

When the Stimulus Is Predicted and What the Stimulus Predicts: Alternative  
Accounts of Habituation

Geoffrey Hall

University of York and University of New South Wales

Gabriel Rodríguez

University of Basque Country

Author Note

Geoffrey Hall, Department of Psychology, University of York, and School of Psychology, University of New South Wales; Gabriel Rodríguez, Facultad de Psicología, Universidad del País Vasco (University of Basque Country).

Correspondence concerning this article should be addressed to Geoffrey Hall, Department of Psychology, University of York, York, YO10 5DD, UK. E-mail: [geoffrey.hall@york.ac.uk](mailto:geoffrey.hall@york.ac.uk)

**© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xan0000237**

## Abstract

Wagner's fully elaborated theory of learning (e.g., Vogel, Ponce, & Wagner, 2019) was founded on an initial analysis of the mechanisms responsible for habituation (Wagner, 1976, 1979). Central to its explanation of long-term habituation was the proposal that a predicted stimulus, one signaled by some other event as a consequence of associative learning, would be less effective at activating its central representation. We review evidence (from studies of the role of context in habituation and latent inhibition, of preexposure to the event to be used as an unconditioned stimulus in conditioning, and of conditioned diminution effects) taken to support this explanation. We argue that the evidence is less than convincing and consider instead an alternative account that interprets habituation as reflecting a reduction in the effective salience of a stimulus that is determined by a learning process akin to extinction, in which the critical factor is that the stimulus is presented followed by no consequences. The application of this account to the phenomena dealt with by Wagner's model is considered and further implications are discussed.

When the Stimulus Is Predicted and What the Stimulus Predicts: Alternative  
Accounts of Habituation

Starting in 1976, Allan Wagner produced a series of publications, chapters in edited volumes for the most part (Wagner, 1976, 1978, 1979, 1981, 1985; also Brandon & Wagner, 1989, 2001; Mazur & Wagner, 1982; summed up in Vogel, Ponce, & Wagner, 2019), in which he advanced and developed a comprehensive account of conditioning that has a justifiable claim to be regarded as “the standard model” of the phenomenon (Roitblat, 1987; Hall, 1991). It built upon the Rescorla-Wagner (1972) model. To begin with, the Rescorla-Wagner model was, in essence, just an equation expressing the informal psychological notion that the predictive power of a conditioned stimulus (CS) (or alternatively the extent to which the occurrence of an unconditioned stimulus, US, is surprising) would determine the acquisition of associative strength. Wagner’s theorizing provided a structure, a collection of nodes representing stimuli, connected by excitatory or inhibitory links, that specified the *mechanisms* by which the phenomena described by the Rescorla-Wagner model might be generated.

Development of this basically simple structure generated an explanatory theory of ever-increasing power and complexity. Thus, for example, short-term memory could be interpreted in terms of residual activity in a node that had just been activated by its stimulus, long-term memory in terms of the activation of nodes by way of associative links established previously by certain patterns of co-activation of nodes. Concepts from the study of human cognition (such as rehearsal and priming) were brought within the scope of animal learning theory. Specification of patterns of nodal activation allowed a principled account of

when learning would be inhibitory rather than excitatory. The puzzling fact that the nature of the response generated by conditioning procedures might sometimes be opposed to that produced by direct presentation of a US was dealt with by the suggestion that associatively generated activation could be different in nature from that produced by direct presentation of a stimulus. Appreciating that stimuli of the sort we blithely describe as USs are complex events that have the important ability to activate emotional/motivational systems allowed expansion of the theory to deal with emotional as well as cognitive aspects of conditioning. And the fact that any event that the experimenter describes as “a stimulus” will consist a complex of elements paved the way for an elaborated analysis of discrimination learning that accommodated the fact that in some circumstances behavior appears to be controlled by configures, rather than simple cues.

Given the later development of the theory – its ability to provide an account of information processing in memory, of complex discrimination, of emotional as well as cognitive learning – it is easy to forget that the starting point for this theory, the foundation of this edifice, was an account of the simplest form of learning of all. Wagner’s theorizing was built upon his analysis of habituation (Wagner, 1976; see also Wagner & Vogel, 2010). In what follows we present an assessment of this aspect of the theory (in particular of its account of long-term habituation). Having noted some problems with it, we offer an alternative interpretation of long-term habituation, discuss some new predictions arising from this alternative, and consider the implications of our account for the theoretical analysis of associative learning more generally. What we offer is not novel, but rather it is an attempt to make explicit what has

previously only been implied in our previous work (Hall & Rodríguez, 2017, 2019).

### **Long-Term Habituation: Assessment of Wagner's Account**

Central among the observations that led to the Rescorla-Wagner (1972) model was the fact that a signaled (and thus, expected) US was poor at generating conditioning. The phenomenon of blocking provided a prime example. The explanation offered for blocking in terms of Wagner's developing theory (e.g., Wagner, 1976, 1978, 1979) was that the US representation (later referred to as the node) was less susceptible to being activated by its appropriate stimulus when it had already been activated (in this case associatively, by a previously trained CS).

The step taken by Wagner (1976), making the link to habituation, was to note the parallel with another situation in which a US appears to be reduced in its effectiveness. With repeated, fairly closely spaced presentations, a stimulus becomes increasingly poor at evoking its usual response (the UR, unconditioned response) – that is, habituation occurs. Wagner then proposed a common source for these phenomena with his suggestion that the state induced in a node by associative activation (anticipation of an event) was the same as that induced immediately after its presentation (short-term memory of an event). A node in this state (referred to as a secondary state of activation, A2, by Wagner, e.g., Wagner, 1981) was less able to respond to application of its stimulus. Habituation in the short-term is readily explained by the presence of the A2 state engendered by the stimulus that has been presented just a short time previously. For long-term habituation (the decrement in responding evident when a test is given hours or days after original training) it is necessary to

assume that some association has been formed during original training that induces the A2 state in the relevant stimulus node. The usual assumption has been that habituation training establishes an association between the target stimulus and the context in which it is presented, so that contextual cues become capable of evoking the A2 state.

There is some limited evidence addressing (and questioning) a central feature of Wagner's (1976) account -- the notion that the state that follows presentation of an event is the same as that generated by a predictor of the event (see, e.g., Linwick & Overmier, 2006). For the most part, however, assessment of the account has focused on its most direct novel prediction -- that long-term habituation will be specific to the context in which training was given. We present next a brief review of this work. We then go on to discuss a range of other related phenomena that have been put forward as supporting the general notion that a predicted stimulus is less effective than an unexpected one.

### **Context-Specificity of Long-term Habituation**

Hall (1991) presented a review of the work then available on the context-specificity of habituation. The conclusion (described as "less than kind" by Wagner & Vogel, 2010) was that habituation was not dependent on context as required by the theory. Subsequent work allows us to be a little kinder, although support for the theory remains less than full.

We should begin by noting that the failure, in some cases, of an habituated response to be restored by a change of context (a recent example is provided by the work of Pilz, Arnold, Rischawy, & Plappert, 2014) can be accommodated by a minor extension of Wagner's theory that still remains faithful to the basic notion that a stimulus is rendered less effective when it is

predicted. For the habituation procedure, the context of training is the most obvious candidate for the predictive cue; but, particularly for complex stimuli, it is quite possible that within-stimulus associations could play this role. Even a simple event like the presentation of a tone is a complex with properties of onset, pitch, duration, intensity, and so on. Perception of one aspect could come to signal the others and thus reduced their ability to evoke a UR even when the context is not that used in training. This extended version, which maintains the basic principle of Wagner's account, has been adopted and developed by McLaren, Kaye, and Mackintosh (1989; see also McLaren & Mackintosh, 2000) as their account of salience change.

Theoretically more critical are demonstrations of sensitivity to context. As the studies reviewed by Hall (1991) showed, there is no doubt that, in some circumstances, a response can return when, after habituation training, the stimulus is presented in a new context. But this does not require the conclusion that habituation depends on an association between the stimulus and the context in which training was given. The response might be restored because the new context, if it is unfamiliar, might sensitize the subject, raising its level of arousal, and making it responsive even to an habituated stimulus. Again, presenting the stimulus in a new context could well modify the way in which that stimulus impinges on the subject. Such generalization decrement would allow a response to be evoked by turning the stimulus into a novel event, effectively different from that given the initial training. More recent (i.e., since 1991) demonstrations of context sensitivity that are susceptible to explanation in these terms are found in studies by Tomsic, Massoni, and Maldonado (1993), Kruse, Stripling, and Clayton (2004), and by Chiandetti and Turatto (2017). To

deal with these issues, it is necessary to perform the test in a context with which the subject is already familiar and with a stimulus that is unlikely to be modified by the change of context. Of experiments that meet these criteria several (e.g., Hall & Channell, 1985; Hall & Honey, 1989) have found no evidence of restoration of the habituated response. But a null result cannot be decisive, and a more recent study by Jordan, Strasser, and McHale (2000) has shown convincingly that a rat's behavioral orienting response (OR) can be restored by a change of context.

Jordan et al. (2000) also found evidence that habituation of stimulus-evoked suppression of behavior might be sensitive to context, but, as Hall and Rodríguez (2017) have suggested, this may be simply a consequence of the concurrent change in the likelihood of the OR. And evidence that ORs, in particular, might be especially sensitive to context change effects comes from another source – studies directed, ostensibly, at the issue of recognition memory. In these (see Robinson & Bonardi, 2015, for a recent review), in what has become a standard procedure, the subject (usually a rodent) is placed in an open arena and allowed to explore a novel object placed in it. Contact (sniffing, touching, etc) is recorded and is found to decline with time. If on retest some time later the subject shows little exploration, it is asserted that a memory of the object is maintained (or equivalently that the OR is still habituated). Critically, for our purposes, it is reliably found that when given a retest in a different (but familiar) arena, the exploratory response returns. At least for cases in which the response studied is an OR, Wagner's account of long-term habituation appears to hold good.

### **Context and Latent Inhibition**



We have been concerned so far with assessing habituation directly, that is, by assessing the ability of a given stimulus to evoke its UR. But Wagner's theory holds that a stimulus node that has been "primed" into the A2 state will be less effective more generally. Thus, among other things, a preexposed stimulus will be less able to function as a CS if it subsequently employed in that role in a standard conditioning procedure; that is, the occurrence of latent inhibition is predicted, at least if the conditioning is given in the context in which exposure to the to be-CS was given. The effect should be absent or less powerful if the context is changed.

On this matter, in contrast to the varied results on the context-specificity of the habituated UR, there is a consensus. A change of context reliably attenuates or abolishes the latent inhibition effect (for reviews see, e.g., Hall, 1991; Holmes & Harris, 2010). The point is clearly made by results reported by Hall and Channell (1985). As we have already noted, this study found no evidence of restoration of the habituated response to a light when the stimulus was presented in a different context; but latent inhibition was abolished when, in a subsequent stage of training, the light was used as a CS in that context. The reason why habituation failed to show context sensitivity can be debated (see Honey, Good, & Manser, 1998; Honey, Iordanova, & Good, 2010) but it is clear that latent inhibition will show sensitivity to context change even when habituation itself does not.

Although demonstration of the context-sensitivity of the latent inhibition effect appears encouraging for Wagner's theory, there is reason to doubt that the effect is produced by the mechanism proposed by the theory -- that is, by way of a direct excitatory association between the context and the stimulus. In brief,

here are three lines of evidence that argue against the theory. First, latent inhibition has been shown to be enhanced when subjects are given prior exposure to the context (Hall & Channell, 1986), a procedure that might be expected to hinder the subsequent formation of a context-stimulus association. Next, Hall and Minor (1984; see also Baker & Mercier, 1982) investigated a related procedure in which exposure to the context alone was given after initial latent inhibition training. This had no effect on the magnitude of the latent inhibition effect obtained in a subsequent conditioning phase, in spite of the fact that such a procedure might be expected to allow extinction of the context-stimulus association on which latent inhibition is postulated to depend. Finally, Westbrook, Jones, Bailey, and Harris (2000), were able to confirm that latent inhibition was attenuated (i.e., a substantial conditioned response, CR, was established) when conditioning was given in a context other than that used for preexposure. This finding is, of course, consistent with the account proposed by Wagner. But Westbrook et al. went on to demonstrate that effect did not depend simply on the abolition of latent inhibition when the CS is unpredicted, as the Wagner theory would suppose. For when the subjects were returned to the original context in which preexposure had occurred the CR was reduced; that is, evidence of a latent inhibition effect was now found. Although conditioning in the second context appeared normal it was not able to express itself normally in a context in which nonreinforced presentations had occurred previously.

Results of this sort lead to the conclusion that the influence of context in these procedures is not (or is not solely) mediated by a direct association between the context and the stimulus presented in it. Rather the context may be thought to function as a conditional cue that facilitates the activation or effect of

associations formed in its presence. Hall and Mondragón (1998) interpreted this as a form of occasion-setting in which the context controls the functioning of the direct CS-US link. Equivalently, Bouton (e.g., 1993, 2004) has suggested that the context acts as a retrieval cue for the memory of the association formed in its presence, which, for latent inhibition would be the memory that no US had followed the target stimulus (Westbrook & Bouton, 2010).

### **The US-Preexposure Effect**

Just as a primed stimulus should, according to Wagner's theory be less effective as a CS, so also it should be less able to function effectively as a US. Preexposure to the event to be used as a US should retard subsequent conditioning, provided the context remains the same for the exposure and conditioning phases. The reality of this US-preexposure effect is well established (Kamin, 1961; Randich & LoLordo, 1979a; see Randich & LoLordo, 1979b, for a review of the early work), and there is no doubt that it can be attenuated by changing the contextual cues between the preexposure and conditioning phases. Randich and Ross (1985) present evidence for the role of the context in the US-preexposure effect in conditioning with shock as the US (see also Randich, 1981). And De Brugada, Hall, and Symonds (2004; see also Hall, 2009) reported a series of experiments on the effects of prior exposure to a nausea-inducing US on the subsequent conditioning of a flavor aversion. These experiments showed that this version of the US-preexposure effect was totally abolished when the cues associated with the US were changed between phases. (In this case the critical cues were supplied, not by the general experimental context but by those associated with the injection procedure used to supply the US.)

Two of the possible explanations for the US-preexposure effect, offered by Randich and LoLordo (1979b), were habituation and blocking. The former was taken to be the consequence of some form of nonassociative, adaptation process; the latter was the suggestion that cues signaling the US during the preexposure phase (usually those arising from the context) would become associated with the US and block conditioning when an explicit CS was introduced. From the point of view of Wagner's theory, of course, this distinction is inappropriate -- habituation (at least the long-term version of relevance to the procedures used in these experiments) is taken to be essentially the same phenomenon as that responsible for blocking. But to acknowledge the role of blocking in the US-preexposure effect is not to accept the Wagnerian account of long-term habituation -- there are, after all, alternative accounts of blocking (e.g., Mackintosh, 1975; Miller & Matzel, 1988; Pearce & Hall, 1980) that do not suppose that the effect depends on the reduced effectiveness of a signaled US. The results of most significance for our present purposes, therefore, would be any that demonstrate that the US-preexposure effect can be obtained when the contribution of blocking can be eliminated.

This issue was investigated thoroughly several decades ago, and we will briefly outline some findings that call into question the context-blocking account. The evidence offered by Randich and LoLordo (1979a) on the basis of experiments using the conditioned suppression paradigm is not conclusive, but it is suggestive. First, they noted that baseline rates of response in the presence of the contextual cues were largely unaffected by shock preexposure -- not what would be expected if the effect depended on shock preexposure establishing the context as a fear-evoking CS capable of blocking. Baker, Mercier, Gabel, and

Baker (1981) made the same point in a series of experiments in which the fear governed by the context was manipulated in a variety of ways but which failed to show any clear relation between context fear and the retardation of conditioning produced by shock preexposure. Randich and LoLordo (1979a) also noted that, according to the context blocking account, signaling the shock during preexposure would be expected to reduce context conditioning and thus attenuate or abolish the US-preexposure effect. Such an effect of signaling has sometimes been obtained (e.g., Baker et al., 1981; Randich, 1981), but it is clear that the effect can be found even when a signal is used in preexposure. Randich found that the retardation effect was quite unaffected by the signaling procedure when a low intensity US was used (the signal was effective when the US was of higher intensity).

Perhaps more telling than any of these points is the observation that exposure to a shock US alone can be effective in attenuating the magnitude of a CR, even when it is given *after* conditioning has been conducted (e.g., Randich & Haggard, 1983). This treatment can be expected only to increase the associative strength of the context. Thus, in the absence of special pleading, an associative theory of the effects of US presentations must predict that the magnitude of the CR will be enhanced as the strength possessed by the CS will be expected to summate with the extra strength possessed by the context. Although a simple associative explanation can be ruled out, we should acknowledge that the source of this effect is uncertain and likely to be multiply determined; a full explanation would need to accommodate the fact that the effect critically depends on the timing of presentation of the shocks and their intensity (see, e.g., Randich &

Rescorla, 1981). None the less, Randich and Haggard concluded that their results were best accommodated by nonassociative accounts of habituation.

### **Conditioned Diminution of the UR and CR**

Kimble and Ost (1961) observed, in a study of human eyeblink conditioning, that the magnitude of the UR declined over the course of conditioning. This phenomenon, which has been referred to as conditioned diminution of the UR (Kimmel & Pennypacker, 1962), was readily confirmed in a range of further studies (e.g., Baxter, 1966; Donegan, 1981; Grings & Schell, 1969; Kimmel, 1967). Critically, some of these studies included the control procedures required to demonstrate that the diminution was greater in the conditioning procedure than for the case in which the US was unsignaled. This outcome is to be expected, if, as is supposed by Wagner's (e.g., 1981) theorizing, the effectiveness of a stimulus is reduced when it is predicted or primed.<sup>1</sup>

The fact that in some experimental procedures this effect is not obtained – and indeed is sometimes apparently reversed, with the UR being enhanced rather than diminished by the presence of a CS (e.g., Brandon, Bombace, Falls, & Wagner, 1991; Donegan, 1981, Experiment 1; Leaton & Cranney, 1990) – does not seriously challenge the priming account. As training progresses the CS will, of course, come to evoke a CR, and if this response is the same as, or similar to, the UR its occurrence could obscure any diminution in the ability of the US to evoke the UR. Again, the conditioning procedure could allow the CS to evoke an emotional or motivational state that enhances the ability of even an habituated US to evoke its response. Thus, Leaton and Cranney measured the startle response to an auditory stimulus, and interpreted the enhanced responding they obtained as an instance of fear-potentiated startle. Similarly, Brandon et al. (see

also Wagner & Brandon, 1989) argue that the conditioned emotional response established by the use of an aversive (shock) US will potentiate both the eyeblink response to a paraorbital shock and the startle response evoked by an airpuff to the ear.

The other side of the coin is that demonstrations of diminution of the UR, when they do occur, cannot be taken as unambiguous support for the priming account. At its simplest, the problem is that the CR that will be established over the course of conditioning might interfere at a peripheral level, reducing the apparent strength of the UR, even when the US is otherwise fully effective. And several theorists, including Wagner (1981) himself (see also, e.g., Siegel, 2008; Young & Fanselow, 1992), have proposed that, at a central level, the nature of the CR is (or sometimes can be) to oppose the response that the US is “trying” to evoke.

Evidently, as Wagner and Vogel (2010) assert, a new experimental approach is required “to disentangle the conflicting influences” at work in studies of conditioned diminution. They offer a set of studies (by Brandon, Bell, and Wagner) that made use of the rabbit’s eyeblink response. These confirm that the UR evoked by a paraorbital shock is greater when the shock is preceded by a CS that has previously signaled that shock rather than a stimulus that has not been paired with the shock. Thus there was no evidence of conditioned diminution, but this effect might, of course, be obscured by the conditioned emotional state evoked by the CS. To control for this Brandon et al. (cited in Wagner & Vogel) trained a different CS as a signal for shock to the *other* eye. Such a stimulus would be expected to generate the same emotional state as the other, and to elevate responding above the level of the untrained stimulus. But a

difference between them, with the UR being less intense after presentation of the “correct” CS than after presentation of the other, would be consistent with the view that the correct CS was capable of generating diminution of the UR. The effect depends critically on the intensity of the shock and the duration of the CSs, but in some circumstances this was the result obtained.

Given the difficulties involved in clearly establishing the reality of conditioned diminution of the UR, it is worthwhile to consider a related phenomenon that speaks to the same issue, and that we may call “conditioned diminution of the CR”. If it is generally the case that the effectiveness of a stimulus is reduced when it is expected, then signaling the occurrence of a stimulus trained as a CS should reduce its ability to evoke its CR. As was true for conditioned diminution of the UR, there are complications when it comes to obtaining demonstrations of such an effect. If subjects are trained with the sequence A-B-US, then omission of A (leaving B unsignaled) should allow an increase in the CR to B. Although in a number of experiments the reverse result (a lesser response to B) has been obtained, this outcome is not decisive. A reduced response to B could be a consequence of generalization decrement (B having been experienced previously only with the after-effects of A present). Again, as Wagner himself has argued (Brandon & Wagner, 1991; Bombace, Brandon, & Wagner, 1991), stimulus A could supply a conditioned motivational background that enhances the ability of B to evoke its CR, an effect that might mask enhancement of the effectiveness of the unsignaled B. These considerations make any successful demonstration of conditioned diminution of the CR all the more intriguing.



A possible candidate for such a demonstration is found in the study by Terry and Wagner (1975) where the result is interpreted as indicating difference in short-term memory for surprising and expected events in the rabbit eyeblink preparation. A simpler and more direct demonstration is provided in experiments by Honey, Hall, and Bonardi (1991; see also Hall & Mondragón, 1998; Honey, 2000) on appetitive conditioning in rats. In their basic procedure rats received training with two visual cues, A and B (different lights) and two auditory cues, X and Y (noise and a tone), experiencing the sequences A-X-food, and B-Y-food. The CR measured was the tendency to approach the site of food delivery in the presence of the auditory cues (X and Y). On the critical test trials the rats received presentations of the auditory cues preceded by the “wrong” lights; that is, of A-Y and of B-X. On these trials the CR was more frequent than on orthodox trials when the auditory cue was preceded by its usual visual cue. It will be noted that generalization decrement consequent on the auditory cue being preceded by the wrong visual cue would tend to produce the opposite result – a reduction in the effectiveness of the latter cue. And any conditioned responses established to A and B (including emotional/motivational states) would be equated given that A and B had been treated equivalently during the training stage. These results are thus supportive of the proposal that a primed stimulus (one that is signaled, predicted, or expected) will be less effective than one that is not primed.

### **Conclusions**

Wagner’s (e.g., 1976, 1979) proposal that many instances of habituation are a consequence of “self-generated priming” is not controversial. The terminology was novel, but the idea that the application of a stimulus might

produce some short-term, fatigue-like state in the system connecting US to the UR has been widely accepted (e.g., Groves & Thompson, 1970; Thompson & Spencer, 1966; see also Thompson, 2009). Long-term habituation requires some other mechanism to produce a more permanent form of learning and Wagner turned to an associative process. His proposal that activating a stimulus representation or node would allow it to form an association with other nodes that are concurrently activated is similarly common ground. What is less obvious is his assertion that the state produced by associative activation is the same as that generated after the stimulus itself has been presented – that expectation of the occurrence of an event is psychologically equivalent to remembering that an event has just occurred. This brave theoretical step paid many dividends; but, as we have seen, direct evidence in support of the proposition is not fully convincing.

Given this background we turn now to an alternative view of (long-term) habituation that starts from the notion that the critical difference between a novel stimulus and one that has been repeatedly presented is that repeated presentation allows the animal to learn that the stimulus is without consequence. We present an account of a formal theory that tries to express this intuition, consider some new predictions, and evaluate the extent to which it can accommodate those observations that are taken to support the Wagnerian alternative.

### **What The Stimulus Predicts: Habituation and Extinction**

The essence of Wagner's (1976) theory of long-term habituation is that the effectiveness of a stimulus depends on what precedes it – more generally, on how well it is predicted. But we should not overlook what, it might be argued, is

a more obviously important aspect of the habituation procedure – the fact that the stimulus itself predicts nothing, that no event follows its presentation. In this, habituation is like experimental extinction. In both procedures a stimulus (a CS in the case of extinction) is presented repeatedly followed by no consequence; and in both the outcome is that the response elicited by the stimulus grows weaker or less probable.

The parallel between extinction and habituation is obvious and has often been noted. Humphrey himself, whose studies of *Helix* can be seen as the start of modern work on the topic, devoted some pages of his important book (Humphrey, 1933) to consideration of the parallel. (For more recent discussions of the parallel see Kling & Stevenson, 1970; McSweeney & Swindell, 2002; also Westbrook & Bouton, 2010, who look at the parallel between extinction and latent inhibition). And Thompson and Spencer (1966) whose influential review set out the framework for most subsequent discussions of habituation noted that the nine critical characteristics they identified for habituation can also be seen in extinction. They went on to say, however, that “to assert that habituation is really extinction does not of course constitute any kind of explanation for either process” (Thompson & Spencer, 1966, p. 29). What we present next, therefore, is a formal statement of a model for the process of extinction. We are then able to assess the extent to which its principles can be applied to, and generate an explanation of, habituation.

### **Extinction as No-US Learning**

Excitatory conditioning is taken to establish a CS-US association. One, increasingly popular, view of the effects of omitting the US, is that it generates a different form of association (often referred to, inelegantly, as a CS-no US

association) that opposes the effects of the first (e.g., Bouton, 1993; Konorski, 1967; Pearce & Hall, 1980; Rescorla, 2001). We will outline the way in which this notion is expressed in terms of the Pearce-Hall model, beginning with a brief account of how the model deals with excitatory conditioning.

The model supposes that excitatory associative strength ( $V$ ) is generated when a US and CS co-occur according to:

$$\Delta V = S\alpha\lambda \quad (1)$$

where  $\lambda$  represents the intensity of the US, and  $S$  the salience of the CS (taken to be directly related to its physical intensity)<sup>2</sup>. The values of these parameters may be assumed to determine the vigor of such responses as are controlled by these stimuli. The parameter  $\alpha$ , referred to as CS associability, reflects the aspect of attention that determines the readiness with which the CS enters into association. Asymptote is reached not because the US comes to be fully predicted (as in the various versions of Wagner's model) but because the CS fully predicts its consequences. The value of associability for a given CS changes with conditioning according to

$$\alpha^n = |\lambda - \Sigma V|^{n-1} \quad (2)$$

so that the value of  $\alpha$  on trial  $n$  is determined by the (absolute value) of the discrepancy between the value of  $\lambda$  and the summed associative strength ( $\Sigma V$ ) of all CSs that were present on the previous trial (trial  $n-1$ ). Thus learning will stop as  $\alpha$  falls to zero with the increase in  $\Sigma V$ . The  $\alpha$  parameter has been referred to as "attention for learning" (Hall & Rodríguez, 2017, 2019), and there is no requirement to suppose that its value will influence performance. It has been observed, however, in studies of the rat's overt OR to a visual cue, that the

frequency of this response will track the changes in  $\alpha$  expected on the basis of this model (Pearce & Hall, 1992).

If, after excitatory conditioning, the US is withheld (extinction), inhibitory learning occurs. This involves the formation of new association between the CS and some representation of no US; activation of the latter will oppose the effects of excitation of the US representation. This learning follows the same basic rules as excitatory conditioning, as shown in equation 3:

$$\Delta V_i = S\alpha\lambda_i \quad (3)$$

where  $V_i$  is inhibitory associative strength (i.e., the strength of the V- no US association) and  $\lambda_i$  the inhibitory reinforcer. The value of the inhibitory reinforcer will depend on the degree of surprise (or frustration or relief, for motivationally significant USs) generated when an anticipated event fails to occur. This will depend, therefore, on the excitatory strength acquired in acquisition, as follows:

$$\lambda_i = \Sigma V - \Sigma V_i \quad (4).$$

Finally, the occurrence of inhibitory learning requires us to amend equation 2 as follows:

$$\alpha^n = |\lambda - (\Sigma V - \Sigma V_i)|^{n-1} \quad (5).$$

The basic principle, that the value of  $\alpha$  declines when a stimulus predicts its consequences, remains unchanged.

### **Habituation as No-Event Learning**

Extension of this theory to the case in which a single stimulus is presented repeatedly (i.e., the habituation procedure) was considered, in the context of latent inhibition, by Hall and Rodríguez (2010a). Our starting point was the assumption that even a novel stimulus would not be truly neutral but

would evoke the expectation of some consequence. That is, our basic proposal was that a novel stimulus will activate an excitatory association with some representation of the occurrence of some other event: that  $V_{\text{event}}$  exists and has a positive value. Given that no event follows, the principles just described for extinction can be expected to operate in this case; that is, inhibitory learning will occur to counteract the inaccurate expectation that an event will follow the stimulus. In terms of the parallel with extinction after conditioning, the stimulus should come to activate an expectation of “no event” that will inhibit that for “event”.

More formally, but exactly paralleling our account of extinction, we suppose that the strength of the no event expectation grows over trials according to :

$$\Delta V_{\text{no event}} = S\alpha\lambda_{\text{no event}} \quad (6).$$

As was the case for the inhibitory reinforcer of equation 4, the value of  $\lambda_{\text{no event}}$  depends on the degree to which an event is expected; that is:

$$\lambda_{\text{no event}} = (V_{\text{event}} - V_{\text{no event}}) \quad (7).$$

And, as before, we assume value of  $\alpha$  will change as the stimulus comes to predict (the absence of) consequences. The parallel to equation 6 is:

$$\alpha^n = |\lambda_{\text{event}} - (\Sigma V_{\text{event}} - \Sigma V_{\text{no event}})|^{n-1} \quad (8).$$

To the extent that the response evoked by a novel stimulus is determined by the properties of what it predicts (just as the response to a CS depends, in part, on that evoked by its US), this account already gives us a possible source of habituation -- whatever response is evoked by the activation of the “event” representation will no longer occur, as that representation will be rendered inactive. But this cannot be whole story, as different novel stimuli (all of which

are assumed to activate the “event” representation) will have their own characteristic URs. The decline of such responses over the course of a series of habituation trials also needs to be accounted for. Hall and Rodríguez (2019) took the step of suggesting that experience with a stimulus will change not only associability ( $\alpha$ ) but also its salience ( $S$ ).

A novel stimulus will have a given initial level of  $S$  that will determine the attention paid to it a perceptual level, and also its ability to evoke responding. A reduction in  $S$  would thus equate to a reduction in the ability of a stimulus to evoke its UR. We now want to argue, therefore, that the effective salience (the value of  $S$ ) of a stimulus will decline as a consequence of the extinction process that occurs during a series of stimulus-alone presentation. The initial salience of a stimulus may be assumed to depend on its ability to activate the expectation of the occurrence of some consequent event, but this expectation will diminish with nonreinforced presentations. Hall and Rodríguez (2019) chose to express this in the following equation:

$$S^n = |V_{\text{event}} - V_{\text{no event}}| \quad (9)$$

by which the salience of a stimulus on trial  $n$  is equated with the net strength with which it activates the expectation that some event is going to occur on that trial. As  $V_{\text{no event}}$  grows, so the value of  $S$  declines and with it the likelihood of a UR. This formulation is compatible with the initial notion of the Pearce-Hall (1980) model that salience is dependent on physical intensity. We assume that an intense stimulus will readily activate an expectation of a consequent event and will thus have a high initial salience.<sup>3</sup>

In summary, we propose that a novel stimulus will activate the expectation of a consequent event. Repeated presentation of the stimulus will

result in inhibitory learning, eventually eliminating this expectation. As a consequence the effective salience of the stimulus declines. The result will be a stimulus that fails to activate any response controlled by the original expectation; and, because of its loss of salience, the stimulus will be rendered less able to evoke its own UR. We now need to assess how this account fares in dealing with the phenomena that we considered in our assessment of Wagner's account. These were phenomena that Wagner sought to explain in terms of the possible signaling role of contextual cues (in the US-preexposure effects, long-term habituation, latent inhibition), and possibly parallel phenomena in which a discrete cue served as the signal (conditioned diminution effects).

### **Context Effects**

The sensitivity of US-preexposure effects to the context in which the preexposure is given is readily explained in terms of blocking by contextual cues. This explanation can thus be readily derived from Wagner's theories, but as we have noted, it is also available to any theorist with an account of blocking. Interpretation of the role of context in latent inhibition and habituation is more critical in assessing rival theories of habituation.

We have already outlined (and endorsed) the proposal that the context in which associative learning occurs can come to act as an occasion-setter, promoting the effectiveness, or facilitating the retrieval, of associations formed in its presence. It has been argued (e.g., Bouton, 1993, Nelson, 2002) that inhibitory (or perhaps second-learned) associations are particularly susceptible to coming under contextual control and are thus particularly likely to be reduced in effectiveness when the context is changed. Bouton's analysis has focused primarily on extinction (i.e., on the effects of inhibitory learning after excitatory



conditioning), but the same analysis should, according to our account, hold for the inhibitory learning produced by simple exposure to a stimulus (i.e., by habituation training).

As Hall and Rodríguez (2010a) have pointed out, an explanation of the context-sensitivity of latent inhibition follows immediately from this account. Recall that the value of associability is given by:

$$\alpha^n = |\lambda_{\text{event}} - (\Sigma V_{\text{event}} - \Sigma V_{\text{no event}})|^{n-1} \quad (8)$$

and that the value of  $\alpha$  will be at or close to zero after stimulus exposure. If a change in context means that the effects of inhibitory learning cannot be retrieved, then the  $V_{\text{no event}}$  term will be reduced in value, and the resulting discrepancy will mean that the value of  $\alpha$  will be restored, so that excitatory conditioning will occur if the CS is paired with the US in a different context (i.e., latent inhibition will be attenuated). Changes in the value of  $S$ , in effective salience, will also contribute to this effect. The value of  $S$  will be low after training, as given by:

$$S^n = |V_{\text{event}} - V_{\text{no event}}| \quad (9).$$

But a failure to retrieve  $V_{\text{no event}}$  as a consequence of a change of context will allow  $S$  to be restored and promote the acquisition of associative strength (as given by equation 1).<sup>4</sup>

We have not previously considered the implications of these factors for the effects on the UR that are to be expected when the context is changed after habituation training. To do this we must begin by acknowledging that to refer to *the* UR is an oversimplification and that any stimulus is likely to evoke a range of responses. Any novel stimulus will evoke the complex of responses that is referred to as the OR, which will include behavioral orienting and a set of

changes mediated by the autonomic nervous system (Sokolov, 1963). This can be distinguished (see, e.g., Graham, 1979) from the pattern of autonomic and behavioral change that constitutes the defensive response (DR) that is evoked by sudden onset, intense, stimuli. To the extent that the vigor of a UR is determined by the intensity of the stimulus (as will be the case for DRs) then a loss of responsiveness with habituation training is to be expected, given that the decline in  $S$  produced by habituation training amounts to a change in effective stimulus intensity. It is problematic for this account that a change of context, which should restore the value of  $S$ , does not reliably result in restoration of the UR. We can only suppose that the effects produced by a change of context are often too slight to produce an observable effect on behavior, at least when the UR being measured is a defensive response. The picture may be different when the response studied is an orienting response for which, as we discussed previously, there is good evidence of restoration with a change of context.

We may assume that, even for an OR, the  $S$  parameter must have some value for a response to be obtained, and any increase in  $S$  as a consequence of a change of context will be expected to promote responding. But we have also argued that the value of the associability parameter,  $\alpha$ , will contribute to the likelihood of occurrence of an OR, and the value of this parameter is also enhanced when a change of context reduces the value of  $V_{\text{event}}$ . With two factors operating to reverse the effects of habituation training and to restore the response, our account comfortably accommodates the fact that a dishabituation effect with context change is particularly likely to be obtained when the response is an OR.

It would be remiss not to acknowledge a problem faced by the account just outlined. It is that several experiments, some with invertebrate subjects (Rankin, 2000; Tomsic, Pedreira, Romano, Hermitte, & Maldonado, 1998), one with rat subjects (Jordan et al., 2000), and one studying a version of the OR in human subjects (Turatto, Bonetti, & Pascucci, 2018), have shown that a period of exposure to the context, interposed between habituation training and a subsequent test, can result in recovery of an habituated response. This is consistent with the suggestion that habituation depends on the strength of a context-stimulus association that extinguishes when the context is presented alone. Such a treatment would not be expected to reduce the effectiveness of the occasion-setting properties of the context; there is substantial evidence to show that mere exposure to an occasion setter alone is not enough to eliminate its occasion-setting properties (e.g., Holland, 1992; Rescorla, 1986). If we are to maintain our interpretation, we need to assume that the effect of exposure to the context depends on some process other than extinction; that, for instance such exposure allows further habituation to, and loss of salience by, contextual cues and that this renders the target stimulus more effective when it is next presented. We must acknowledge, however, that the proposal that habituation depends on a direct context-US association is much more comfortable with these effects than is our proposal that allows the context only occasion-setting properties.

### **Conditioned Diminution Effects**

Emphasis on the role of context arises from the fact that the basic procedure for habituation involves presenting the stimulus alone without any signal to warn of its occurrence. A theory that supposes the habituation effect to

depend on the fact that the stimulus is predicted must rely on a role for contextual cues. But when it comes to testing the central notion of a theory of this sort a direct test can be arranged by providing an explicit cue that precedes the presentation of the target stimulus. Manipulation of this cue should allow a test of theory. As we have seen, interpretation of effects of this sort, under the heading of conditioned diminution of the UR, have often proved difficult to interpret, but clear results emerge from the procedure that we have called conditioned diminution of the CR.

In the version described above (by Honey et al., 1993) rats were trained with the sequences A-X-food, and B-Y-food, and tested with A-Y and B-X. The CRs evoked by Y and X were enhanced on these test trials compared with trials when the original arrangement was used. This result, above all others, seems to demand an explanation in terms of the diminution of the effectiveness of a signaled stimulus, by virtue of its being expected. But, given the lack of support for this general notion from the other behavioral phenomena we have discussed, it seems worthwhile to consider an alternative that can be derived from proposals of Hall and Rodríguez (2019). From this perspective, the focus should be not on stimuli X and Y, but on stimuli A and B. Our perspective emphasizes the importance of what a stimulus predicts, and in the test procedure used here, the consequences of X and Y are unchanged. A and B, on the other hand, are followed by unpredicted events on test, something that should, according to our theorizing, enhance the effectiveness of these stimuli and thus promote the occurrence of such conditioned responses as they control. An enhancement of responding recorded on the on the test can thus be expected, given that A and B

will be more effective in eliciting the response of approaching the food tray where food pellets are delivered following the occurrence of X or of Y.

At the risk of being unduly speculative, we will note the possibility that an analysis of the type just offered could be applied in explanation of the results described by Wagner and Vogel (2010). Recall that in these studies it was shown that the UR to a paraorbital shock was less when signaled by the “correct” CS rather than a CS that had been used to predict a shock to the contralateral eye. It was argued that the emotional state (a state likely to potentiate occurrence of the UR) would be equated for the two CSs, thus allowing the reduced UR in the presence of the correct CS to be unambiguously interpreted as an example of conditioned diminution. Our account challenges this assumption. Specifically we would suggest that the treatment given to the “incorrect” CS over the course of the test trials would, given the discrepancy between the test and initial training procedures, result in a restoration of stimulus effectiveness that had been lost in initial training. The outcome would be that the incorrect CS would be better able to evoke the emotional state established during conditioning. A heightened state of fear would promote occurrence of the eyeblink UR. The experimental result can thus be attributed, not to a diminution of the UR to the correct CS, but to a potentiation of responding in the presence of the incorrect CS.

### **Further Predictions**

We have been concerned so far to show that an account of long-term habituation that stresses the role of learning about what the stimulus predicts can successfully accommodate features of the phenomenon that have been interpreted in terms of how well the stimulus is predicted. We now consider evidence relating to effects that appear to be uniquely predicted by the view that

habituation depends on the consequences of the stimulus. This comes from consideration of the effects of manipulating the nature and scheduling of events following presentation of the target stimulus. Normally, of course, in the standard habituation procedure, no event follows; in the procedures to be described next we consider the case in which a salient event can follow.

Analysis of the effects produced by posttrial events played an important role in the early development of Wagner's theory of learning. The study by Wagner, Rudy, and Whitlow (1973), showing that a salient posttrial event could reduce the effectiveness of a CS-US pairing in producing conditioning, was especially influential. The interpretation was that the posttrial event disrupted the processing necessary for association formation. For some experimental procedures it has been shown that the occurrence of a posttrial event can modify the effect of presenting just a single stimulus on the trial – that is, habituation can be attenuated (e.g., Green & Parker, 1975; Shanks, Preston, & Stanhope, 1986). The explanation that emerges directly from Wagner's theorizing (e.g., Wagner, 1976; 1981) is that the posttrial event disrupts formation of the association between context and target stimulus. But the arguments presented above, that challenge this associative account of habituation, prompt us to consider an alternative interpretation. The result is equally compatible with the interpretation that the changes in stimulus properties (in  $S$  and  $\alpha$ ) produced by stimulus presentations will occur more readily when the learning required is extinction of the expectation that an event will follow (i.e., when the target stimulus is presented alone) than when the new learning is the formation of an excitatory association with the subsequent event.

In order to distinguish between these rival accounts it will be useful to look at a different comparison -- between the case in which the posttrial event follows the target stimulus on every trial and one in which it is presented on only some trials. According to Wagner's (1976, 1981) theory both of these procedures should retard the development of long-term habituation by interfering with the development of the context-stimulus association. Which of these procedures will be more effective in this regard is not clear. Interference on every trial might be expected to be more effective than interference on only some trials; on the other hand, a posttrial event that occurs on only some trials will retain its surprising qualities and might, therefore, be especially effective on the trials on which it does occur. But the comparison of Wagner's account with that of Hall and Rodríguez (2010a, 2019) does not depend on resolving this issue. Rather it derives from the capacity of the latter to make separate predictions about changes in the effective salience and the associability of the target stimulus.

According to our account, the habituation procedure will generate changes in two separate aspects of the stimulus -- in its associability (the  $\alpha$  parameter) and its effective salience (the S parameter). (We have sometimes referred to these as attention for learning and attention for performance respectively; Hall & Rodríguez, 2017, 2019.) As we have seen, changes in these parameters obey different rules. Associability declines when the stimulus is followed reliably by a consequence; it is maintained when the consequence varies from trials to trial. Salience, by contrast, declines when the stimulus is followed by no event; it will be maintained when an event follows the stimulus and will be better maintained the stronger the association with its consequence,

and thus better maintained when the consequence occurs on all trials. Figure 1 presents simulations (using the equations presented previously) of changes in  $S$  and  $\alpha$  over a series of 40 trials in which target stimulus A is followed by another event on all trials (labeled consistent in the figure) or on 50% of trials (labeled inconsistent). The starting value for  $\alpha$  was set to a moderate value (.5); the initial salience of A was set to .4, that of its consequence to .8, these values being chosen in an effort to match the assumed properties of the stimuli used in the experiments to be reported next.

As the figure shows, the values of  $S$  and  $\alpha$  decline over trials (the effectiveness of the consequence will itself decline over trials as it is followed by no event). Critically, however we see that the value of  $\alpha$  is maintained at a higher level in the inconsistent condition whereas the value of  $S$  is higher in the consistent condition. We have argued previously that different URs are likely to be differently sensitive to different properties of a stimulus – that a defensive UR will be readily evoked by a stimulus high in salience whereas a high value for  $\alpha$  would generate a strong OR. This proposition allows for an empirical test of the implications of the effects shown in the Figure 1 simulations.

The upper panel of Figure 2 shows the results of an experiment reported by Hall and Rodríguez (2010b; using data from Hancock, 2007) that investigated habituation to a shock in rats. Rats were given 20 trials of preexposure to a weak shock. For rats in the consistent condition the shock was followed on every trial by a 60-s presentation of a loud noise; rats in the inconsistent condition received the noise after a random 50% of shock presentations. The next stage of training tested the properties of the shock by using it as the US in a conditioning procedure. Our assumption was that a shock that had lost effective salience



would function less well as a reinforcer in this procedure. The results in Figure 2 show the acquisition of conditioned suppression to a light CS over the course of 6 sessions of conditioning (with two trials per session). It is evident that suppression was acquired more readily in the consistent than in the inconsistent condition. This accords with our prediction (upper panel of Figure 1) that the consistent arrangement would be more effective than the inconsistent in preserving the salience of the shock.

As the lower panel of Figure 1 shows, a quite different pattern of result is to be expected when the response studied is an OR (given the assumption that the OR is primarily determined by the value of  $\alpha$ ). In this case the loss of the response should be greater in the consistent condition. This prediction has been confirmed in experiments both with people and with rats. Lovibond (1969) reported a study of the human OR (the skin conductance response) evoked by presentation of a light. For some subjects the light was followed by a tone. Habituation proceeded readily when the tone was presented on all trials, but it was much attenuated when the tone occurred on a random 50% of trials. For rats the OR of rearing before and approaching a signal light in a Skinner box has been extensively studied by Pearce and his colleagues (e.g., Kaye & Pearce, 1984; Pearce, Wilson, & Kaye, 1988; Swan & Pearce, 1988). The lower panel of Figure 2 shows the results of the simplest of these (from Kaye & Pearce, 1984). Over the course of 14 days rats received presentations of a 10-s light and the proportion of presentations evoking an OR was scored. The light was followed by a poststimulus event, the presentation of a food pellet -- on all trials for rats in the consistent condition, on a random half of the trials for rats in the inconsistent

condition. It is evident that the OR declined more readily in the latter condition, matching the theorized change in  $\alpha$  shown in the lower panel of Figure 1.

### Conclusions

It is a fact of everyday experience that signaling the upcoming occurrence of an event can change the reaction to it – if you are told that a loud noise is just about to occur your response will be different from when the noise occurs without a signal. The signal can allow the organism to prepare for the noise that is about to occur, and an appropriate conditioned response to the signal (a response that could be peripheral – putting your hands over your ears – but may well be a more subtle central equivalent) will reduce the impact of the noise. This notion (in a rather more sophisticated form) is central to Wagner's (e.g., 1976, 1979) theory of habituation, and to his account of learning more generally. By accepting the reality of a process of this sort we are accepting that the account of learning that we have been developing here, with its stress on what the stimulus predicts, cannot be the whole story. A fully comprehensive account of learning will need to incorporate mechanisms by which the effectiveness of a stimulus presentation will be determined both by the *predictiveness* of the stimulus and also by its *predictability*.

To acknowledge this is not to accept that an association between signal and stimulus is *the* mechanism responsible for (long-term) habituation. Our review of the evidence provides little support for Wagner's proposal that associative activation of a stimulus node is, in itself, effective in reducing the responsiveness of that node to external stimulation. We do not reject the idea that the node becomes less responsive (although we describe the phenomenon as a reduction in the effective salience of the stimulus, this amounts to the same

thing). But we suggest (and provide preliminary evidence in support of the view) that the change depends not on the stimulus being predicted but on learning about what it predicts. This general idea is not particularly novel – the notion has been central to a range of theories of learning that find a role for changes in attention to the stimuli (e.g., Mackintosh, 1975, Pearce & Hall, 1980). These theories had a single attentional construct (associability) that could change with experience. What we have done (in common with some more recent theories, e.g., George & Pearce, 2012; Le Pelley, 2004) is to propose that there is more than one form of attention and that changes to different forms are likely to obey different rules. The rules we have described suggest that attention for learning (associability) will depend on how well or badly a stimulus predicts its consequences; attention for performance (effective salience) is determined simply by the strength of the expectation that some consequence will follow.

As we have already acknowledged, a strength of Wagner's theorizing was that it proposed mechanisms (rather than mere equations), generated real-time predictions, and led on to a range of predictions and explanations concerning (associative) learning phenomena quite generally. What we have offered ourselves is sadly inadequate in some of these respects. It is worth noting, therefore, that the essence of the alternative approach to habituation that we have been discussing here – the proposal that changes in the properties of a stimulus can be determined by its consequences – has similarly generated a range of theories of associative learning that have general applicability (e.g., George & Pearce, 2012; Le Pelley, Mitchell, Beesley, George, & Wills, 2016; Mackintosh, 1975; Pearce & Hall, 1980).

## Footnotes

1. Studies of conditioned diminution of the UR with human subjects have for the most part used an electric shock US. Formally equivalent experiments have been conducted with two neutral stimuli (S1 followed by S2, e.g., a tone and a light) and with the OR to S2 as the critical response measure (e.g., Siddle & Spinks, 1992). These show that, after prior exposure to the S1-S2 sequence, the response to S2 is enhanced when it is presented without S1 as a precursor (Siddle, Broekhuizen, & Packer, 1990), a result that could be taken to be an instance of conditioned diminution (S1 being the CS and S2 the US). But as Mackintosh (1987, 1988) has pointed out, this result can be interpreted as being an instance of generalization decrement, S2 on test being effectively a novel stimulus, as it is presented for the first time in the absence of the aftereffects of S1.

2. We acknowledge that this is a simplification, and that salience will be determined by more than simple physical intensity. It is evident, for example, that a reduction in intensity, such as a reduction from bright to dim, could be a salient event. And the effectiveness of a given event can clearly differ across species according to their different phylogenetic histories (compare the reaction to a snake of a chimpanzee and a mongoose).

3. In the original Pearce-Hall (1980) model the intensity of the CS determined the value not only of  $S$ , but also of  $\alpha$ , with the latter but not the former changing with experience. In simulations of the present version (Hall & Rodríguez, 2019)

we have eliminated this duplication using a common starting value of  $\alpha$  for all stimuli, regardless of salience.

4. Evidence taken to demonstrate the role of the loss of salience by the preexposed stimulus in a (human) latent inhibition procedure is offered by Rodríguez, Aranzubia-Olasolo, Liberal, Rodríguez-San Juan, and Hall (2019).

## References

- Baker, A. G., & Mercier, P. (1982). Extinction of the context and latent inhibition. *Learning and Motivation, 13*, 391-416.
- Baker, A. G., Mercier, P., Gabel, J., & Baker, P. A. (1981). Contextual conditioning and the US preexposure effect in conditioned fear. *Journal of Experimental Psychology: Animal Behavior Processes, 2*, 109-128.
- Baxter, R. (1966). Diminution and recovery of the UCR in delayed and trace classical GSR conditioning. *Journal of Experimental Psychology, 71*, 447-451.
- Bombace, J. C., Brandon, S. E., & Wagner, A. R. (1991). Modulation of a conditioned eyeblink response by a putative emotive stimulus conditioned with hindleg shock. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 323-333.
- Bouton, M. E. (1993). Context, time and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114*, 80-99.
- Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.
- Brandon, S. E., & Wagner, A. R. (1991). Modulation of a discrete Pavlovian conditioned reflex by a putative emotive Pavlovian conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 299-311.
- Brandon, S. E., Bombace, J. C., Falls, W. A., & Wagner, A. R. (1991). Modulation of unconditioned defensive reflexes by a putative emotive Pavlovian conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 312-322.

- Chiandetti, C., & Turatto, M. (2017). Context-specific habituation of the freezing response in newborn chicks. *Behavioral Neuroscience*, *131*, 437-446.
- De Brugada, I., Hall, G., & Symonds, M. (2004). The US-preexposure effect in lithium-induced flavor-aversion conditioning is a consequence of blocking by injection cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 58-66.
- Donegan, N. H. (1981). Priming-produced facilitation or diminution of responding to a Pavlovian unconditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 295-312.
- George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning & Behavior*, *40*, 241-254.
- Green, K. F., & Parker, L. A. (1975). Gustatory memory: Incubation and interference. *Behavioral Biology*, *13*, 359-367.
- Graham, F. K. (1969). Distinguishing among orienting, defensive, and startle reflexes. In H. D. Kimmel, E. H. van Holst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 137-167). Hillsdale, NJ: Erlbaum.
- Grings, W. W., & Schell, A. M. (1969). UCR diminution in trace and delay conditioning. *Journal of Experimental Psychology*, *79*, 246-248.
- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review*, *77*, 419-450.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Clarendon Press.
- Hall, G. (2009). Preexposure to the unconditioned stimulus in nausea-based aversion learning. In S. Reilly & T. R. Schachtman (Eds.), *Conditioned taste aversion: Behavioral and neural processes* (pp. 58-73). New York, NY: Oxford University Press.

- Hall, G., & Channell, S. (1985). Differential effects of contextual change on latent inhibition and on the habituation of an orienting response. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 470–481.
- Hall, G., & Channell, S. (1986). Latent inhibition and conditioning after context preexposure. *Learning and Motivation*, *16*, 381-397.
- Hall, G., & Honey, R. C. (1989). Contextual effects in condition, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 232–241.
- Hall, G., & Minor, H. (1984). A search for context-stimulus associations in latent inhibition. *Quarterly Journal of Experimental Psychology*, *36B*, 145-169.
- Hall, G., & Mondragón, E. (1998). Contextual control as occasion setting. In Schmajuk, N. A., & Holland, P. C. (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 199-222). Washington, DC: American Psychological Association.
- Hall, G., & Rodríguez, G. (2010a). Associative and nonassociative processes in latent inhibition: An elaboration of the Pearce-Hall model. In R. E. Lubow & I. Weiner (Eds.), *Latent inhibition: Cognition, neuroscience and applications to schizophrenia* (pp. 114-136). Cambridge: Cambridge University Press.
- Hall, G., & Rodríguez, G. (2010b). Attentional learning. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour*. (pp. 41-70). Oxford: Oxford University Press.
- Hall, G., & Rodríguez, G. (2017). Habituation and conditioning: Salience change in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*, 48-61.



- Hall, G., & Rodríguez, G. (2019). Attention to perceive, to learn, and to respond. *Quarterly Journal of Experimental Psychology*, *72*, 335-345.
- Hancock, A. J. (2007). *The consistency of outcome effect* (Doctoral thesis, University of York, UK).
- Holland, P.C. (1992). Occasion setting in Pavlovian conditioning. *Psychology of Learning and Motivation*, *28*, 69-125.
- Holmes, N. M., & Harris, J. A. (2010). Latent inhibition. In Mitchell, C. J., & Le Pelley, M. E. (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 99-130). Oxford: Oxford University Press.
- Honey, R. C. (2000). Associative priming in Pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, *53B*, 1-23.
- Honey, R. C., Good, M., & Manser, K. L. (1989). Negative priming in associative learning: Evidence from a serial-habituation procedure. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 229-237.
- Honey, R. C., Hall, G., & Bonardi, C. (1993). Negative priming in associative learning: Evidence from a serial conditioning procedure. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 90-97.
- Honey, R. C., Iordanova, M. D., & Good, M. (2010). Latent inhibition and habituation: Evaluation of an associative analysis. In R. E. Lubow & I. Weiner (Eds.), *Latent inhibition: Cognition, neuroscience and applications to schizophrenia* (pp. 163-182). Cambridge: Cambridge University Press.
- Humphrey, G. (1933). *The nature of learning*. London: Kegan Paul.
- Jordan, W. P., Strasser, H. C., & McHale, L. (2000). Contextual control of long-term habituation in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 323-339.

- Kamin, L. J. (1961). Apparent adaptation effects in the acquisition of a conditioned emotional response. *Canadian Journal of Psychology, 15*, 176–188.
- Kaye, H., & Pearce, J. M. (1984). The strength of the orienting response during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 99–109.
- Kimble, G. A., & Ost, J. W. P. (1961). A conditioned inhibitory process in eyelid conditioning. *Journal of Experimental Psychology, 61*, 150-156.
- Kimmel, E. (1967). Judgments of UCS intensity and diminution of the UCR in classical GSR conditioning. *Journal of Experimental Psychology, 73*, 532–543.
- Kimmel, H. D., & Pennypacker, H. S. (1962). Conditioned diminution of the unconditioned response as a function of the number of reinforcements. *Journal of Experimental Psychology, 64*, 20–23.
- Kling, J. W., & Stevenson, J. G. (1970). Habituation and extinction. In G. Horn & R. A. Hinde (Eds.), *Short-term changes in neural activity and behaviour* (pp. 41-61). Cambridge: Cambridge University Press.
- Konorski, J. (1967). *Integrative activity of the brain*. Chicago: University of Chicago Press.
- Kruse, A. A., Stripling, R., & Clayton, D. F. (2004). Context-specific habituation of the *zenk* gene response to song in adult zebra finches. *Neurobiology of Learning and Memory, 82*, 99-108.
- Leaton, R. N., & Cranney, J. (1990). Potentiation of the acoustic startle response by a conditioned stimulus paired with acoustic startle stimulus in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 16*, 279–287.

- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology, 57B*, 193–243.
- Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N. & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin, 142*, 1111-1140.
- Linwick, D. C., & Overmier, J. B. (2006). Associatively activated representations of food events resemble food outcome expectancies more closely than they resemble food-based memories. *Learning & Behavior, 34*, 1-12.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276–298.
- Mackintosh, N. J. (1987). Neurobiology, psychology and habituation. *Behaviour Research and Therapy, 25*, 81-97.
- Mackintosh, N. J. (1988). Comparator theories of habituation: A reply. *Biological Psychology, 27*, 65-67.
- McLaren, I. P. L., Kaye, H., Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.G.M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102-130). Oxford: Clarendon Press.
- McLaren, I. P. L. & Mackintosh, N. J. (2000). Associative learning and elemental representations. I: A theory and its application to latent inhibition and perceptual learning. *Animal Learning & Behavior, 26*, 211-246.
- McSweeney, F. K., & Swindell, S. (2002). Common processes may contribute to extinction and habituation. *Journal of General Psychology, 129*, 364-400.

- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. *Psychology of Learning and Motivation*, 22, 51-92.
- Nelson, J. B. (2002). Context specificity of excitation and inhibition in ambiguous stimuli. *Learning and Motivation*, 33, 284-310.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Pearce, J. M., & Hall, G. (1992). Stimulus significance, conditionability, and the orienting response in rats. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (pp. 137-160). Hillsdale, NJ: Erlbaum.
- Pilz, P. K. D., Arnold, S. W., Rischawy, A. T., & Plappert, C. F. (2014). Longterm-habituation of the startle response in mice is stimulus modality, but not context specific. *Frontiers in Integrative Neuroscience*, 7, Article 103.
- Randich, A. (1981). The US preexposure phenomenon in the conditioned suppression paradigm: A role for conditioned situational stimuli. *Learning and Motivation*, 12, 321-341.
- Randich, A., & Haggard, D. (1983). Exposure to the unconditioned stimulus alone: Effects on retention and acquisition of conditioned suppression. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 147-159.
- Randich, A., & LoLordo, V. M. (1979a). Preconditioning exposure to the unconditioned stimulus affects the acquisition of a conditioned emotional response. *Learning and Motivation*, 10, 245-277.

- Randich, A., & LoLordo, V. M. (1979b). Associative and nonassociative theories of the UCS preexposure phenomenon: Implications for Pavlovian conditioning. *Psychological Bulletin*, *86*, 523–548.
- Randich, A., & Rescorla, R. A. (1981). The effects of separate presentations of the US on conditioned suppression. *Animal Learning & Behavior*, *9*, 56-64.
- Randich, A., & Ross, R. T. (1985). Contextual stimuli mediate the effects of pre- and postexposure to the unconditioned stimulus on conditioned suppression. In P. D. Balsam & A. Tomie (Eds.), *Context and learning* (pp. 105-132). Hillsdale, NJ: Erlbaum.
- Rankin, C. H. (2000). Context conditioning and habituation in the nematode *Caenorhabditis elegans*. *Behavioral Neuroscience*, *114*, 496-505.
- Rescorla, R. A. (1986). Extinction of facilitation. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*, 16–24.
- Rescorla, R. A. (2001). Experimental extinction. In R. R. Mowrer & S. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 119-154). Hillsdale, NJ: Erlbaum.
- Robinson, J., & Bonardi, C. (2015). An associative analysis of object memory. *Behavioural Brain Research*, *285*, 1-9.
- Rodriguez, G., Aranzubia-Olasolo, M., Liberal, U., Rodriguez-San Juan, F., & Hall, G. (2019). Loss of salience as a source of latent inhibition in human associative learning. *Quarterly Journal of Experimental Psychology*, *72*, 1047-1054.
- Roitblat, H. L. (1987). *An introduction to comparative cognition*. New York: Freeman.

- Shanks, D. R., Preston, G. C., & Stanhope, K. (1986). Effects of distractor familiarity on habituation of neophobia. *Animal Learning & Behavior*, *14*, 393-397.
- Siegel, S. (2008). Learning and the wisdom of the body. *Learning & Behavior*, *36*, 242-252.
- Siddle, D. A. T., Broekhuizen, D., & Packer, J. S. (1990). Stimulus miscuing and dishabituation: Electrodermal activity and resource allocation. *Biological Psychology*, *31*, 229-243.
- Siddle, D. A. T., & Spinks, J. A. (1992). Orienting, habituation, and the allocation of processing resources. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (pp. 227-262). Hillsdale, NJ: Erlbaum.
- Sokolov, E.N. (1963). *Perception and the conditioned reflex*. New York, NY: Macmillan.
- Swan, J. A., & Pearce, J. M. (1988). The orienting response as an index of stimulus associability in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 292-301.
- Thompson, R. F. (2009). Habituation: A history. *Neurobiology of Learning and Memory*, *92*, 127-134.
- Tomsic, D., Massoni, V., & Maldonado, H. (1993). Habituation to a danger stimulus in two semiterrestrial crabs: Ontogenetic, ecological and opioid modulation correlates. *Journal of Comparative Physiology A*, *173*, 621-633.
- Tomsic, D., Pedreira, M. A., Romano, A., Hermitte, G., & Maldonado, H. (1998). Context-US association as a determinant of long-term habituation in the crab *Chasmagnathus*. *Animal Learning & Behavior*, *26*, 196-209.

- Turatto, M., Bonetti, F., & Pascucci, D. (2018). Filtering visual onsets via habituation: A context-specific long-term memory of irrelevant stimuli. *Psychonomic Bulletin & Review*, *25*, 1028-1034.
- Vogel, E. H., Ponce, F. P., & Wagner, A. R. (2019). The development and present status of the SOP model of associative learning. *Quarterly Journal of Experimental Psychology*, *72*, 346-374.
- Wagner, A. R. (1976). Priming in STM: An information-processing mechanism for self-generated or retrieval-generated depression in performance. In T. J. Tighe & R. N. Leaton (Eds.), *Habituation: Perspectives from child development, animal behavior, and neuropsychology* (pp. 95-128). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wagner, A. R. (1979). Habituation and memory. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation* (pp. 53-82). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wagner, A. R., & Brandon, S. E. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In S.B. Klein & R.R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the scope of traditional learning theory* (pp. 149-190). Hillsdale, NJ: Erlbaum.
- Wagner, A. R., Rudy, J. W., & Whitlow, J. W. (1973). Rehearsal in animal conditioning. *Journal of Experimental Psychology*, *97*, 407-426.
- Wagner, A. R., & Vogel, E. H. (2010). Associative modulation of US processing: Implications for understanding of habituation. In N. Schmajuk (Ed.),

*Computational models of conditioning* (pp. 150-185). Cambridge: Cambridge University Press.

Westbrook, R. F., & Bouton, M. E. (2010). Latent inhibition and extinction: Their signature phenomena and the role of prediction error. In R. E. Lubow & I. Weiner (Eds.), *Latent inhibition: Cognition, neuroscience and applications to schizophrenia* (pp. 23-39). Cambridge: Cambridge University Press.

Westbrook, R. F., Jones, M. L., Bailey, G. K., & Harris, J. A. (2000). Contextual control over conditioned responding in a latent inhibition paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 157–173.

Wilson, P. N., Boumphrey, P., & Pearce, J. M. (1992). Restoration of the orienting response to a light by a change in predictive accuracy. *Quarterly Journal of Experimental Psychology*, 44B, 17-36.

Young, S. L., & Fanselow, M. S. (1992). Associative regulation of Pavlovian fear conditioning: Unconditional stimulus intensity, incentive shifts, and latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, 18, 400–413.



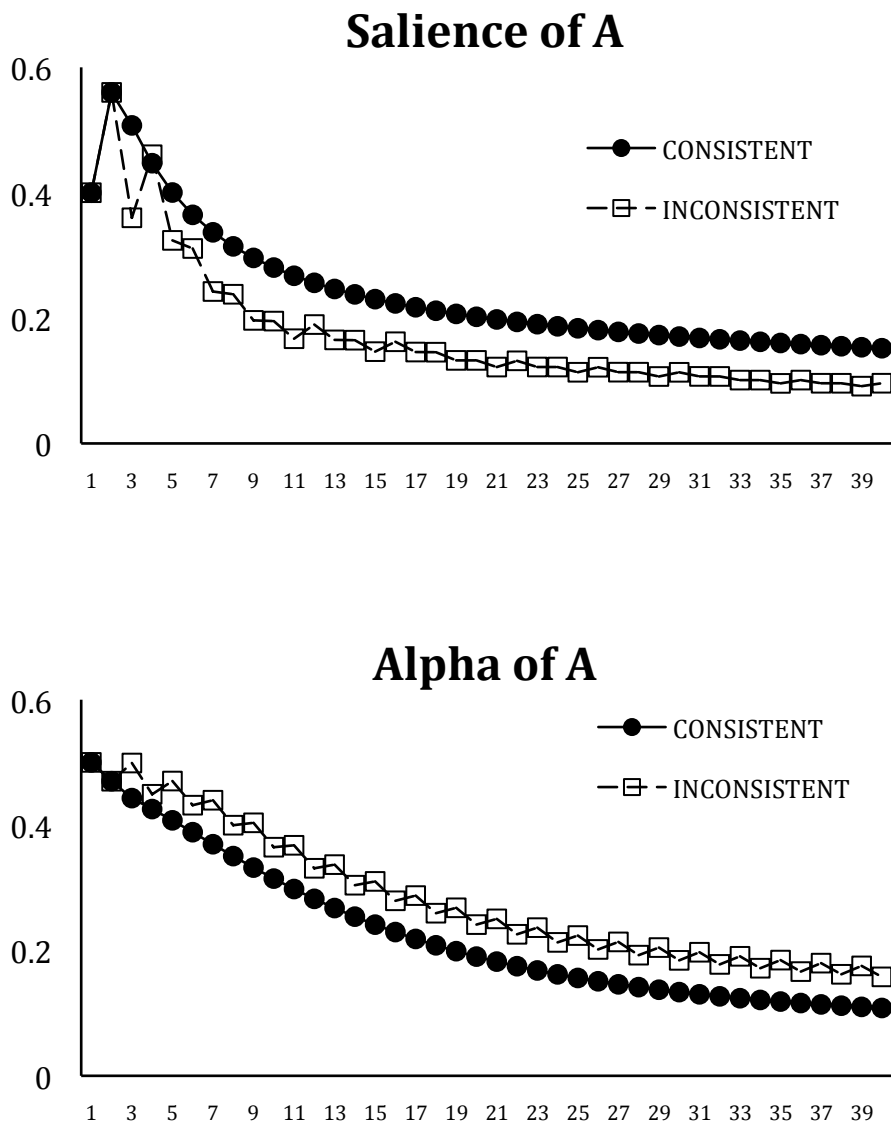


Figure 1. Simulations, using the Hall-Rodríguez (2010a, 2019) model, of changes in the salience ( $S$ ) and associability ( $\alpha$ ) of stimulus A over a series of trials in which A is consistently followed by another, salient, stimulus (Consistent condition), or is followed by this stimulus on 50% of trials.

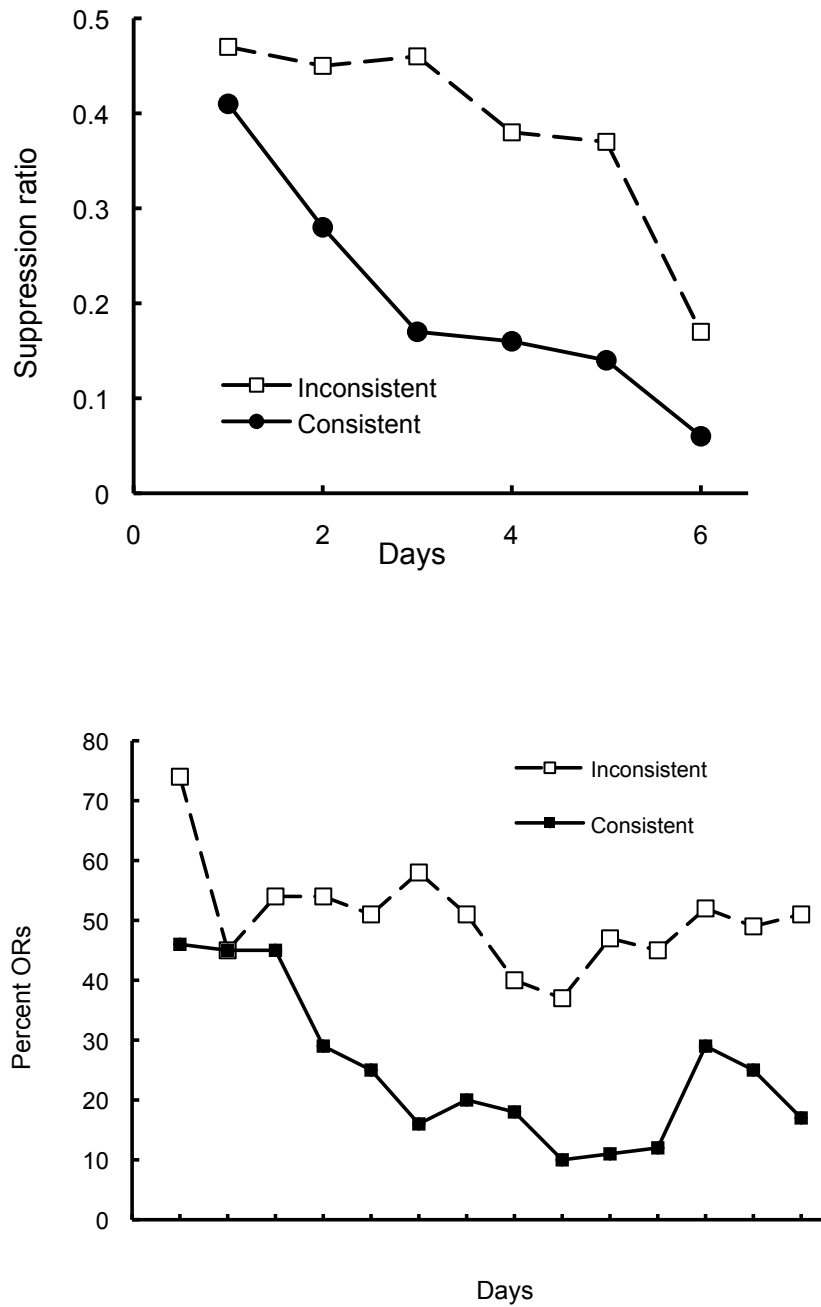


Figure 2. Upper panel: Acquisition of conditioned suppression with a shock US for rats given prior exposure to the shock in which the shock was followed on all trials by a loud noise (Consistent condition), or on 50% of trials (Inconsistent condition). Data from Hancock (2007). Lower panel: Orienting responses (ORs) in rats to a light followed by food on all trials (Consistent condition) or on a random 50% of trials (inconsistent condition). Data from Kaye and Pearce (1984).