A Bayesian model for a Pavlovian-instrumental transfer hypothesis

Emilio Cartoni
Laboratory of Computational Embodied Neuroscience
Istituto di Scienze e Tecnologie della Cognizione
Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR)
Via San Martino della Battaglia 44, I-00185 Roma, Italy
emilio.cartoni@yahoo.it

Francesco Mannella
Laboratory of Computational Embodied Neuroscience
Istituto di Scienze e Tecnologie della Cognizione
Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR)
Via San Martino della Battaglia 44, I-00185 Roma, Italy
francesco.mannella@istc.cnr.it

Stefano Puglisi-Allegra
Dipartimento di Psicologia
Sapienza Università di Roma
Centro Daniel Bovet
Via dei Marsi 78, I-00185 Roma,Italy
stefano.puglisi-allegra@uniroma1.it

Gianluca Baldassarre
Laboratory of Computational Embodied Neuroscience
Istituto di Scienze e Tecnologie della Cognizione
Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR)
Via San Martino della Battaglia 44, I-00185 Roma, Italy
gianluca.baldassare@istc.cnr.it

Abstract

A Pavlovian conditioned stimulus (CS) associated with a reward can enhance an instrumental response directed to the same or other rewards. This effect is called Pavlovian-instrumental transfer (PIT). In recent years, lesion studies using rats have gained insight into its neural substrates dissociating between specific PIT (where CS and instrumental response share the same reward) and general PIT (where they do not) (Corbit and Balleine, 2005, 2011). Despite these advances, the functional differences between specific and general PIT and how Pavlovian cues interact with instrumental response are still not clear. Here we try to explain Pavlovian-instrumental transfer effects by using a latent causes Bayesian model. Previous work in the Pavlovian conditioning literature (Courville et al., 2005) suggests that during Pavlovian conditioning rats do not simply learn associations between two events (CS and reward); instead, they actually try to figure out the real hidden causes behind them by constructing a latent cause model. We expanded that view to include instrumental actions and so explain the interactions between Pavlovian conditioning and instrumental conditioning. Our model correctly reproduces both the presence of specific and general PIT and the absence of general PIT when the CS is associated to the reward of another instrumental action. By framing the PIT effects explanation in Bayesian terms, our model offers a new integrated view on their functional mechanisms and new testable predictions.

Keywords: Pavlovian-instrumental transfer, Bayesian network, latent causes

Acknowledgements

This research has received funds from the European Commission 7th Framework Programme (FP7/2007-2013), “Challenge 2 - Cognitive Systems, Interaction, Robotics”, Grant Agreement No. ICT-IP-231722, Project “IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots”.

1 Introduction

A Pavlovian conditioned stimulus (CS) associated with a reward can enhance an instrumental response directed to the same or other rewards. This effect is called *Pavlovian-instrumental transfer* (PIT). As an example, in a typical experimental scenario, a rat is trained to associate a sound (CS) with the delivery of food. Later, the rat is subject to an instrumental training where it learns to press a lever to get some food (without the sound being present). Finally the rat is presented again with the opportunity to press the lever, this time both in the presence and absence of the sound. The results show that the rat will press the lever more in the presence of the sound than without, even if the sound has not been previously paired with lever pressing. The Pavlovian sound-food association learned in the first phase has somehow transferred to the instrumental situation, hence the name Pavlovian-instrumental transfer.

In recent years, this effect has been further subdivided into specific and general PIT. Specific PIT happens when the CS is paired with the same reward of the instrumental action. Instead, general PIT happens when the CS is paired with a different reward. In both cases the presence of the CS leads to higher instrumental responding, however, different neural substrates are involved (Corbit and Balleine, 2005, 2011). Specific PIT involves the *basolateral amygdala* and the *shell* part of *nucleus accumbens*. General PIT involves *central amygdala* and the *core* part of *nucleus accumbens*. While most of the Pavlovian-instrumental transfer experiments have been done with rats, PIT is also present in human subjects and seems to involve the same brain structures (Prévost et al., 2012). Despite these advances in associating PIT effect to specific brain areas, the specific neural mechanisms underlying it are still unknown.

At the functional level, the picture is not fully clear either.

Both Pavlovian and instrumental learning are often thought about in associationist terms. In associationist terms, Pavlovian conditioning leads to learning stimulus-outcome associations while instrumental conditioning can lead to associations between responses and their outcomes. One straightforward way of explainining PIT could be then in terms of a stimulus-outcome-response (S-O-R) chain. According to this view, during Pavlovian learning the subject learns a stimulus-outcome association (S-O); while during instrumental training both a R-O (response-outcome) association and its inverse O-R (outcome-response) association are learned. In the PIT test phase, hearing the sound (S) triggers the activation of the food outcome representation (O) thanks to the S-O association; this representation in turn activates its associated response through the O-R association, thus increasing instrumental responding. In general PIT, however, the outcome in the S-O association is not the same outcome of the O-R association. This case should then be explained by referring to the general motivating properties of a rewarding outcome instead of its specific sensory properties, so that the CS presence can still enhance instrumental responding. Even if the S-O-R chain and the "general motivating effect" were good explanations for specific and general PIT, some data would still remain unexplained. In particular, they would not explain why there is no general PIT effect when the conditioned stimulus is associated with a reward given by a different instrumental response than the one currently available. In this case, one would expect that, as the CS-evoked reward is different compared to the one currently available through instrumental action, a non-specific (general) PIT effect should happen. The absence of any enhancing PIT effect in this particular condition is currently attributed to a non-well defined inhibitory effect (Corbit and Balleine, 2005, 2011).

With our model, we try to unify the explanations of specific PIT, general PIT and the absence of general PIT into one single integrated view. Within this integrated view, framed in Bayesian terms, we also provide a more detailed functional explanation of these PIT effects and new testable predictions. As neural activity can also be thought in terms of probabilistic computation, hopefully this framework will also spark new ideas to bridge the gap between the functional level and the yet to be discovered neural mechanisms.

2 A Bayesian model of PIT

In the Pavlovian literature, Pavlovian learning is usually thought as S-O learning. That is, Pavlovian learning is simply learning the association between a stimulus (S) and an outcome (O). In particular, associationist models usually focus on the predictive property of S. For example, models such as Rescorla-Wagner or Pearce try to explain how S comes to predict outcome O. However, a different approach exists. For example, Courville et al. (2005) describe Pavlovian learning using a Bayesian generative model with hidden latent causes. In this model, the agent does not simply try to learn how often O occurs after S. Instead, the agent tries to learn the whole generative model: that is, it tries to learn the hidden cause that makes both S and O appear. Indeed, this is also a more rational strategy by the agent, as in Pavlovian experiments, S does not really cause O, but the two simply occur together because of a common cause (the experimental setup). By using this model, Courville et al. explain many phenomena that are otherwise not accounted for by classical associationist models such as Rescorla-Wagner or Pearce ones (Courville et al., 2005).

Inspired by this work, we are extending it to the instrumental realm to try and explain PIT. As in Courville et al. (2005), our model is composed by a sigmoid belief network. Activations of each node in the network represent the probability of that event happening and are computed using the following equation:

$$P(S_i|w, x) = (1 + exp(-(w^i)^T x - w_{bias}))^{-1} \tag{1}$$

where x is a vector of the activations of node ancestors and w^i is a vector of the weights of incoming links between the ancestors and node S_i. Given x and w^i the probability of each node S_i being active is the sigmoid function of the sum of the activations of its ancestors x (latent causes and the action node) each multiplied by their respective link weights to that node (w^i), plus a bias weight w_{bias}. In other words, each weight represents the influence of the presence of a node on its child nodes. This influence can be both positive or negative, rising or lowering the chances of the child node of happening. The bias weights represent the probability of each node of happening when its ancestors are not active.

We expanded Courville et al.'s model by adding an action node. This represents the act of pressing a lever. The action node is linked to each food reward, but the weight of each link in the model is initially unknown: the agent has to learn during trials how his actions can influence the presence of food and how latent causes influence the presence of both food and other observables (levers and sounds). In Courville et al.'s work, the model structure is learned (i.e. how many hidden causes are present and the number of their links). For simplicity, in the current model only link weights are learned, while the number of hidden causes and their links are given. Learning link weights means that the model has to discover how the hidden causes are affecting the probabilties of each observable. As in Courville et al. (2005), learning is done by using the Monte Carlo Markov Chain method (MCMC). Using MCMC, the model is first fitted to data representing pavlovian and instrumental conditioning trials. Then, the resulting model with weights based on the pavlovian and instrumental "experience" is tested to see if it shows PIT effects in PIT testing conditions (see next section). Indeed the trained model can account for the various PIT effects. We are working on developing a version of the model where the number of hidden variables and the number of their links are also learned, since the agent (usually a rat) does not know in advance how many causes will be present in the experimental scenario.

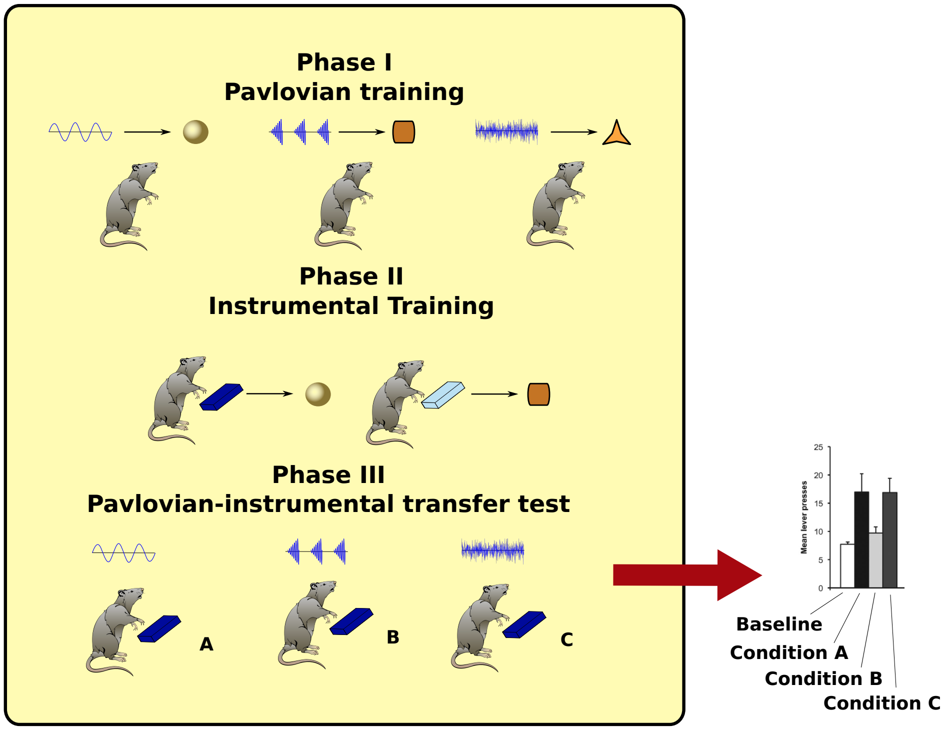

Figure 1: PIT paradigm to distinguish specific and general PIT. The histogram, which shows typical results, is from Corbit and Balleine (2011).

3 Results

The PIT paradigm that distinguishes specific and general PIT works as follows (see also Figure 1). It involves three phases: a Pavlovian phase, an instrumental phase and a test phase. In the Pavlovian phase, three sounds are associated with delivery of three different foods. In the instrumental phase, the rat undergoes two separate trainings: in each of these two trainings it learns to press a lever for food. A non-continuous reinforcement schedule is used, so the relationship between pressing the lever and obtaining food is probabilistic. For example, if a random-ratio RR20 schedule is used, the rat will get the reward about once every 20 lever presses. The rewards used for the instrumental phase are two of the foods previously used in the Pavlovian training. In the final phase PIT is tested. The rat is presented with one of the levers and each of the three sounds are played separated by some interval. When there is no sound, the rat will press the lever with a certain frequency (baseline). During two of the three sounds it will press the lever more than the baseline (PIT effect). Specifically, it will press more when it hears either the sound associated with the same reward of the lever (specific PIT) or the sound associated with a reward not used in the instrumental phase (general PIT). The sound associated with the reward of the other lever will not augment instrumental responding to the tested lever. We will now describe how our model behaves in each phase and how it accounts for these findings.

3.1 Pavlovian phase

During Pavlovian training, the rat sees that sound S1 and food F1 are correlated, so it assumes that an hidden cause H1 is generating both events. This happens for all the three Pavlovian trainings, thus generating positive weights between each of the three hidden causes (H1, H2, H3) and its pair of sound and food events (see Figure 2).

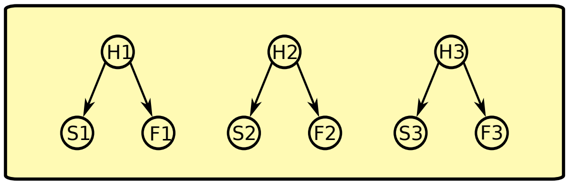

Figure 2: Pavlovian phase: positive links between latent variables and co-occurring sounds and foods are established.

3.2 Instrumental phase

During instrumental learning, the rat sees that the presence of a lever L1 and food F1 availability are correlated, so it assigns a positive weight between H4 and L1 and between H4 and F1 (see Figure 3). Food delivery, however, depends also on the action of lever pressing (A), so a positive link between action and food is also formed. The same happens for the other instrumental learning with lever L2 and food F2. Remarkably, the rat also learns a negative association (dashed line) between the “instrumental” hidden cause and the other food: in other words, it knows that when H4 is active and lever L1 is present, lever pressing (A) will not obtain food F2.

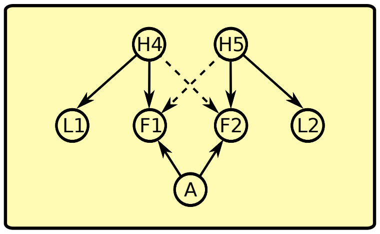

Figure 3: Instrumental phase: instrumental latent causes, which give rise to the presence of the levers, interact with action to make food available. Instrumental latent causes also inhibit each other’s food availability (negative links depicted as dashed lines).

3.3 Test phase

The test phase involves four possible conditions depending on the presence of different sounds togheter with one of the levers. This will give rise to different patterns of activations in the model (see Figure 4).

Baseline: in the baseline condition, lever L1 is presented alone. The rat will press it with some frequency knowing from previous instrumental learning that when L1 is present, it can get food F1 by lever pressing (action A).

Condition A - specific PIT: from the presence of S1 and L1, the rat can infer the presence of causes H1 and H4, which both predict F1 in the near future. This motivates the rat to press the lever more than when L1 is present alone, without any sound, as there are now increased chances of getting F1 in the immediate future.

Condition C - general PIT: the presence of lever L1 and sound S3 implies that causes H4 and H3 might be present and that foods F1 and F3 will appear in the future. While food F3 is not a direct effect of lever pressing A, its predicted presence might nevertheless motivate the rat to press the lever more than the baseline condition. This is a different kind of motivation from Condition A: instead of agumenting the probablity of the food targeted by the action, it adds a new food reward to the scene. Indeed experimental evidence has proved that these two enhancement effects are mediated by different neural substrates (Corbit and Balleine, 2005, 2011).

Condition B - absence of general PIT: this condition is similar to condition C, but in this case food F2, evoked by sound S2 (through H2) is inhibited by cause H4. Thus only food F1 remains predicted and no enhancement is found compared to the baseline.

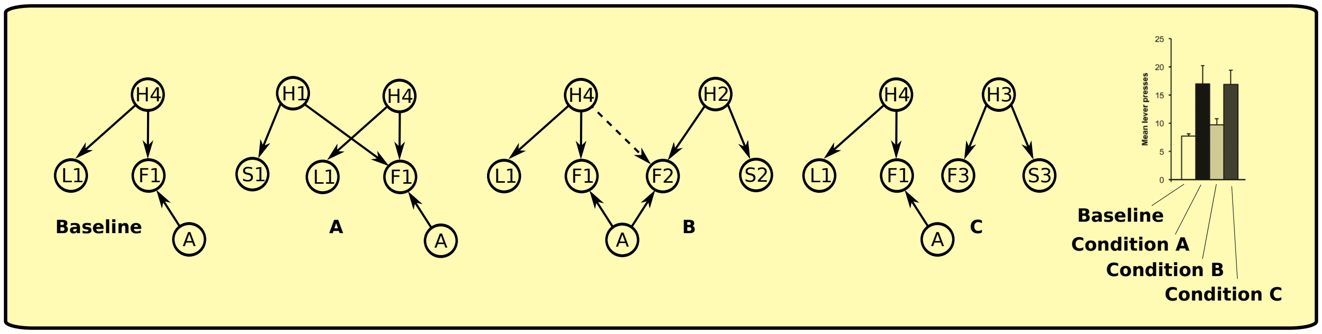

Figure 4: PIT test phase: hearing different sounds gives rise to different interactions in the model. Only relevant nodes from the previous learning phases are shown in each condition. The model can match typical experimental results, such as the histogram from (Corbit and Balleine, 2011).

4 Conclusions

The Bayesian model proposed here replicates basic findings on specific and general PIT. These two effects, which are known to be based on different neural substrates, also work in separate ways in our model. Specific PIT (condition A) enhances the predicted chances of obtaining the action-associated reward, while general PIT (condition C) motivates the action by placing an additional reward in the near future. We also explain the lack of PIT effects on condition B by invoking an inhibitory effect by one instrumental latent cause on the reward of the other instrumental latent cause. This might also be thought as some kind of “context” effect (i.e. in the context signalled by lever L1, you cannot get food F2). If our model is correct, it should be possible to derive and test specific predictions for each effect: for example, if specific PIT enhances the chances of getting a reward, it should have less or no effect if the reward is already certain.

The current implementation of the model relies on some simplifying assumptions, such as the model being “given” to the rat, which then has just to learn its link values - while in reality the rat would also have to guess and learn its structure, such as how many latent causes are out there. Besides improving and generalizing the model, a further challenge awaits: the model currently offers a functional view of PIT but no specific insight on its neural mechanisms. We plan to develop further models that bridge this gap between the functional and the neural implementation level. Studying PIT is a window on both Pavlovian and instrumental learning and can give insights on both habitual and goal-directed behavior in general. Knowledge gained on Pavlovian-instrumental transfer can shed light not only on normal behavior but also on pathological behaviors, such as addiction: a stronger than normal PIT is infact believed to be involved in the ability of drug cues to cause relapse into drug abuse.

References

Corbit, L. H. and Balleine, B. W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *The Journal of Neuroscience*, 25(4):962–70.

Corbit, L. H. and Balleine, B. W. (2011). The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *The Journal of Neuroscience*, 31(33):11786–94.

Courville, A. C., Daw, N. D., and Touretzky, D. S. (2005). Similarity and discrimination in classical conditioning : A latent variable account. *Advances in Neural Information Processing Systems*, 17.

Prévost, C., Liljeholm, M., Tyszka, J. M., and O’Doherty, J. P. (2012). Neural correlates of specific and general Pavlovian-to-Instrumental Transfer within human amygdalar subregions: a high-resolution fMRI study. *The Journal of Neuroscience*, 32(24):8383–90.