a detailed account of SOP. Suffice it to say that, in SOP, stimuli activate memory nodes for which transitional probabilities are dictated by decay functions (traces); that learning rules separately account for excitatory and inhibitory links depending on the particular level of activation of the stimulus traces; and that behaviour is explicitly dealt with through weighted response-generation rules. Regardless of its merits[vi], it is difficult to assess the explanatory and predictive power of SOP due to its representational and algorithmic complexity.

Both Rescorla and Wagner's model and SOP share the assumption that when two or more stimuli are presented at the same time of conditioning, each element may enter into an association with the reinforcer that follows (an unconditioned stimulus US). In general, such elemental theories further assume that responding in the presence of the compound is determined by the sum of the associative strengths of the constituents. As an alternative, configural theories are based on the assumption that conditioning with a compound results in a unitary representation of the compound entering into a single association with the reinforcer. Responding in the presence of the compound is then determined by its own associative strength, together with any associative strength that generalizes to it from similar compounds that have also taken part in conditioning. Configural models have proved to be particularly useful when studying conditional associations where a stimulus comes to control responding to a conditioned stimulus (CS) in a manner that is independent of its direct association with the US (Honey and Watt, 1998) or forming a configural cue that becomes associated with the US (Wilson and Pearce, 1989, 1990). Contrarily, elemental accounts tend to focus on the modulatory properties of the conditional cue over the CS-US association (Holland, 1983, Bonardi, 1991) or over the US representation (Rescorla, 1985).

Anyhow, it seems that research in associative learning suffers from various problems, namely:

1. There is no model that satisfactorily accounts for all the phenomena under study;

2. Some models, such as Mackintosh's and Pearce and Hall's, predict opposite changes in the associability of a stimulus as a consequence of the very same procedure (for example, in partial reinforcement). Likewise, elemental models predict that when two compounds (AB, CD) are trained their associative strength will be the same that the one observed when novel compounds (AC, BD) are tested. Configural theories predict that the associative strengths of trained and novel compounds will differ;

3. Some models explain phenomena that alternative models cannot but, in turn, fail to explain other phenomena that the second can. For instance, latent inhibition, that the Pearce-Hall model predicts, cannot be explained in Rescorla and Wagner's whereas over-expectation, on the other hand, can be explained by the latter but not by the former. Similarly, configural theories can account for feature discrimination effects but cannot predict summation effects, exactly the opposite of what elemental theories account for;

4. There are phenomena that are still waiting for a model to be dealt with. For example, it is not obvious (at least not without making use of *ad hoc* arguments) how to explain spontaneous recovery.

3. Mathematical Models of Associative Learning

Associative theories of associative learning have been mathematically expressed as quantitative models in the form of (sets of) equations. In the traditional syntactic view of mathematical models, equations are taken as formal models in which variables and their relations explicitly denote the phenomena under study.

In particular, Rescorla and Wagner use a simple difference equation (the well-known delta rule) to express the change in associative strength across discrete trials. Moreover, keeping the learning rates of the stimuli constant makes the relation between the US and the CSs and the relation between the CSs themselves linear (although, of course, the learning rate decreases as learning progresses).

Despite it is obviously true that, in so doing, temporal phenomena are left unexplained, we should keep in mind that Rescorla and Wagner's model is an abstraction of change, and that models of dynamic systems do not need to be dynamic themselves. Hence, recent attempts to express Rescorla and Wagner's model as a set of differential equations can be considered as merely mathematical pastimes (*e.g.*, Yamaguchi, 1999).

On the other hand, continuous (a.k.a. real-time) models like SOP are, at least in theory, useful when it comes to making accurate predictions about inter-stimulus intervals effects[vii]; however, the introduction of such extra parameters does not seem to necessarily contribute to the understanding of the fundamental issues in associative learning.

Finally, Pearce's model just adds a similarity function specified in terms of the proportion of elements that the stimuli share.

All in all, mathematical models of associative learning have so far been used as a means to make calculations through elementary algebra or differential analysis. The problem with adopting this narrow version of mathematical model is that it does not provide us with tools to address the above-mentioned limitations. For example, if the meaning of a mathematical model is in the linguistic expression it takes (that is, if there is a unique isomorphism between phenomena and algorithms) then either (a) we cannot explain how a theory can be expressed in different sets of equations or (b) we will not be sure about the effect the addition of a simple parameter may have. Paraphrasing (Chakravartty, 2001), theories and models can be (are) given linguistic formulations but theories and models should not be identified with such formulations.

4. Computational Models of Associative Learning

In general, the use of computational models of associative learning has followed the connectionist trend and borrowed from computational neuroscience several techniques, mainly Artificial Neural Networks (ANNs, see Volge *et al.*, 2004, and Balkenius & Morén, 1998, for two good surveys). It is claimed that such models are adequate models of associative phenomena for four main reasons:

- Firstly, computational models are considered (material and/or formal) analogue models of associative learning. The underlying reasoning (expressed in, *e.g.*, Dayan & Abbot, 2001, and Enquist & Ghirlanda, 2005) is that (a) ANNs model by analogy natural neural networks and that (b) psychological process, including associative learning, are ultimately embedded in natural neural networks; hence, indirectly, ANNs model associative learning.

  However appealing this line of argumentation may be, it shows serious flaws. It is widely acknowledged that ANNs do not resemble natural neural networks in any fundamental way; besides, there is no strong evidence suggesting that electrical or chemical neural activity and associative learning are related. That a version of Dirac's rule can be taken as a model both of neural plasticity and long-term potentiation effects (the Hebbian rule, Hebb, 1949) and of association formation (Rescorla and Wagner's rule) cannot be considered as proof of any common underlying structure and should not be used to reduce psychological phenomena to their alleged neural substratum[viii].

Likewise, that Rescorla and Wagner's rule is essentially identical to the Widrow-Hoff rule (Widrow & Hoff, 1960) for training *Adeline* units and that, in turn, such a rule can be seen as a primitive form of the generalised delta rule for backpropagation only tells us that, computationally speaking, associative learning follows an error-correction algorithm. What a computational model does not tell us, however, is which underlying psychological processes (attention, motivation, etc.) intervene in associative learning or how the physical characteristics of the units involved (*e.g.*, the salience of the stimuli) affect such processes.

It seems, therefore, that neither (a) nor (b) stands and that, consequently, the computational models proposed on such premises are not adequate models of associative learning.

- A more general argument is that ANNs are connectionist models according to which information is not stored explicitly in symbols and rules but rather in the weights (strengths) of the connections; learning would consist of changes in such weights. It is claimed, rightly, that these are precisely the assumptions associative learning models are based upon and, hence, wrongly, that ANNs are a natural candidate to model associative learning.

This quite straightforward argument is, in fact, a fallacy: As connectionists (at least implementational connectionists) themselves concede the way we represent learning, either as continuous changes of weighted connections or as the result of discrete symbolic processing, is *a matter of convenience* and therefore irrelevant to the study of the structures involved.

So, what makes ANNs the preferred computational model of associative learning or, more generally, the preferred tool in computational neuroscience (or, for that matter, in computational physics or computational biology)? Massive parallelism. Not surprisingly, the New Connectionism landmark paper introduced the Parallel Distributed Processing paradigm in cognition (Rumelhart *et al*., 1986). And, in which applications the use of massive parallelism is an asset? In applications where traditional techniques that can be verified and validated fail, in other words, when we have lots of unstructured data and/or when the algorithms are intractable or unknown. Now we are in a position to introduce the two remaining reasons for using ANNs as computational models of associative learning.

- The third argument is that ANNs are powerful tools in finding patterns and models of data. After all, ANNs are just a set of statistical methods (with a misleading name). Although they are certainly not the simplest, fastest or most efficient data mining techniques (see, *e.g*., Mitchie *et al*., 1994), ANNs have proved useful when standard methods fail and a bottom-up, data-driven approach is needed. It is, however, necessary to point out (a) that associative learning is not so data intensive as other areas like genetics where there is an obvious need for statistical tools (see, for example, Hastings & Palmer, 2003), and (b) that, in any case, computational models may lead to an increase in empirical adequacy but not necessarily to a better understanding of psychological phenomena and of their underlying psychological mechanisms.

- This brings us to the fourth argument. ANNs can be used, of course, not as models of phenomena and data but simply to solve problems that cannot be solved analytically – or when *in silico* experiments are needed. Indeed, it is a common practice to use sheer computational power to simulate with state spaces the dynamics of non-linear (chaotic or not) systems such as population growth or the weather. The point is, however, that associative learning does not seem to be one of such systems unless, of course, we view it from a behavioural regulation approach according to which animals adjust their long-term behaviour so as to reach an optimal (bliss or equilibrium) point (Timberlake, 1980).

To sum it up, computational models of associative learning should be taken with caution and, indeed, with caution have been taken by most psychologists[ix]. Surely, they may provide us with complementary idealized models of psychological phenomena and with powerful statistical tools to construct models of psychological data but they are not the appropriate instruments to address the issues referred to in section 2.

Our contention is that what we are lacking in the field of associative learning and behaviour is the identification of invariant structures that underlie specific (psychological, mathematical and computational) models. That is, we need to study psychological symmetries. Crucially, symmetries can be formalized mathematically as operations satisfying the conditions for forming various algebraic structures –typically groups. We propose to employ abstract algebra to explore models of psychological theories from a non-syntactical view (as Physics and Biology have done).

More specifically, our objective is to apply the *Erlangen Program* to Psychology, that is, to carry out a comparative review of recent researches in associative learning according to psychological symmetries.

5. Symmetries

Generally speaking, symmetries define *invariance*, that is, impunity to possible alterations. They refer to the fact that parts of a whole are equivalent (interchangeable) under a group of operations. Interestingly, the fact that the parts that are related by means of an equivalence relation corresponds to the fact that the family of operations transforming the parts into each other while leaving the whole invariant satisfies the conditions for constituting a group (*i.e.*, the existence of the identity and inverse operations, and the closure of the product associativity). Consequently, it has traditionally been assumed that *group theory is the language of symmetries.*

What is more important, in group theory the objects do not need to be mathematical objects or physical, biological or psychological objects. Objects and their elements can be any abstraction (shapes, phrases, laws, mathematical equations and even theories). And the transformations or operations under which the whole remains invariant can be any operation (from a rotation over an axis to a specific conditioning procedure). This is because groups act on operations not on elements or objects. This feature makes groups a powerful tool to study symmetries independently of a particular theory or expression[x].

The study of symmetries flourished in the XIX century, originally as an instrument to solve algebraic equations: It was the young E. Galois who first understood that groups opened a new general way of finding the (invariant) structures that underlie the number and form of the solutions for equations of arbitrary degrees. This had an immediate effect in Physics: C. G. Jacobi developed a procedure for transforming step by step the Hamiltonian formulation of the dynamical equations of mechanics into new ones that are simpler but perfectly equivalent. In geometry, F. Klein (Klein, 1872) proposed the *Erlangen Program* to classify various geometries (Euclidean, affine, and projective) with respect to geometrical properties that are left invariant under rotations and reflections. It was also in Göttingen where E. Noether proved the connection between symmetries and conservation laws[xi] (Noether, 1918).

In fact, we can view the history of the theories of modern Physics in terms of their symmetries and groups. Newtonian classical mechanics was based on Galilei's transformations formalized in the Galilei group; Einstein and Poincaré's special theory of relativity unified seemingly contradictory mechanical and electromagnetic phenomena of the hand of Lorentz's transformations and Lorentz's groups; finally, Einstein and Hilbert's general theory of relativity explained gravity, the most symmetrical of field theories so far, under the group of all diffeomorphisms of a space-time.

It has been, however, with quantum mechanics when symmetry groups have become an indispensable tool in Physics (see Weyl, 1928, for a starter). Internal symmetries (*i.e.*, those which act on fields while at the nuclear level and cannot be reduced to "classical" spatiotemporal symmetries), both global and gauge, can only be fully understood when studied through the groups their representations form. In particular, the Standard Model classifies all elementary particles and their interactions according to their flavor, charge and color symmetries (the $SU(3) \otimes SU(2) \otimes U(1)$ group), and, in so doing, it unifies electromagnetism, QED and QCD and explains electroweak interactions through spontaneous symmetry breaking.

Indeed, Biology has also benefited from the study of symmetries, particularly in molecular biology, physical chemistry and crystallography.

Why is it then that symmetries take such a prominent part in Natural Sciences? As argued in (Brading & Castellani, 2003):

(a) First, we attribute symmetry properties to *theories and laws* (symmetry principles). It is natural to derive the laws of nature and to test their validity by means of the laws of invariance, rather than to derive the laws of invariance from what we believe to be the laws of nature;

(b) At the same time, we may derive specific consequences with regard to particular *phenomena* on the basis of their symmetry properties (symmetry arguments). Pierre Curie (Curie, 1894) postulated a necessary condition for a given phenomenon to happen, namely, that it is compatible with the symmetry conditions established by a principle.

More specifically, symmetries play several inter-related roles in science:

- *Normative role:* One the one hand, symmetries furnish a kind of selection rule. Given an initial situation with a specified symmetry, only certain phenomena are allowed to happen; on the other side, it offers a falsification criterion for (physical) theories: A violation of Curie's principle may indicate that something is wrong with the (physical) description. That is, symmetries can be viewed as normative tools, as constraints on theories –the requirement of invariance with respect to a transformation group imposes several restrictions on the form the theory may take, limiting the types of quantities that may appear in the theory as well as the form of its fundamental equations;

- *Unification role:* Symmetries can be used as a heuristic to compare and unify theories, resulting from the possibility of unifying different types of symmetries by means of a unification of the corresponding transformation groups. Likewise, we can use symmetries to analyse whether or not different theories are, in fact, equivalent; and even if theories turn out to be incomparable (it seems, after all, that Rescorla and Wagner's model and SOP correspond to different algebraic structures –Rescorla and Wagner's model to groups, SOP to Lie groups), we will at least have a tool to formally show they are so;

- *Classificatory role:* Classifications can be used (apart from collecting stamps) to identify gaps in the theories but also to predict the existence of new phenomena –as were quarks, predicted by M. Gell-Mann as a consequence of symmetries in matter particles. This applies when new phenomena can be predicted exclusively in terms of symmetry and when the predictions so postulated are coherent with those of existing models;

- *Explanatory role:* Symmetries are also explanatory in that phenomena can be explained as consequences of symmetry arguments. We know that the symmetry elements of the causes must be found in their effects and that the converse is not true. That is, the effects can be (and often are) more symmetric than their causes. In grouptheoretic terms this means that the initial symmetry conditions are lowered into (more constrained) sub-groups: The symmetry has been broken.

Of course, organising our knowledge using symmetries does not prove anything. Symmetries (and group theory) provide us with very powerful abstract tools to analyse the structure of psychological models. But they are just abstract tools after all. In any empirical science, the ultimate proof rests on experimental evidence. Nonetheless, perhaps paradoxically, here it is precisely where the full strength of symmetries shows: Not from the models of theories built on symmetry principles but from the intimate connection (through symmetry arguments) between such models and observed phenomena.

If we look back to the problems faced by psychological models of associative learning as listed in section 2, we find that they relate to deficiencies that the symmetry roles could be use to resolve. The first shortcoming, that no model accounts for all associative learning phenomena, refers to a lack of explanatory power in such models; the second one, that contradictory rules explain the same phenomena, claims for a normative approach; the third one, that models are partial, relates to the need for unifying principles where different theories that cover disjoint phenomena find common grounds and are made compatible; and the fourth one, that some phenomena remain unaccounted for, identifies a classification problem. It seems, therefore, that symmetries may be useful in solving such problems. All is left is to do is to find the psychological symmetries.

6. In search of Psychological symmetries

Although there is not a universally accepted "law of learning", all psychological models coincide in assuming that learning takes place when a (relatively permanent) change in behaviour happens as a consequence of some experience. Now, we need to know whether such law establishes *sufficient* symmetry conditions for the occurrence of the observed phenomena –or, in other words, we have to investigate whether the observed phenomena describe *necessary* conditions for the law to hold (invariantly) true. Unfortunately, a glimpse at the literature suggests it does not:

(a) That the sensory and motivational features of the stimuli as well as their novelty and relevance affect learning are well documented facts (Kamin and Schaub, 1963; Pavlov, 1927; Jenkins and Moore, 1973; Randich and LoLordo, 1979; Lubow, 1989; Garcia and Koelling, 1966);

(b) Procedurally, the idea that learning is context-specific is also gaining ground (Bouton, 1993; Bouton and Swartzentruber, 1986; Hall and Mondragón, 1998); also, different results emerge depending on the order in which stimuli are presented during training and on the number (single or compound) and representation (elemental or configural) of the cues themselves (see, *e.g.*, Pearce and Bouton, 2001 for a survey).

This first setback may not challenge our search for psychological symmetries though. It could we argued that, after all, we should expect that the parameters in (a) affected the pace of the learning (accelerating or decelerating the learning process, *i.e.*, strengthening or weakening the links between nodes/stimuli as time goes), defining, in the extreme, explicit symmetry breaks[xii]. Unfortunately, the study of complex phenomena in (b) does not only tell us that the learning rate changes in different experimental conditions. What these results tell us is that the *rules* of learning themselves fluctuate depending on such factors and, consequently, the do not reflect any genuine object of invariance.

Not surprisingly, a mathematical analysis of the above-mentioned issues reveals that each of them violates one of the conditions for group formation: Associativity[xiii]. This is rather worrying since associativity is *the* key condition for symmetry. It tells us that the concatenation of two different operations gives the same result, and *that* gives us much more information and reflects much more structure than the boring commutativity[xiv].

Let's illustrate this point with an example: Both elemental models and configural models of stimulus encoding anticipate that after conditioning is given to two compounds (say, AB and CD) responding to them will be greater than responding to the constituent elements. However, they differ in their expectations for responding to different compounds formed with the same elements (for example, AD and BC); elemental theories expect it to be as large as that to the trained compounds whereas Pearce's configural theory expects some generalization decrement and, as a consequence, responding should be as small as that to the elements. That is, elemental theories assume associative invariance under different compounds; Pearce's theory, on the other hand, assumes invariance under elements *per se* and new configurations. Unfortunately, evidence suggests (see, *e.g.*, Rescorla, 2003) that neither interpretation is complete: In agreement to Pearce's theory, novel compounds perform less than original compounds but, in agreement to elemental theories, novel compounds perform better than their separate elements. Symmetries, therefore, seem elusive.

Should we conclude, on this basis, that there are no symmetries in associative learning? Perhaps, we can try a different approach and investigate this issue through a representative case study, a model that embodies the fundamental laws of associative learning. Few would disagree that Rescorla and Wagner's model is such a model. Now, Rescorla and Wagner's model is based on five basic assumptions (see Miller *et al.*, 1995), namely:

(1) The associative strength of a stimulus depends on the summed associative value of all the CSs present on a given trial;
(2) Excitation and inhibition are represented by opposite signs on a single dimension of associative strength and, consequently, are assumed to be mutually exclusive;
(3) Associability of a given stimulus (α) is constant, that is, associability is not subject to changes as function of experience;
(4) New learning is invariant to any prior associative history (path independence). Past associative status of a cue, *per se*, is assumed to influence neither behaviour nor future changes in associative status;
(5) Differences in behaviour reflect differences in associative strength, that is, there is a monotonically positive relationship between associative value and a relevant response.

A simple analysis of these five premises tells us that only the first one is *a*symmetric. It states that the associative strength of each CS present on a specific trial does not independently gravitate toward the asymptotic value of the US (λ). If it were so, then the associative strength of each CS would be invariant to the presence of other stimuli. This assumption (that is at the heart of cue competition) has proven to be the most innovative feature of Rescorla and Wagner's model.

The rest of assumptions are, in fact, symmetry postulates: Symmetry between excitation and inhibition (2), invariance of associativity to experience (3) and to learning history (4), and symmetry between learning and performance (5). Sadly, countless observations refute in a consistent manner such assumptions[xv]. The important point is that such failures do not come from Rescorla and Wagner's disregard for parametric features. The disproving phenomena do not refer to specific values that the context, time/schedule or stimulus characteristics may take but rather are the result of fundamental assumptions on the structure of conditioning.

We can attribute this unsuccessful search for symmetries in associative learning to alternative causes:

- Is it that the laws of associative learning are simply wrong? This does not seem to be the case. Despite the problems referred to in section 2, psychological models of associative learning have been confirmed experimentally so as not to doubt their general validity. As stated by (Spreat & Spreat, 1982, from Bouton, 2006), "*much like the law of gravity, the laws of learning are always in effect*";

- Or is it that associative learning phenomena (and the theories in which they are modelled) do not show any underlying structure, at least not in the form of symmetries? Again, this is dubious. As we have seen in the previous section, symmetry has proved to be just too powerful a principle in the study of Nature as not to be found in Psychology;

- Or is it that the formalization of symmetries in the notion of group is too constraining and that associative learning shows, to some extend, symmetries that should be expressed with a subtler concept? Is there any notion in abstract algebra that provides us with the required flexibility to represent associative learning phenomena and theories and, at the same time, preserves the properties that have made groups so popular in Physics and Biology? Yes, there is: The notion of groupoid.

Indeed, it seems that groups do not provide us with the right ontology to deal with the type of symmetries that associative learning may show. Each associative learning theory could perhaps be modelled as a unidimensional single-object category, in other words, as a group. Yet, the problem is that groups (and the theories so modelled) are not expressive or flexible enough and, consequently, are prompt to generate inherently limited classifications and contradictory explanations. Besides, as groups are independent from each other and do not form more general structures there seems to be no need for a meta-syntax that would regulate the relations between different theories and, potentially, unify them.

7. Groupoids

We have seen that mathematicians (and physicists) tend to think of the notion of symmetry as being virtually synonymous with the theory of groups (symmetry groups). In fact, though groups are indeed sufficient to characterize homogeneous structures, there are plenty of objects which exhibit what we clearly recognize as symmetry, but which admit few or no nontrivial automorphisms. It turns out that the symmetry, and hence much of the structure, of such objects can be characterized algebraically (and categorically) if we use groupoids and not just groups (see Brown, 1987, and Weinstein, 1996, for two formal introductions to groupoids).

Intuitively, a groupoid should be thought of as a group with many objects, or with many identities. A groupoid with an object is essentially just a group. So the notion of groupoid is an extension of that of group. This apparently innocuous distinction between one-object structures (groups) and many-objects structures (groupoids) is actually crucial. The homomorphisms defined in groups are always automorphisms (homomorphisms of the object to itself). In other words, as groups are one-object categories, *all* morphisms can be composed with all other morphisms. From this, the algebraic conditions for the formation of groups (closure, unique identity, unique total inverse and *total* associativity) follow directly. On the other hand, groupoids, can only compose morphisms (isomorphisms in their case) with the appropriate domains and co-domains. Algebraically, a groupoid is a set with a *partially* defined binary operation (that *is* associative *where* defined) and a total inverse function.

What is important to get from this mathematical mumbo-jumbo is (a) that, in groupoids, associativity is partially defined, allowing us to investigate variable symmetries (symmetry groupoids) and (b) that, in groupoids, isomorphisms are defined over sets of base points (fundamental groupoids), permitting us to study more symmetries. Indeed, groupoids show new structures that do not show at a group level –more specifically, in groupoids, the inverse relation, although total, is defined over paths; besides, groupoids lead to higher dimensional algebras and help us move between n-categories through natural transformations, limits and co-limits.

Summarizing, groupoids present three very useful properties: (1) Partial associativity, (2) path reversibility, and (3) hierarchism. How this relates to our study of symmetries in associative learning behaviour?

(a) To start with, the very idea of associative learning can be nicely expressed as morphisms (associations) defined over objects (stimuli or nodes), that is, as categories;

(b) Building iteratively up categories (3 above) may allow us to gain knowledge about hierarchical processes –associative processes between associative processes (Bonardi, 2001; Mondragón, Bonardi and Hall, 2003)–, in particular, about the role of context and occasion setters (Bouton, 1994);

(c) Also, the isomorphisms that define groupoids (unlike all or nothing equivalence relations that define groups) permit us to introduce partial symmetries (1 above) that may explain results where novel compounds seem to elicit less response than the original trained compounds but more than each separate element;

(d) Finally, the ability to look at intermediate processes (2 above) may be very useful in determining the causes for non-responding: Failure to express or failure to acquire (or to retrieve) information?

More generally, the theory of groupoids does not differ widely in spirit and aims from the theory of groups. The recognition of the utility of groupoids gives gains over the corresponding groups without any consequent loss. Our contention is that the above-described characteristics make groupoids an ideal candidate to fill in the symmetry roles that, we have argued, would help solve the problems outlined in section 2: Groupoids provide us with a multi-object language defined over paths along with rules of variance and rules of transformation with which to study both internal and external symmetries. In other words, the language of groupoids gives us the required expressiveness and flexibility to attack classification and explanation problems; and its syntax would allow us to solve normative and unification problems.

Admittedly, the debate over whether groupoids are useful or unmotivated abstractions is still going on (Corfield, 2003). Nevertheless, since they were introduced by H. Brandt in 1926, groupoids have been used in a wide area of mathematics as well as in theoretical physics[xvi], neurosciences, biodynamics and networks and logic and computer science (*e.g.*, Ramsay & Renault, 1999, and the excellent http://math.ucr.edu/home/baez/).

References

Balkenius, C., & Morén, J. (1998). Computational models of classical conditioning: a comparative study. In Mayer, J.-A. , Roitblat, H. L., Wilson, S. W., & Blumberg, B. (Eds.), *From Animals to Animats 5*. Cambridge, MA: MIT Press.

Bonardi, C. (1991). Blocking of occasion setting in feature-positive discriminations. *The Quarterly Journal of Experimental Psychology B. Comparative and physiological psychology, 43,* 431-448.

Bonardi, C., & Ward-Robinson, J. (2001). Occasion Setters: Specificity to the US and the CS-US Association. *Learning and Motivation, 32,* 349-366.

Bouton, M.E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning, *Psychological Bulletin*, 114, 80-99.

Bouton, M. E., (1994). Context, Ambiguity, and Classical Conditioning. *Current Directions in Psychological Science, 3,* 49-53.

Bouton, M. E. (2006). *Learning and Behavior. A Contemporary Synthesis*. Sunderland, MA: Sinauer Associates.

Bouton, M.E., & Moody, E.W. (2004). Memory processes in classical conditioning. *Neurosciences and Behavioral Reviews*, 28, 663-674.

Bouton, M. E., & Swartzentruber, D. (1986). Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 12,* 333-350.

Brading, K., & Castellani, E. (Eds.) (2003). *Symmetries in Physics: Philosophical Reflections*. Cambridge, England: Cambridge University Press.

Brown, R. (1987). From groups to groupies: a brief survey. *Bulletin of the London Mathematical Society, 19*, 113-134.

Chakravartty, A. (2001). The semantic or model-theoretic view of theories and scientific realism. *Synthese, 127,* 325-345.

Corfield, D. (2003). *Towards a Philosophy of Real Mathematics*. Cambridge, England: Cambridge University Press.

Curie, P. (1894). Sur la symétrie dans les phénomènes physiques. Symétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique, 3rd series, 3,* 393-417.

Dayan, P., & Abbott, L.F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: The MIT Press.

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. In H. Pashler & R. Gallistel (Eds.), *Stevens' Handbook of Experimental Psychology* (Third ed., Vol. 3: Learning, motivation, and emotion, pp. 497-533). New York: John Wiley & Sons.

Enquist, M., & Ghirlanda, S. (2005). *Neural Networks and Animal Behaviour*. Princeton, NJ: Princeton University Press.

Estes, W.K. (1950). Toward a statistical theory of learning. *Psychological Review, 57,* 94-104.

Garcia, J., & Koelling, R.A. (1966). Relation of cue to consequence in avoidance learning, *Psychonomic Science, 4,* 123-124.

Hall, G. (2002). Associative structures in Pavlovian and instrumental conditioning. In H. Pashler, S. Yantis, D. Medin, R. Gallistel an J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology*, Volume 3, pp 1-45. Hoboken, NJ: John Wiley and Sons.

Hall, G., & Mondragón, E. (1998). Contextual control as occasion setting. In N.A. Schmajuk and P.C. Holland (Eds.), *Occasion Setting: Associative Learning and Cognition in Animals*. Washington CD: American Psychological Association.

Hastings, A., & Palmer, M.A. (2003). A bright future for biologists and mathematicians? *Science, 299 (5615),* 2003-2004.

Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Holland, P. C. (1983). Occasion-Setting in Pavlovian Feature Positive Discriminations. In M.L. Commons, R.J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative Analyses of Behavior* (4 ed., pp. 183-206). New York: Ballinger.

Honey, R.C., & Watt, A. (1998). Acquired relational equivalence: implications for the nature of associative structures. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 325-334.

Hull, C.L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. New York: Appleton-Century-Crofts.

Icke, V. (1995). *The Force of Symmetry*. Cambridge: Cambridge University Press.

Jenkins, H.M., & Moore, B.R. (1973). The form of autoshaped response with food or water reinforcers. *Journal of Experimental Analysis of Behavior, 20,* 163-181.

Kamin, L.J., & Schaub, R.E. (1963). Effects of conditioned stimulus intensity on the conditioned emotional response. *Journal of Comparative and Physiological Psychology, 56,* 502-507.

Klein, F. (1872). Vergleichende Betrachtungen über neuere geometrische Forschungen, *Mathematische Annalen, 43* (1893) pp. 63-100 (Also: Gesammelte Abh. Vol. 1, Springer, 1921, pp. 460-497). An English translation by Mellen Haskell appeared in *Bull. N. Y. Math. Soc* 2 (1892-1893): 215--249.

Lederman, L. M., & Hill, C.T. (2004). *Symmetry and the Beautiful Universe.* Amherst, New York: Prometheus Books.

Lubow, R.E. (1989). *Latent Inhibition and Conditioned Attention Theory*. Cambridge, England: Cambridge University Press.

Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychological Review, 82,* 276-298.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117,* 363-386.

Mitchie, D., Spiegelhalter, D.J., & Taylor, C.C., (Eds.) Elder, J.F., IV. (Rev) (1994). Machine Learning, Neural, and Statistical Classification. *Journal of the American Statistical Association, 91,* 436-438.

Mondragón, E., Bonardi, C., & Hall, G. (2003). Negative priming and occasion setting in an appetitive Pavlovian procedure. *Learning and Behavior, 31,* 281-291.

Noether, E.A. (1918). Invariante variationsprobleme, *Nachr. v. d. Ges. d. Wiss. zu Götingen* 1918, 235-257. English translation: M.A. Tavel, Reprinted from "Transport Theory and Statistical Mechanics" 1(3), 183-207 (1971).

Pavlov, I.P. (1927). *Conditioned reflexes*, (G.V. Anrep, trans.), London: Oxford University Press.

Pearce, J.M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94,* 61-7.

Pearce, J.M., & Bouton, M.E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111-139.

Pearce, J.M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532-552.

Ramsay, A., & Renault, J. (Eds.) (1999). *Groupoids in Analysis, Geometry, and Physics*. Contemporary Mathematics 282. Providence, Rhode Island: American Mathematical Society.

Randich, A., & LoLordo, V.M. (1979). Associative and non-associative theories of the UCS preexposure phenomenon: Implications for Pavlovian conditioning. *Psychological Bulletin, 86,* 523-548.

Rescorla, R.A. (1985). Conditioned inhibition and facilitation. In R.R. Miller and N.E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 299-326). Hillsdale, NJ: Erlbaum.

Rescorla, R.A. (2003). Elemental and configural encoding of the conditioned stimulus. *The Quarterly Journal of Experimental Psychology, 56B(2),* 161-176.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rumelhart, D.E., McClelland, J.L. & the PDP Research Group (Eds.) (1986). *Parallel Distributed Processing*, Vol. 1, Foundations. Cambridge, MA: The MIT Press.

Schmajuk, N.A., & DiCarlo, J.J. (1992). Stimulus configuration, occasion setting, and the hippocampus. *Psychological Review, 99,* 268-305.

Shanks, D.R. (1995). *The Psychology of Associative Learning*. Cambridge, England: Cambridge University Press.

Sutton, R.S., & Barto, A.G. (1987). A temporal-difference model of classical conditioning. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Earlbaum.

Timberlake, W. (1980). A molar equilibrium theory of learned performance. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 1-58). New York: Academic Press.

Volge, E.H., Castro, M.E., & Saavedra, M.A. (2004). Quantitative models of Pavlovian conditioning. *Brain Research Bulletin, 63,* 173-202.

Wagner, A.R. (1981). SOP: a model of automatic memory processing in animal behaviour. In N.E. Spear & R.R. Miller (Eds.), *Information Processing in Animals: Memory Mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

Wasserman, E.A., & Miller, R.R. (1997). What's elementary about associative learning? *Annual Review of Psychology, 48,* 573-607.

Weinstein, A. (1996). Groupoids: Unifying internal and external symmetry (a tour through some examples). *Notices of the American Mathematical Society, 43(7),* 744-752.

Weyl, H. (1928). *Gruppentheorie und Quantenmechanik*. Leipzig: S. Hirzel Verlag. English translation: The Theory of Groups and Quantum Mechanics. New York: Dover Publications.

Widrow, G., & Hoff ,M.E. (1960). *Adaptive switching circuits*. Institute of radio engineers, Western Electronic show & convention, Convention record, Part 4, 96-104.

Wilson, P. N., & Pearce, J. M. (1989). A role for stimulus generalization in conditional discrimination learning. *The Quarterly Journal of Experimental Psychology B, 41,* 243-273.

Wilson, P. N., & Pearce, J. M. (1990). Selective transfer of responding in conditional discriminations. *The Quarterly Journal of Experimental Psychology B, 42,* 41-58.

Yamaguchi, M. (1999). New methods for solving Rescorla-Wagner model. *Behaviour Research Methods, Instruments and Computers, 3(4),* 684-688.

[i] After all, behaviour be used to measure learning but it does not constitute learning itself.

[ii] That learned behaviour is expressed as (learned) reflexes (classical conditioning) or operant responses (instrumental conditioning) does not really matter.

[iii] It is worth noting, however, that this fact is entirely due to methodological issues over experimental control: Whereas in classical conditioning stimuli can be precisely manipulated by the experimenter, only the animals can control their own responses in instrumental conditioning.

[iv] We are considering only associative models of associative learning.

[v] Although all associative models are connectionist it has become fashionable to label as connectionists only those that recur to the use of complex networks.

[vi] Significantly, SOP provides us with a mechanism to represent learning about absent stimuli.

[vii] Special mention goes to temporal-difference models (Sutton & Barto, 1987), models that have had a huge impact in machine learning.

[viii] Of course, this is not to say that (computational) neurosciences and their models of associative learning constitute a Byzantine exercise. Quite clearly, any psychological phenomena can be analysed at a biological level, pretty much like any biological phenomena can be analysed at a physical level and like any physical phenomena can be analysed at an informational level – *it for bit* ! It should be clear, however, that different levels of analysis of the same reality are fit for different purposes. Psychological models provide us with the right level of analysis for the study of psychological phenomena, *i.e.*, when we need to explain how complex organisms learn from experience and adapt their behaviour accordingly. The relationships between the different levels should not conceal that (a) the

levels do not form an explanatory hierarchy where the ones at the bottom (allegedly, the neural ones) are more fundamental and, thus, more "real" than the others; and that (b) the models (particularly, the formal models) at the different levels may (and indeed more often than not do) differ. That is, we should keep in mind that (methodological or ontological) reductionism does not apply between the levels.

[ix] Perhaps the only genuine contribution of (multilayer) ANNs to the theory of associative learning is the configural account of occasion setting, as opposed to the more conventional modulatory approach (*e.g.*, Schmajuk & DiCarlo, 1992).

[x] Readers are referred to Icke, 1995, and Lederman & Hill, 2004, for two popular introductions to the world of symmetries and groups.

[xi] Noether's theorem(s) is unsurpassed in its significance and beauty (and a great illustration of Wigner's "unreasonable" effectiveness of Mathematics in the Natural Sciences). Among other things, it tells us that if the Lagrangrian of a system does not change under certain transformations then we can assert that there exist quantities that are preserved: Symmetry under time translation implies conservation of energy; symmetry under spatial translations implies conservation of momentum; and symmetry under rotation implies conservation of angular momentum.

[xii] Unlike spontaneous symmetry breaking that define phase transitions, explicit symmetry breaking results from interactions (noise) that introduce corrections to dynamical equations that are not manifestly invariant under the symmetry group considered.

[xiii] It is easy to see that in all four types of complex phenomena described above, that is, those that vary with the context, the order and number of the stimuli and the way they are represented, identity and inverse conditions do hold. Any conditioning is preserved if nothing

is done; besides, to undo the effects of any training procedure all we have to do is to re-train the animals in the original condition. The identity element and the inverse element are unique and the same.

[xiv] Incidentally, associative learning has shown to be stubbornly non-Abelian: “Associative symmetry” phenomena and the basic distinction between latent inhibition and extinction are just two examples of non-commutativity.

[xv] In defence of Rescorla and Wagner, it must be said that they themselves expressed their doubts about these four assumptions. For instance, it is hard to believe that Rescorla and Wagner really mistook extinction for unlearning or that they were ignorant of silent learning phenomena. It should also be noted that alternative models based on contingencies do not seem to improve the landscape. Although it has been proved that the non-pairings of CS and US influence behaviour as do pairing of CS and US we know that the four inter-event combinations do not contribute equally to the acquired behaviour (*i.e.*, they have equal normative weights but not equal psychological weights, Wasserman & Miller, 1997).

[xvi] It is interesting to note the view of Connes that Heisenberg discovered quantum mechanics by considering the groupoid of quantum transitions rather than the group of symmetry.