

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/87789/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Dwyer, Dominic M. and Waldmann, Michael R. 2016. Beyond the information (not) given: representations of stimulus absence in rats (*Rattus norvegicus*). *Journal of Comparative Psychology* 130 (3) , pp. 192-204. 10.1037/a0039733 file

Publishers page: <http://dx.doi.org/10.1037/a0039733> <<http://dx.doi.org/10.1037/a0039733>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 Beyond the information (not) given: Representations of stimulus absence in rats (*Rattus*
2 *norvegicus*).

3
4 Running header - Uncertainty & associations in rats

5
6 Dominic M. Dwyer (Cardiff University & University of New South Wales)

7 and

8 Michael R. Waldmann (University of Göttingen)

9
10 July 2015 – Accepted for publication in the Journal of Comparative Psychology

11
12 NOTE: This is a copy of the authors' pre-print manuscript prior to acceptance for publication
13 in The Journal of Comparative Psychology. This article may not exactly replicate the final
14 version published in The Journal of Comparative Psychology. It is not the copy of record.
15 The final version is copyright to The Journal of Comparative Psychology, this version can be
16 found at: <http://dx.doi.org/10.1037/a0039733>

17
18
19 For correspondence: DwyerDM@cardiff.ac.uk or michael.waldmann@bio.uni-goettingen.de

20

21

22

23

Abstract

24 Questions regarding the nature of non-human cognition continue to be of great interest within
25 cognitive science and biology. However, progress in characterising the relative contribution
26 of “simple” associative and more “complex” reasoning mechanisms has been painfully slow
27 – something that the tendency for researchers from different intellectual traditions to work
28 separately has only exacerbated. This paper re-examines evidence that rats respond
29 differently to the non-presentation of an event than they do if the physical location of that
30 event is covered. One class of explanation for the sensitivity to different types of event
31 absence is that rats' representations go beyond their immediate sensory experience and that
32 covering creates uncertainty regarding the status of an event (thus impacting on the
33 underlying causal model of the relationship between events). A second class of explanation,
34 which includes associative mechanisms, assumes that rats represent only their direct sensory
35 experience and that particular features of the covering procedures provide incidental cues that
36 elicit the observed behaviours. We outline a set of consensus predictions from these two
37 classes of explanation focusing on the potential importance of uncertainty about the
38 presentation of an outcome. The example of covering the food-magazine during the
39 extinction of appetitive conditioning is used as a test-case for the derivation of diagnostic
40 tests that are not biased by preconceived assumptions about the nature of animal cognition.

41

42 Keywords: Causal model, renewal, secondary reinforcement, ambiguity

43

44 “And no man, when he hath lighted a lamp, covereth it with a vessel, or putteth it under a
45 bed: But he putteth it on a stand.” *Luke, Ch. 8, V 16.*

46

47 *Putting lamps under bushels*

48 While a lamp under a bushel casts just as little light as an unlit lamp, the status of the
49 unlit lamp is clear, while that of the covered lamp is uncertain – it may be lit or unlit.

50 Although probably not the typical message taken from this parable, it exemplifies the fact
51 that, considered rationally, there is a clear difference between the absence of an event, and the
52 absence of information about that event. One goal of the present article is to examine recent
53 research on the capacities of rats to reason about hidden objects as a test case for examining
54 distinctions between higher-level cognitive processes and basic associative mechanisms. But
55 before turning our attention to these empirical concerns we will comment, relatively briefly,
56 on the sometimes rancorous debate concerning the commonalities and differences between
57 human and non-human animal cognition.

58 Comparisons between human and non-human animal cognition have attracted great
59 interest in cognitive science and biology in the past decades. Perhaps the dominant tradition
60 has been to assume that non-human animals are convenient systems in which to study simple
61 processes (e.g. of learning and memory), and their underlying biological substrates,
62 untrammelled by the more complex reasoning and rule-based processes possessed by
63 humans. This view has been challenged by recent evidence which suggests that animals
64 might, in addition to simple associative processes, also have far richer ways of representing
65 the causal texture of their environment (e.g., Blaisdell, Sawa, Leising, & Waldmann, 2006;
66 Fast & Blaisdell, 2011; Leising, Wong, Waldmann, & Blaisdell, 2008; Murphy, Mondragon,
67 & Murphy, 2008; Waldmann, Schmid, Wong, & Blaisdell, 2012). However, the potentially
68 far-reaching implications of these studies depend on the idea that behaviours consistent with
69 complex cognitive mechanisms are indeed the result of such complex mechanisms, and

70 cannot be explained as emergent properties of more simple (in particular associative)
71 mechanisms (Burgess, Dwyer, & Honey, 2012; Dwyer, Starns, & Honey, 2009; Kutlu &
72 Schmajuk, 2012). A fundamental shortcoming of this debate is that it is not entirely clear
73 how higher-level cognitive processes can theoretically and empirically be distinguished from
74 basic associative mechanisms. We present here a new proposal for making this distinction.

75 In the literature, different proposals have been discussed on how to distinguish higher-
76 level cognition from associative processes. The traditional view, inspired by behaviourism,
77 was that cognitive but not associative theories postulate information processing mechanisms
78 operating on mental representations of the world. This distinction is no longer pertinent
79 because many modern associative theories assume that animals possess mental
80 representations, and characterise learning as the formation of associative links between these
81 representations. A prime example of this is the idea that classical conditioning reflects the
82 formation of an excitatory association between mental representations of a conditioned
83 stimulus (CS) and an unconditioned stimulus (US) – an idea included in essentially all
84 accounts of associative learning regardless of their differences concerning the details of the
85 learning algorithm involved (e.g., Esber & Haselgrove, 2011; Harris, 2006; Le Pelley, 2004;
86 Mackintosh, 1975; Pearce, 2002; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner,
87 1981). While contemporary associative theory does include (and require) mental
88 representations, it should be recognised that these are informationally “thin” representations,
89 held to consist essentially as copies or traces of aspects of the sensory and motivational
90 stimulation produced by experience of the stimulus (Heyes, 2012). In particular, associative
91 theories do not allow that either their representations or the links between them have semantic
92 content – that is their truth value cannot be assessed. In this sense “thick” representations are
93 effectively propositional (i.e. they can be expressed as a statement with a truth value – e.g.
94 “The light is on” – which is either true or false, and also allows the possibility “I don’t

95 know”). In contrast, as a copy or trace of the activation produced by the stimulus, thin
96 representations accord to nothing more than the set of nodes/elements that are activated by
97 experience with the stimulus (or activated through associative links). Therefore, it makes no
98 sense to ask whether the activation is “correct”, it is merely a matter of whether activation
99 exists and to what degree. Although the fact that contemporary associative theory admits
100 mental representations at all removes one classical divide between associative processes and
101 complex cognition, the commitment to thin mental representations has one critical
102 consequence: It requires associative theory to deal only with the sample of events
103 experienced by an organism and the activation of the representations that occur as a result of
104 this experience.

105

106 Levels of Representation

107 Our main focus in this article is on causal representations. Predicting and explaining
108 events on the basis of observations and interventions is arguably one of the most important
109 cognitive competencies that allow organisms to adapt to the world. There are a vast number
110 of competing theories specifying the cognitive mechanisms underlying this competency. As
111 a first approximation, we would like to propose two different classes of theories that can be
112 distinguished on the basis of the postulated representations of the world. Of course, within
113 each class there are numerous competing variations that have been the focus of extensive
114 research.

115 Level 1: Sample-based theories:

116 The basic assumption underlying this class of theories is that causal representations
117 use representations of temporally ordered observed events (cues, outcomes) and that the goal
118 of learning is to capture the statistical relations between these events. Thus, the key
119 assumption for our purposes is that Level 1 accounts assume that organisms do not (or

120 cannot) look beyond the observed sample of events. The sample of learning events is what
121 organisms know about the particular aspect of the world they observe.

122 One of the key topics within this class of theories is to investigate which statistical
123 rules organisms actually use to represent the observed covariations. A large number of such
124 rules have been proposed both within cognitive theories (e.g., Hattori & Oaksford, 2007;
125 Perales & Shanks, 2007) and within associative theories (e.g., Dickinson, 2001; Le Pelley,
126 Oakeshott, Wills, & McLaren, 2005; Shanks & Dickinson, 1987). One thing all these
127 otherwise competing theories have in common is that they compute some index of
128 covariation from the learning sample, which encapsulates the effective strength of the causal
129 relation. Indeed, the fact that some associative and cognitive models make identical
130 predictions under some circumstance – see for example relationship between the output of the
131 Rescorla-Wagner model and delta-P metric discussed by Shanks (1995) – implies that these
132 models often capture the same functional relationships between experienced events
133 perspective (for a more detailed analysis of the implications of examining learning at a
134 functional level see De Houwer, Barnes-Holmes, & Moors, 2013; De Houwer et al., this
135 volume). In the present context, it is most important that such theories do not include a role
136 for any awareness about the fallibility of experiences of the world (e.g., absence of evidence)
137 or of the representations themselves (e.g., dreams, hallucinations vs. experiences of real
138 events). The fact that many associative models are based around error-correction
139 mechanisms does mean that they calculate a prediction error between the associative
140 activation of representational nodes and the activation produced by experience of events.
141 However, this is an algorithmic comparison and does not require the organism to have a
142 meta-representational appreciation of the current internal associative model, the current
143 external input, and the relationship between them. In short, sample-based theories do not

144 assume a meta-representational understanding by the organism of the distinction between its
145 representation of the world and the world that produces that representation.

146 Various research paradigms view human and non-human organisms as focusing on
147 samples, unable to go beyond the information given. In causal research, associative theories
148 are a prime example of this class of theories. Indeed, the fact that associative theories are
149 characterised by a reliance on thin mental representations of stimuli and the links between
150 them requires that they must focus on an organism's sample of experience. Thin
151 representations do not allow an assessment of truth value, so there is no way in which the
152 mental representation activated by a stimulus (or its activation through memory or associative
153 means) can be evaluated as accurately corresponding to the outside world or not¹. Moreover,
154 thin representations ascribe no content to an associative link other than as a means for
155 specifying the degree to which activity of one representation will influence the degree of
156 activation in a representation to which it is associatively linked. As such associative accounts
157 do not explicitly distinguish between causal and non-causal relationships between events.

158 According to this sample-based class of theories, organisms encode the presence and
159 absence of temporally ordered events and learn statistical covariations between these events.
160 The strength of these covariations determines inferences or behaviour. Rule-based theories of

¹ It is instructive to note here Holland's (1990) work showing that stimulus representations activated associatively ("images" in his terminology) can elicit some of the same processing that occurs when the stimulus itself is presented. The same body of work also established that the processing of retrieved images is not exactly the same as that for experienced events – so there is clearly some distinction between retrieved and directly activated stimulus representations. However, when only thin representations are assumed then this distinction in what is activated by experience (the world) and through association (the image) is literally just that, a difference in what is activated – only from the outside can the different sets of activated elements be related to which set accords to the real world. As we will see later, recent model-based accounts are very different in assuming that there is some ability to distinguish the model from the experience.

161 causal reasoning are another example (for a review, see, Waldmann & Hagmayer, 2013).
162 These theories debate which exact covariation rule organisms employ. But as in the
163 associative framework, statistical covariations are based on what is observed in a sample. In
164 social psychology, there is also a variant of the sample view (see, Fiedler, 2012; Fiedler &
165 Juslin, 2006). Here the claim is that judgmental biases are often caused by distortions in the
166 observed or retrieved sample of experiences. Fiedler (2012) argues that humans are largely
167 unable to understand and correct statistical distortions in the sample. He has labelled this
168 deficit “metacognitive myopia.”

169 Level 2: Causal Models:

170 This class of theories assumes that organisms go beyond the information given when
171 learning about causal relations to make inferences about an underlying unobservable causal
172 model (see Waldmann, Hagmayer, & Blaisdell, 2006). Of course, going beyond the sample
173 is not an all-or-none feature. There are different degrees of inferences transcending the
174 sample, and different organisms may differ in the extent to which they are capable of going
175 beyond the information given (for an example within causal model theory, see Waldmann,
176 Cheng, Hagmayer, & Blaisdell, 2008).

177 A key difference between causal and associative theories concerns the links between
178 causes and effects. Causal links, often depicted as arrows, are directed from cause to effect.
179 In associative theories, temporal order determines whether an association is excitatory or
180 inhibitory, but this alone does not result in the explicit representation that the first event
181 caused the second. Indeed, causal and temporal order can be dissociated (e.g., Waldmann,
182 2000; Waldmann & Holyoak, 1992). For example, physicians often observe the symptoms
183 (i.e., effects) prior to diagnosing the cause. The exact meaning of the causal arrows differs
184 across theories, but the general assumption is that causal processes are unobservable and need
185 to be inferred based on observations and prior knowledge. For example, Cheng’s (1997)

186 power PC theory assumes that people are capable of inferring the power of a cause based on
187 covariation and background assumptions. Power is a point estimate of the unobservable
188 probability of the cause generating or preventing a specific effect in the hypothetical absence
189 of background factors.

190 A less abstract account assumes hidden forces and causal mechanisms that transfer
191 some kind of conserved quantity (such as linear momentum or electric charge to take
192 examples from physics) between causes and effects (see Waldmann & Hagmayer, 2013, for a
193 review). Although causal mechanisms can sometimes be elaborated as chains of observable
194 variables, the variables within the chain are connected via arrows that code some kind of
195 hidden flow of a conserved quantity (Dowe, 2000). Mechanism theories do not necessarily
196 assume elaborate knowledge, as it is well known that human laypeople often have no or only
197 very sketchy knowledge of the exact relationships between events (Rozenblit & Keil, 2002).
198 The assumption rather is that people understand a relation between two events as causal if
199 they assume that there is some kind of mechanism that links the events, even if the details of
200 this mechanism are largely unknown.

201 A more recent development in causal model theory goes one step further in separating
202 observed samples from underlying unobservable generating models. Inspired by Bayesian
203 statistical inference, it is assumed that a rational approach to causal inference would require
204 taking into account the fact that samples are noisy reflections of the hidden generating causal
205 models. Thus, depending on statistically relevant factors, such as sample size, samples carry
206 more or less *uncertainty* about the structure and the parameters of the causal model.

207 According to this view, organisms are mainly interested in a faithful representation of the
208 characteristics of the causal model, and therefore need to take into account uncertainty when
209 making inferences. A number of studies have demonstrated that human subjects are indeed

210 sensitive to statistical uncertainty (Griffiths & Tenenbaum, 2009; Lu, Yuille, Liljeholm,
 211 Cheng, & Holyoak, 2008; Meder, Mayrhofer, & Waldmann, 2014)².

212

213 *Testing the Level of Representation*

214 Level 1 associative and Level 2 causal model theories are often pursued in separation.

215 A typical research strategy of those interested in either class of account is to design studies
 216 that test between competing theories within their class – while questions of between-class
 217 comparisons tend to be considered most seriously only after publication when conclusions are
 218 challenged externally. For example, it is not uncommon for alternative associative Level 1
 219 “killjoy” (Shettleworth, 2010) accounts to be developed in a post-hoc fashion after novel
 220 patterns of behaviour had been discovered based on predictions of Level 2 theories. In this
 221 light it is rather unsurprising that progress in this area often appears meagre: if for nothing
 222 else than publication lag “conversations” in the literature are incredibly slow. In addition
 223 there is often a strong bias for Level 2 theorists to interpret data that is consistent with
 224 predictions of their complex accounts as evidence for their theory without considering the
 225 possibility that level 1 accounts of the same data might be available (this is especially
 226 prevalent when human subjects are involved). When alternative Level 1 accounts are
 227 considered, this consideration is often constrained by a lack of familiarity with contemporary
 228 associative theory. On the other hand, the emergent properties of Level 1 theories are not
 229 always apparent without considering the exact experimental situation and by themselves

² The nomenclature we have adopted (Level 1 vs Level 2) is entirely abstract and we admit that this may appear uninformative, but the choice was quite deliberate. While we focus here on the nature of the representations assumed at each level and the differences in terms of the explicit role of causal relationships, the distinction between these two classes of model goes beyond causality (as our subsequent discussion of theory of mind illustrates). Thus the abstract nomenclature avoids overly-restrictive characterisations of the model classes we are discussing.

230 Level 1 theories commonly provide little guide to the investigation of the sort of phenomena
231 predicted by Level 2 theories. For example, it was only after Couchman, Coutinho, Beran,
232 and Smith (2010) published their analysis of delayed feedback as supporting a (Level 2)
233 metacognition account of primate behaviour in a discrimination task that Le Pelley (2012)
234 was able to simulate their experimental procedures with a (Level 1) reinforcement learning
235 account. Similarly, the demonstration that rats' behaviour can diverge as a function of
236 whether a cue appears as a result of their actions or not followed from the prediction from a
237 (Level 2) causal model account suggesting a critical difference between seeing and doing
238 (Blaisdell et al., 2006). Only following the publication of the experimental methods used to
239 produce this demonstration could Kutlu and Schmajuk (2012) examine the possibility that
240 their associative model might be able to simulate the observed behaviour³. Thus, Level 1
241 theorists often need to await progress within Level 2 theories before they can address the
242 question of whether the discovered phenomena genuinely require complex representations or
243 can also be explained by a Level 1 account. One possible response to these systemic
244 problems is the direct collaboration between researchers from different theoretical
245 perspectives.

246 Of course, developing an alternative Level 1 account for a phenomenon generated by
247 Level 2 research is only the first step. Although considerations of simplicity enshrined in
248 Morgan's Canon (Morgan, 1894) have often led researchers, at least from the associative
249 camp, to favour Level 1 over Level 2 theories, it should be remembered that the Canon is (at
250 best) a guide to interpretation and does not have any logically probative status (for a more
251 detailed discussion of this point, see Heyes, 2012). Indeed, any heuristic arguments that

³ This far from a one-way relationship as demonstrated by the example of Bayesian reasoning accounts (e.g., Gopnik et al., 2004; Griffiths & Tenenbaum, 2009) developed to explain cue-competition effects such as backward blocking that were first reported in the associative literature.

252 might be applied – from considerations of parsimony to appeals to predictive or explanatory
253 scope – cannot on their own conclusively decide between Level 1 and Level 2 accounts. As
254 ever in science, empirical data are paramount, and thus the most productive research strategy
255 is to develop competing Level 1 and Level 2 accounts of a phenomenon and then deploy
256 experimental paradigms that allow differentiation between them.

257 But before moving to consider a test case for a targeted empirical comparison of
258 Level 1 and Level 2 theories, we should emphasise that they are not necessarily mutually
259 exclusive. In cognitive psychology, two-process theories (see, Evans, 2012) have become
260 increasingly popular. One example, related to our target phenomenon, is the two-process
261 model of theory of mind inferences by Apperly and Butterfill (2009). A typical task in this
262 domain is the Sally scenario in which the protagonist Sally hides an object, which in her
263 absence is transferred to a different location. The key finding is that children younger than 4
264 seem unable to understand that Sally will look at the place she has hidden the object
265 regardless of the current location. When asked where she will go, young children tend to
266 point to the actual location of the object. Fully understanding this situation requires the
267 competency to have meta-representations that separate reality from (possibly erroneous)
268 mental representations. Many researchers argued that young children as well as animals lack
269 such meta-representational capacities. In the last decade, however, researchers using more
270 implicit habituation paradigms have demonstrated some level of understanding of this task
271 even in infants (Onishi & Baillargeon, 2005). Apperly and Butterfill therefore postulate two
272 separate processes that may underlie the responses in the different tasks. Whereas infants
273 may only understand that agents look for something where they have seen it last, older
274 children may reason with more complex meta-representations, which in the beginning stages
275 of reasoning leads to the observed errors. According to the two-process view, some species
276 may only be capable of reasoning with the simpler process, whereas others may have both

277 types of processes at their disposal. Critically however, even for these sort of two-process
278 accounts, the question remains as to whether a particular behaviour is (or can be) supported
279 by the simpler process or only the more complex one. So the importance of determining the
280 representational level at which an organism is functioning remains germane even from the
281 perspective of dual-process accounts.

282

283 *Hidden Events: A Simple Test Case for Sensitivity to Uncertainty*

284 The present article will discuss a fairly simple potential indicator of uncertainty,
285 uncertainty about the status of events. Level 2 causal model accounts would differentiate
286 between two possible causes for the failure to experience an expected event: Either the event
287 is really absent in the world, or the event is present but access to it is being prevented in some
288 fashion. Waldmann et al. (2012) examined a test-case for this possibility in the extinction of
289 Pavlovian appetitive conditioning. In their experiments, rats were presented with three
290 learning and test phases. In Phase 1, an association between a cue (CS), a light, and sucrose
291 (US) was established through a Pavlovian conditioning procedure (a 10s light was presented
292 and the offset of the light followed by 10s access to a sucrose-filled dipper)⁴. In Phase 2, the
293 extinction phase, the cue was paired with the experience of absence of sucrose (the light was
294 presented in advance of the empty dipper – i.e. the dipper arm was raised for 10s, but the
295 trough did not contain sucrose, so no primary reward was presented). Then in Phase 3, the
296 degree of extinction was tested by presenting the light cue without sucrose (again, the empty
297 dipper continued to be presented). The crucial manipulation involved Phase 2. In one

⁴ The food magazine was positioned above a trough containing sucrose solution. A mechanical dipper arm, with a small cup on the end, was immersed in this solution. Sucrose access was provided by raising the arm so that the cup protruded through a hole in the base of the food magazine for 10s before being lowered again. The rats could not access either the dipper arm or the sucrose except when it was raised.

298 condition, the No-Cover condition, rats could directly observe that sucrose was actually
299 absent from the food magazine, whereas in the alternative Cover condition a metallic plate
300 was placed over the magazine preventing rats from accessing it. The test phase showed that
301 rats differentiated between these conditions with greater test phase responding to the CS in
302 the Cover than the No-Cover condition. Moreover, it was not merely the presence of the
303 metallic plate that controlled responding, because a control condition where the plate was
304 included without preventing access to the food magazine did not prevent extinction.

305 As noted above, the causal model account would interpret this finding as evidence
306 that rats are capable of differentiating between two possible causes of the absence of sucrose
307 in the extinction phase: Either the sucrose is really absent, or it is present but access is
308 blocked. This inference requires an understanding of uncertainty of the status of events. In
309 other words, initial training experience should create a *light causes sucrose* model. The
310 transition from the rewarded training phase to the non-rewarded extinction phase could
311 potentially create an ambiguity in a causal understanding of the situation – has the causal
312 relationship changed, and the light no longer causes sucrose to appear, or is the relationship
313 still intact but the sucrose has for some other reason not been observed? This ambiguity
314 would be emphasised when access to the usual source of sucrose delivery was prevented
315 during extinction – although the light is still experienced without sucrose, both possible
316 causal structures are still consistent with the experience because there is no direct
317 disconfirmation of the expected sucrose delivery. Thus a causal model analysis would
318 suggest that covering the sucrose magazine should attenuate the effects of extinction and help
319 preserve the *light causes sucrose* model. In turn, preserving a causal relationship between the
320 light and sucrose should result in higher responding in the test phase - which is exactly what
321 happened (Waldmann et al., 2012). Clearly, a full causal understanding of this situation
322 requires some kind of understanding of the difference between the representations of the

323 world and the actual world. Even in humans, unless people have philosophical training, this
324 differentiation is unlikely to be explicitly available. It suffices that in specific cases absence
325 is distinguished from lack of evidence.

326 Functionally the separation between experience and world has a number of potential
327 advantages for organisms. If experience and the world were collapsed, every instance of
328 disappearance due to another object blocking sight would lead to a fading of the
329 representation of the object although it is still present behind the occluder. Since such
330 experiences are common, the physical representation of the world arising from such
331 inferences would be very different from ours. Work on object permanence with animals
332 seems to indicate that many animals may not think that objects behind an occluder actually
333 disappear from the world (Gómez, 2004, 2005). Similarly, in Waldmann et al.'s (2012) study
334 organisms that only represent present and absent events and do not differentiate between
335 absence in the world and lack of evidence would represent events in Phase 2 (extinction) as a
336 gradual change of contingency. Although this is certainly a possibility, as the No-Cover
337 condition demonstrates, it is not necessarily adaptive to always make this inference. One key
338 feature of causal relations is that they tend to be stable and do not suddenly change (Pearl,
339 2000). Thus, the capacity to distinguish between different causes of experienced absence is
340 potentially adaptive for an organism that has the goal of forming veridical representations of
341 the causal texture of the world and if these veridical representations improve the organism's
342 success in interacting with the world.

343

344 *Associative Accounts of Hidden Events: Renewal and Secondary Reinforcers*

345 As described above, a causal model account based on uncertainty can explain why
346 covering the food magazine during extinction might result in higher levels of responding
347 during test. However, the details of the experiments performed also admit alternative

348 explanations of the same results based entirely on associative Level 1 mechanisms: We will
349 consider one based on response prevention⁵, a second based on renewal theory, and another
350 on a consideration of conditioned reinforcement.

351 Rescorla (2001) notes that there is typically a direct relationship between the amount
352 of non-reinforced responding in extinction and the degree to which such non-reinforcement
353 impacts on future behaviour. For example, following tone-food pairings, presentation of the
354 tone alone will typically result in some degree of responding to the food magazine during an
355 extinction phase, while devaluation of the food reward or satiating the animals reduces the
356 level of extinction phase magazine responding. Even though the number of unrewarded tone
357 alone presentations is unaffected by devaluation or satiation, these treatments which reduce
358 extinction phase magazine responding also reduce the effectiveness of extinction (Holland &
359 Rescorla, 1975). On the basis of such results, Rescorla (2001; see also Colwill, 1991)
360 suggested that learning not to make a particular response may make a critical contribution to
361 the decrement in responding typically observed in extinction. One direct corollary of this
362 idea is that the effects of non-reward in extinction will be reduced if the original response is
363 not produced. In the present circumstances, covering the magazine clearly prevents the target
364 response of magazine entry, and thus prevention of this response should protect it from
365 extinction. Not only does this provide a simple explanation of why test phase responding was
366 be higher after the magazine was covered in the extinction phase, it also explains why
367 introducing a similar metallic cover that did not prevent access to the magazine had little
368 effect.

369 A second associative account of the effects of the magazine cover comes from
370 renewal theory. This approach suggests that extinction should be specific to the context in

⁵ We would thank one of the reviewers of an earlier version of this paper for their suggestion of this possibility.

371 which it occurs, and that extinguished responses should reappear when testing occurs in a
372 situation more akin to the original training context than to the context of extinction (e.g.,
373 Bouton, 2004; Delamater, 2004). In the current situation, the cover provided during
374 extinction could act as a context change, so its removal would comprise a return to the
375 original training context, thus supporting the re-emergence of responding. Thus, according to
376 this view rats would gradually start to represent Phase 2 as a situation in which the light is
377 paired with the absence of sucrose, but expression of this new association would be restricted
378 to the context in which extinction took place. This possibility was acknowledged in the
379 original report of these experiments, and in Experiment 3 of that paper an additional control
380 group was used in which the metal “cover” was inserted into the apparatus during the
381 extinction phase, but did not actually prevent access to the food magazine. This control, in
382 which the presence or absence of a cover could have acted as a cue separating the extinction
383 and test contexts, resulted in performance that was no different to that in the No-Cover
384 condition. However, it may be argued that a cover preventing access to a source of food is
385 more salient than a cover placed elsewhere, in which case a magazine cover would be a more
386 effective contextual cue than one that does not cover the magazine.

387 It should be noted that in all the Cover conditions the sucrose dipper continued to be
388 raised and lowered, but that there was “no noticeable vibrations for the human ear” (p. 983,
389 Waldmann et al., 2012), that could be discerned inside the experimental chamber. That is,
390 covering was assumed to have prevented all access to information about the operation of the
391 dipper during extinction⁶. Thus in the covering situation, the training and test contexts were

⁶ It should be noted that this assumption was not directly tested, and given that rat and human sensory abilities are somewhat different then it is certainly plausible that the rats in Waldmann et al.’s (2012) experiments were able to sense some aspect(s) of the dipper’s operation behind the cover. Although this possibility has no direct impact on the ideas discussed here, it does raise the issue of what predictions the different accounts of the

392 similar in the operation of the dipper but diverged from the extinction context in both respects
393 – while in the No-Cover, and the plate without covering conditions, the extinction and test
394 contexts both included the operation of an empty dipper. In short, covering the magazine in
395 the extinction phase of the experiments produced several potential cues that could have
396 differentiated the extinction and test contexts. This could support the recovery of
397 extinguished responding in the covered condition without reference to any Level 2
398 mechanisms.

399 The final alternative account of the covering data we will consider here relies on
400 secondary reinforcement. Remembering that the training phase of these experiments was
401 based on pairing the light with a sucrose filled dipper, the training phase should establish
402 light-sucrose, light-dipper, and dipper-sucrose associations. It is well known that animals
403 will respond both to cues paired with primary reinforcers - i.e. the sucrose in these studies -
404 and also secondary reinforcers - i.e. any stimulus that is associated with a primary reinforcer
405 (for reviews see, Mackintosh, 1974; Mackintosh, 1983). In these studies the dipper would
406 have accrued secondary reinforcing properties by being paired with sucrose during the
407 training phase. Following this, all groups received light-alone presentations in the extinction
408 phase - presumably extinguishing light-sucrose associations to a similar extent between
409 groups. In the No-Cover condition the empty dipper would also be experienced – resulting in
410 the extinction of the dipper-sucrose associations, and thus the removal of secondary
411 reinforcing properties of the dipper. However, in the Cover condition, the dipper would not
412 be experienced at all during the extinction phase, which would protect the dipper-sucrose
413 associations and preserve the conditioned reinforcement properties of the dipper. In turn, this
414 would allow the dipper to support responding to the light when the light was again paired

covering effect might make regarding “partial” covers (e.g. explicitly preventing vision but not audition).

415 with the dipper in the test phase. In short, the training phase paired the light cue with both a
416 primary (sucrose) and a secondary (the sucrose-paired dipper) reinforcer. Covering the
417 magazine in the extinction phase of the experiments could preserve the secondary reward
418 properties of the dipper compared to the uncovered conditions. The secondary reinforcing
419 properties of the dipper could support additional test-phase responding in the covered
420 condition without reference to any Level 2 mechanisms.

421

422 *Divergent predictions from Level 1 and Level 2 accounts of hidden events*

423 One important feature of the causal uncertainty and renewal/secondary reinforcement
424 accounts of the effects of covering the magazine is that the differences between them relate
425 directly to the nature of the division between Level 1 and Level 2 theories outlined
426 previously. The causal model account suggests that uncertainty produced by the cover would
427 preserve the strength of a *light causes sucrose* model in the face of experiencing the light
428 without sucrose. This goes beyond the direct sample of experience because the fact that
429 sucrose did not follow the light is discounted due to a distinction between absence of sucrose
430 (the No-Cover case) and absence of evidence (the Cover case). That is, the effects seen in the
431 test phase are a product of covering producing uncertainty over whether the sucrose did or
432 did not occur, and thus reducing the effective level of extinction. In contrast, the three
433 associative accounts considered here all related to direct effects of the cover in extinction or
434 its removal at test. The response-prevention account suggests that covering reduces the
435 effects of extinction because the target response could never be produced when the magazine
436 was covered. Both the renewal and secondary reinforcement accounts assume that extinction
437 does occur due to experience of the light without sucrose, but that responding returns in the
438 test phase due to events that happen during that test: For renewal theory, the critical event in
439 the Cover condition is that the context of test is different from that of extinction (it allows

440 access to the magazine and includes an operating dipper – as in training but not extinction);
441 For secondary reinforcement, the critical event is that the rats experience the light paired with
442 the dipper, and in the Cover condition the dipper will be a secondary reinforcer (but not in the
443 No-Cover condition, because then the previous experience of the empty dipper has removed
444 the secondary reinforcing properties of the dipper) – these test phase light-dipper pairings
445 support the re-acquisition of responding to the light. That is, the associative accounts are
446 sample-based as they refer only to events that are actually experienced (or not experienced, in
447 the case of prevented responses). Therefore, empirical tests of the divergence between these
448 accounts speak not only to the particular details of each of them, but also to the more general
449 division between Level 1 and Level 2 processes in the context of this behavioural procedure⁷.

450 Effects of manipulating dipper presentation:

451 Given that the status of the dipper in the extinction and test phases is critical to two of
452 the Level 1 sample-based accounts, while uncertainty concerning the presence of reward is
453 central to the Level 2 causal model account, one empirical test would be to manipulate the
454 presence of the dipper during these phases. That is, to compare the pattern of responses
455 between groups that receive either: (A) training and testing as in the original paper with the
456 empty dipper presented during the extinction and test phases; or (B) with no presentation of
457 the empty dipper during either the extinction or test phases (i.e. the dipper would remain
458 lowered – but not be explicitly removed from the chamber). Table 1 outlines the proposed
459 experiment and summarises the key predictions of each of the accounts for responding to the
460 light at the beginning of the test phase of the experiment. The original experiments included
461 control conditions which received extinction without the magazine cover. Such controls are

⁷ Of course, it is also possible to assess how causal models might account for the direct effects of test phase events, but this would not address our current concern with whether rats are able to go beyond the sample of their experience in terms of the explicit role for uncertainty.

462 needed to establish a baseline for levels of responding after effective experimental extinction,
463 and we would propose including such uncovered controls which would receive extinction and
464 test with or without dipper presentation in the current experiment. Although it is likely that
465 the operation vs. non-operation of the dipper would influence the rate of experimental
466 extinction, we will not considered these control conditions in any detail because (as in the
467 original experiments) the extinction phase would be continued until responding to the light
468 has stopped, and so all theoretical accounts would predict negligible test phase responding.
469 The derivation of the predictions for the critical magazine cover conditions is fleshed out in
470 turn for the causal model, response prevention, renewal, and secondary reinforcement
471 accounts.

472 In both the Dipper Cover and No-Dipper Cover conditions the training phase would
473 produce a *light causes sucrose* model. In the extinction phase, the light occurs alone, but
474 because access to the magazine is blocked the *light causes sucrose* model will be protected
475 because the covering means that the status of the sucrose is uncertain and thus the evidence
476 for sucrose not appearing is partially or totally discounted in terms of relevance to the light-
477 sucrose relationship. Covering might also protect the light-sucrose causal relationship
478 because it leads to the formation of a more complex causal model whereby the light causes
479 sucrose but the action of an external event stops this being expressed (e.g. the cover stops
480 access to the delivered sucrose). In the test phase, the cover is removed – so behaviour will
481 be determined by the *light causes sucrose* model (i.e. moderate to high responding is
482 predicted). Critically, the extinction phases for the Dipper Cover and No-Dipper Cover
483 conditions are the same. In both conditions, the dipper and sucrose are covered during
484 extinction so the causal model at the start of test should be the same. In turn, this same causal
485 model predicts that the response to the light at the start of test would be the same in these two
486 conditions. Of course, as the test phase continues, then the Dipper Cover and No-Dipper

487 Cover conditions will have different experiences. Thus their causal models, and levels of
488 responding, may be expected to diverge across testing: for example, the non-operation of the
489 dipper might support the formation of a more complex causal model whereby the light causes
490 sucrose only through the action of the dipper, which for some reason did not operate (e.g. the
491 dipper was stuck). However, the dipper is operated at the end of the light during training, so
492 at the time of responding is assessed (during the presentation of the light) there is no direct
493 evidence to indicate whether or not the dipper will operate on that trial. So even if
494 responding is dependent on the expectation of dipper operation, this expectation should only
495 decline gradually as the light is encountered without the dipper following immediately
496 afterwards. Irrespective of these issues, responding early in the test phase should remain
497 diagnostic of the strength of the light-sucrose causal relationship at the end of the extinction
498 phase to the extent that causal representations are stable (Pearl, 2000).

499 The predictions of the response-prevention account are simple – in both the No-
500 Dipper Cover and Dipper Cover conditions the cover will prevent the production of magazine
501 entry responses. To the extent that extinction requires the production of the relevant
502 response, then such response prevention will attenuate the effects of extinction, and levels of
503 magazine responding to the light would be predicted to be high at the start of the test phase.

504 As outlined above, the renewal account suggests that the training phase should
505 establish an excitatory light-sucrose association, while presenting the light without the reward
506 in extinction will create an inhibitory light-“no sucrose” association. Responding at test will
507 be determined by the degree to which these two associations are expressed – something that
508 is controlled by the similarity of the extinction and test phase contexts. For the Dipper Cover
509 condition, the test phase and the extinction phase differ in two critical respects, access to the
510 magazine and the operation of the dipper: both of which are absent in the extinction phase
511 and present at test. Thus, the extinction and test contexts are quite different which will

512 attenuate the expression of the inhibitory light-“no sucrose” association formed in extinction
513 and result in responding to the light on the basis of the originally-formed excitatory light-
514 sucrose association – a classic renewal effect. In contrast, for the No-Dipper Cover
515 condition, the test phase and the extinction phase differ with respect to access to the
516 magazine, but are the same with respect to the non-operation of the dipper. Thus, while there
517 will be some difference between the extinction and test contexts in the No-Dipper Cover
518 condition, and thus some degree of renewal would be expected, this should not be as great as
519 in the Dipper Cover condition. As the non-operation of the dipper can only be observed after
520 the first trial, this difference between the Dipper and No-Dipper conditions should emerge
521 across the extinction phase.

522 Finally, the conditioned reinforcement account is based on the potential contribution
523 of the dipper as a secondary reinforcer due to its pairing with sucrose in the training phase of
524 the study. In the Dipper Cover condition, the light is presented in the absence of either the
525 primary or secondary reinforcer during the extinction phase – so by the end of extinction
526 there will be no effective source of primary or secondary reinforcement. However, the
527 secondary reinforcing properties of the dipper will be preserved through the extinction phase
528 because the dipper is never experienced without sucrose. In the test phase, the light will
529 again be presented in conjunction with the dipper, and thus the secondary reinforcing
530 properties of the dipper will support responding to the light (at least for as long as the dipper
531 remains an effective secondary reinforcer). Obviously, this secondary reinforcing effect of
532 the dipper could only be apparent after the first trial of the extinction phase. The No-Dipper
533 Cover condition will also result in the removal of any effective source of primary or
534 secondary reinforcement by the end of the extinction phase, but in this case dipper operation
535 is not reintroduced at the test phase. So test phase responding to the light will be low in this
536 condition.

537 In summary, all accounts predict that, if the dipper continues to be presented, then
538 covering the magazine in extinction will result in higher levels of test phase responding than
539 if the magazine is uncovered in extinction. Two of the associative accounts – renewal and
540 secondary reinforcement – predict that this covering effect will be reduced or removed if the
541 dipper is not presented after the training phase. In contrast, uncertainty within a causal model
542 account and the response prevention account both predict that the effects of covering the
543 magazine will be preserved, at least in the initial trials of the test phase in which the absence
544 of the dipper is not yet apparent.

545 Importantly, these predictions emphasise the test phase as a whole. However it has
546 already been noted that the presence or absence of the dipper might produce changes in the
547 levels of responding across the test phase. We have not considered trial-by-trial effects in the
548 predictions we have described thus far. The predictions of associative theories regarding
549 changes during extinction depend on the assumed learning parameters. Cognitive theories
550 would predict that changes of expectation depend on prior knowledge about causal stability
551 within the learning domain (e.g., physical vs. social). Little is known about these effects.
552 However, the very first trial of the test phase is different from all subsequent trials because
553 the response to the light is assessed before the dipper is presented (or not presented) and so
554 the Dipper versus No-Dipper manipulation cannot influence responding on the first test trial.
555 The impact of this fact is particularly clear in terms of the secondary reinforcement account
556 as it predicts that responding should emerge after only after the light is followed by the
557 dipper. Similarly, the renewal account predicts some responding to the light on the first trial
558 in the Dipper Cover and No-Dipper Cover conditions (because the removal of the cover is a
559 return to part of the training context), but only after the first trial will the Dipper vs No-
560 Dipper manipulation contribute to the context change between extinction and test phases.
561 Therefore, it should be recognised that the theoretical accounts we have presented here do

562 imply that responding could vary in a systematic fashion across trials, and that the different
563 accounts make divergent predictions about such trial-by-trial effects. That said, it should also
564 be acknowledged that the variability in responding that motivates the usual practice of
565 aggregating across multiple trials may make a reliable assessment of such fine-grained
566 predictions difficult in practice.

567 Sign-tracking vs. Goal-tracking:

568 Thus far, we have discussed responding to the light, following light-sucrose pairings,
569 entirely in terms of a single measure – magazine entry. However, Pavlovian conditioning can
570 establish a range of possible responses when a cue stimulus is paired with reward (Boakes,
571 1977). In particular, a distinction is made between sign-tracking, i.e., responding directed
572 towards the conditioned stimulus, and goal-tracking, i.e., responding towards the
573 unconditioned stimulus (for recent examples of this distinction in the context of cues
574 predicting food reward, see Flagel, Watson, Robinson, & Akil, 2007; Meyer et al., 2012). In
575 the present context, the original light to sucrose training should establish both a sign-tracking
576 response (e.g. orientation to the light) and a goal-tracking response (e.g. entry to the sucrose
577 magazine). Clearly, covering the sucrose magazine in extinction will prevent animals from
578 producing the same goal-tracking responses they produced in the training phase, but would
579 have no impact on the production of sign-tracking responses to the light. Therefore, an
580 examination of sign-tracking and goal-tracking responses would shed some light on the
581 mechanisms underpinning the effects of covering the food magazine during extinction. On a
582 practical note, sign-tracking to a light can be assessed by videoing the animals and measuring
583 the number of times the orient to the light. However, many studies of sign- vs goal-tracking
584 have used a retractable lever as the CS (Flagel et al., 2007; Meyer et al., 2012). Here, a lever
585 is inserted and removed from the box just as a light may be turned on and off. Critically, the
586 lever is entirely a signal; there is no need for the rats to press it in order for the reward to be

587 delivered. Despite this, rats will still approach and press the lever, and thus sign-tracking can
588 be measured by the number of lever presses, while goal tracking can continue to be assessed
589 through magazine entry. Table 2 outlines a proposed experiment using these techniques and
590 summarises the key predictions of each of the accounts in terms of sign and goal tracking
591 responses. This experiment would use a lever as the cue in place of the light used in previous
592 experiments to facilitate recording of sign-tracking responses, but all other aspects of the
593 experiment would remain the same. That is, the critical condition involves covering the food
594 magazine in the extinction phase. We will focus our analysis on this condition although a
595 control group receiving extinction without the magazine would still be needed to establish the
596 effects of experimental extinction for comparison purposes. As before, the derivation of
597 these predictions is fleshed out in turn for the causal model, response prevention, renewal,
598 and secondary reinforcement accounts.

599 The predictions of the causal model approach are based on the uncertainty
600 surrounding the appropriate causal structure. However, cognitive theories have not as yet
601 addressed how exactly expectations translate into different types of behaviour. Because the
602 relationship between model-based expectation and behavioural measures have not been the
603 subject of detailed consideration we have assumed here that, for all responses, a simple
604 monotonic function relates the degree of expectation of reward to the level of response⁸.
605 Critically rats that are sign-tracking respond towards to a cue to the extent that it reliably
606 predicts reward, and rats that are goal-tracking respond to the site of reward delivery during
607 the presentation of the cue, again, to the extent that the cue reliably predicts rewards. Thus
608 both sign- and goal-tracking behaviours are determined by the cue to reward relationship. In
609 terms of the causal model account described here this reflects the strength of the *light causes*

⁸ This represents a minimal assumption which allows the causal model approach to reflect the fact that both goal- and sign-tracking behaviours occur. It also focuses our analysis only on the effects of uncertainty regarding sucrose presentation in the extinction phase.

610 *sucrose* model. As described above, this model might be protected from the effects of
611 extinction through the creation of uncertainty about the status of the reward by covering of
612 the magazine. Under these preliminary assumptions, the consideration of uncertainty within
613 the causal model account predicts that both sign- and goal-tracking responses will be affected
614 by covering the sucrose magazine during the extinction phase.

615 As noted above, covering the magazine will prevent goal tracking (i.e. magazine
616 entry) responses, but would not prevent sign-training (i.e. lever press) responses. To the
617 extent that extinction requires the production of the relevant response, then covering the
618 magazine will attenuate the effects of extinction on goal-tracking responses but will not
619 influence the extinction of sign-tracking responses. Therefore, the action of response
620 prevention alone predict that levels of magazine responding to the light would be high at the
621 start of the test phase, while levels of lever press responding would be low.

622 With respect to the renewal account, the local context for the goal-tracking response is
623 the magazine. Covering the magazine is a distinct and salient change to this local context and
624 so the covering manipulation will mean that magazine responses at test will occur in a
625 different context to that experienced during extinction. As described above, this difference in
626 context between extinction and test phases should produce renewal and thus levels of
627 magazine responding (i.e. the goal tracking response) would be expected to be high at test. In
628 contrast, the local context for the sign-tracking response is the lever, which is not directly
629 affected by the covering manipulation. Thus, although the global context will differ between
630 extinction and test due to the presence/absence of the magazine cover, the local context for
631 sign-tracking responding will be the same for extinction and test. This similarity in the local
632 context for extinction and test should act to support generalisation of learning in extinction to
633 the test phase. Thus, while some renewal is expected for sign-tracking responses, this will

634 less than that seen for goal-tracking, and so renewal theory predicts that levels of lever-press
635 responding at test would be moderate.

636 The predictions of the secondary reinforcement account are somewhat less
637 categorical. Both sign- and goal-tracking after covering should relate to the same CS-US
638 relationship – where the effective US here is the conditioned reinforcement provided by the
639 dipper. So if covering preserved the conditioned reinforcing properties of the dipper then
640 both sign- and goal-tracking responses should return after the dipper is paired with the light
641 during test. However, there are large individual differences between animals in the levels of
642 sign- and goal-tracking responses they produce (Flagel et al., 2007; Meyer et al., 2012), and
643 animals that display a preponderance of sign-tracking responses may have a reduced
644 opportunity to interact with the conditioned reinforcer during the test phase. If so, then the
645 conditioned reinforcement account also predicts a greater effect of the covering manipulation
646 on goal-tracking than sign-tracking responses.

647 In summary, how uncertainty is translated into sign- and goal-tracking behaviours has
648 not been specified yet within the class of theories which includes causal model approaches.
649 Under the preliminary assumption that all responses reflect the strength of the underlying
650 *light causes sucrose* model, the causal model account predicts that sign- and goal-tracking
651 responses will both be affected by the magazine covering manipulation because uncertainty
652 about the status of the sucrose reward will protect this causal model. The three Level 1
653 associative accounts all relate to direct effects of the covering manipulation through either
654 preventing only one of the target responses in extinction, having different effects on the local
655 context for lever press and magazine entry responses, or by influencing the interaction with
656 the secondary reward. Thus the response competition and renewal accounts (and to a less
657 certain extent the secondary reinforcement account), predict that goal-tracking responses
658 should be more sensitive to magazine covering in extinction than sign-tracking responses.

659

660 Summary and comparisons to previous approaches

661 In the initial parts of this paper we outlined a distinction between two general classes
662 of theoretical accounts: Level 1 – which refers to accounts that focus on the representations
663 of events as experienced by the organism, and (in associative versions of such account at
664 least) involve only thin, non-semantic representations of events and the links between them;
665 and Level 2 – which refers to accounts that are focused on the idea that sensory experience is
666 the basis for forming models of the events in the world and the nature of the relationships
667 between them (with a particular focus on causal relationships), and thus involve explicitly
668 semantic representations of events. We then considered one test case involving extinction of
669 a classically conditioned CS-US relationship, where covering the food magazine during the
670 extinction phase attenuated the effects of that extinction in a subsequent test. While both
671 Level 1 and Level 2 accounts of the observed behaviour are available, these accounts make
672 divergent predictions about the effects of manipulating the details of how the reward was
673 delivered and the nature of the response assessed. Critically, these divergent predictions
674 speak directly to the level at which the theoretical accounts were based: The Level 1
675 accounts are based only on sensitivity to manipulations influencing the precise events
676 experienced by the animals in the test phase; while the Level 2 account we have considered is
677 focused on how covering the magazine creates uncertainty regarding the presence or absence
678 of the reward, which in turn will impact on how experiencing the absence of sucrose modifies
679 the causal model of the situation that was established during initial training. This influence of
680 uncertainty on the *light causes sucrose* model is explicitly a level 2 account as it clearly goes
681 beyond the direct effects of the sample of events experienced.

682 It should, of course, be noted that while the predictions of the four accounts
683 (uncertainty in causal models, response prevention, renewal, and secondary reinforcement)

684 are clear, it would be entirely possible to make post-hoc revisions or additions to them. For
685 example, a renewal theorist may suggest that the key feature of the context was not the dipper
686 but some other aspect of the magazine. Moreover, it should be emphasised that we have
687 focused the causal model account entirely on the effects that covering might have by inducing
688 animals to go beyond the direct effects of experience through creating uncertainty. But all
689 causal theories, regardless of their sensitivity to uncertainty, also assume Level 1 contingency
690 learning competencies. For example, on a causal account one could assume that the dipper is
691 part of the causal model learned in the acquisition phase (light-dipper-sucrose) so that its
692 absence in the test phase would lead to changes of expectation. These changes would be
693 solely due to Level 1 causal contingency learning which should be unaffected by the cover
694 manipulation in the extinction phase. That said, the current experiments do make a direct
695 comparison between an explanation in terms of uncertainty alone (i.e. an example of a Level
696 2 “beyond the sample” account) and explanations in terms of particular local features of the
697 manipulations (i.e. examples of Level 1 “sample-based” accounts). Thus, while the two
698 experimental manipulations described here do not comprise a definitive and general test of
699 causal model theory and its associative alternatives on their own, they do provide a specific
700 test of whether uncertainty over the presence or absence of reward considered alone is able to
701 explain the behaviour of animals in the current extinction situation.

702 We think it is instructive to compare our current approach – based on directly
703 examining one key (Level 2) aspect of a causal model account – with previous approaches.
704 In addition to the extinction experiments considered here, there are several other
705 demonstrations that preventing rats having access to the source of significant stimulus events
706 results in behaviour that is materially different to the simple non-presentation of those events
707 (Blaisdell, Leising, Stahlman, & Waldmann, 2009; Fast & Blaisdell, 2011). These other
708 covering experiments were discussed by Dwyer and Burgess (2011), but only to present

709 Level 1 associative accounts of the observed behaviours and to dismiss the originally-
710 proposed Level 2 accounts entirely on the basis of an appeal to Morgan's Canon. That is,
711 there was no discussion of how to make an empirically-based comparison between the
712 alternative accounts let alone any report of new or relevant empirical data. So, while the
713 Dwyer and Burgess analysis was of value in providing an existence-proof of an associative
714 account, it makes no progress towards determining whether the behaviour of the rats was
715 under the control of Level 1 or Level 2 mechanisms.

716 In summary, the current paper attempts to approach the investigation of the cognitive
717 mechanisms underpinning the behaviour of human and non-humans animals without bias
718 from preconceived assumptions regarding the prior probability of one account over another.
719 This approach supported the derivation of diagnostic empirical tests focusing on the key
720 feature of the current situation (i.e. the effect of uncertainty) which divided the current
721 theoretical accounts on the basis of the general level of representation they instantiate. Of
722 course, the proof of this particular pudding is in the baking, and we are in the process of
723 preparing to run exactly the studies we outline here.

724

725

726

- 727 Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and
728 belief-like states? *Psychological Review*, *116*(4), 953-970.
- 729 Blaisdell, A. P., Leising, K. J., Stahlman, W., & Waldmann, M. R. (2009). Rats distinguish
730 between absence of events and lack of information in sensory preconditioning.
731 *International Journal of Comparative Psychology*, *22*(1), 1-18.
- 732 Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in
733 rats. *Science*, *311*(5763), 1020-1022.
- 734 Boakes, R. A. (1977). Performance on learning to associate a stimulus with positive
735 reinforcement. In H. Davis & H. M. B. Hurwitz (Eds.), *Operant Pavlovian interactions*
736 (pp. 67-97). Hillsdale, NJ: Lawrence Erlbaum Associates.
- 737 Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*,
738 *11*(5), 485-494.
- 739 Burgess, K. V., Dwyer, D. M., & Honey, R. C. (2012). Re-assessing causal accounts of learnt
740 behavior in rats. *Journal of Experimental Psychology: Animal Behavior Processes*,
741 *38*(2), 148-156.
- 742 Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological*
743 *Review*, *104*, 367-405.
- 744 Colwill, R. M. (1991). Negative discriminative stimuli provide information about the identity
745 of omitted response-contingent outcomes. *Animal Learning & Behavior*, *19*(4), 326-
746 336.
- 747 Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2010). Beyond Stimulus
748 Cues and Reinforcement Signals A New Approach to Animal Metacognition. *Journal*
749 *of Comparative Psychology*, *124*(4), 356-368.

- 750 De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and
751 merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20*(4),
752 631-642.
- 753 Delamater, A. R. (2004). Experimental extinction in Pavlovian conditioning: Behavioural and
754 neuroscience perspectives. *Quarterly Journal of Experimental Psychology Section B-*
755 *Comparative and Physiological Psychology*, *57*(2), 97-132.
- 756 Dickinson, A. (2001). Causal learning: An associative analysis. *Quarterly Journal of*
757 *Experimental Psychology*, *54B*(1), 3-25.
- 758 Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- 759 Dwyer, D. M., & Burgess, K. V. (2011). Rational accounts of animal behaviour? Lessons
760 from C. Lloyd Morgan's canon. *International Journal of Comparative Psychology*, *24*,
761 349-364.
- 762 Dwyer, D. M., Starns, J., & Honey, R. C. (2009). "Causal Reasoning" in rats: A reappraisal.
763 *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(4), 578-586.
- 764 Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and
765 uncertainty on stimulus salience: a model of attention in associative learning.
766 *Proceedings of the Royal Society B-Biological Sciences*, *278*(1718), 2553-2561.
- 767 Evans, J. S. B. T. (2012). Dual-process theories of reasoning: Facts and fallacies. In K. J.
768 Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp.
769 115-133). New York, NY: Oxford University Press; US.
- 770 Fast, C. D., & Blaisdell, A. P. (2011). Rats are sensitive to ambiguity. *Psychonomic Bulletin*
771 *& Review*, *18*(6), 1230-1237.
- 772 Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical
773 inference. In B. H. Ross (Ed.), *Psychology of Learning and Motivation, Vol 57* (Vol.
774 57, pp. 1-55).

- 775 Fiedler, K., & Juslin, P. (2006). *Information sampling and adaptive cognition*. New York,
776 NY: Cambridge University Press; US.
- 777 Flagel, S. B., Watson, S. J., Robinson, T. E., & Akil, H. (2007). Individual differences in the
778 propensity to approach signals vs goals promote different adaptations in the dopamine
779 system of rats. *Psychopharmacology*, *191*(3), 599-607.
- 780 Gómez, J. C. (2004). *Apes, monkeys, children and the growth of mind*. Cambridge, MA:
781 Harvard University Press.
- 782 Gómez, J. C. (2005). Species comparative studies and cognitive development. *Trends in*
783 *Cognitive Sciences*, *9*(3), 118-125.
- 784 Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A
785 theory of causal learning in children: Causal maps and Bayes nets. *Psychological*
786 *Review*, *111*(1), 3-32.
- 787 Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological*
788 *Review*, *116*(4), 661-716.
- 789 Harris, J. A. (2006). Elemental representations of stimuli in associative learning.
790 *Psychological Review*, *113*(3), 584-605.
- 791 Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation
792 detection in causal induction: Model comparison and rational analysis. *Cognitive*
793 *Science*, *31*(5), 765-814.
- 794 Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philosophical*
795 *Transactions of the Royal Society B-Biological Sciences*, *367*(1603), 2695-2703.
- 796 Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action.
797 *Cognition*, *37*, 105-131.

- 798 Holland, P. C., & Rescorla, R. A. (1975). The effect of two ways of devaluing the
799 unconditioned stimulus after first- and second-order appetitive conditioning. *Journal of*
800 *Experimental Psychology-Animal Behavior Processes*, 1(4), 355-363.
- 801 Kutlu, M. G., & Schmajuk, N. A. (2012). Classical conditioning mechanisms can
802 differentiate between seeing and doing in rats. *Journal of experimental psychology.*
803 *Animal behavior processes*, 38(1), 84-101.
- 804 Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A
805 selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*
806 *Section B-Comparative and Physiological Psychology*, 57(3), 193-243.
- 807 Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple
808 reinforcement learning explains uncertainty in nonhuman animals. *Journal of*
809 *Experimental Psychology-Learning Memory and Cognition*, 38(3), 686-708.
- 810 Le Pelley, M. E., Oakeshott, S. M., Wills, A. J., & McLaren, I. P. L. (2005). The outcome
811 specificity of learned predictiveness effects: Parallels between human causal learning
812 and animal conditioning. *Journal of Experimental Psychology-Animal Behavior*
813 *Processes*, 31(2), 226-236.
- 814 Leising, K. J., Wong, J., Waldmann, M. R., & Blaisdell, A. P. (2008). The special status of
815 actions in causal reasoning in rats. *Journal of Experimental Psychology-General*,
816 137(3), 514-527.
- 817 Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian
818 generic priors for causal learning. *Psychological Review*, 115(4), 955-984.
- 819 Mackintosh, N. J. (1974). *The psychology of animal learning*. Oxford, England: Academic
820 Press.
- 821 Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with
822 reinforcement. *Psychological Review*, 82, 276-298.

- 823 Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford: Carendon Press.
- 824 Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic
825 causal reasoning. *Psychological Review*, *121*(3), 277-301.
- 826 Meyer, P. J., Lovic, V., Saunders, B. T., Yager, L. M., Flagel, S. B., Morrow, J. D., &
827 Robinson, T. E. (2012). Quantifying individual variation in the propensity to attribute
828 incentive salience to reward cues. *PLoS ONE*, *7*(6).
- 829 Morgan, C. L. (1894). *An introduction to comparative psychology*. London, England: Walter
830 Scott Publishing Co; England.
- 831 Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*,
832 *319*(5871), 1849-1851.
- 833 Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?
834 *Science*, *308*(5719), 255-258.
- 835 Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural
836 learning. *Animal Learning & Behavior*, *30*(2), 73-95.
- 837 Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the
838 effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*,
839 *87*, 532-552.
- 840 Pearl, J. (2000). *Causality*. Cambridge, UK: Cambridge University Press.
- 841 Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A
842 review and synthesis. *Psychonomic Bulletin & Review*, *14*(4), 577-596.
- 843 Rescorla, R. A. (2001). Experimental extinction. In R. R. Mowrer & S. B. Klein (Eds.),
844 *Handbook of contemporary learning theories* (pp. 119-154). Mahwah, NJ: Lawrence
845 Erlbaum.
- 846 Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in
847 the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F.

- 848 Prokasy (Eds.), *Classical Conditioning II: Current theory and research*. (pp. 64-99).
849 New York: Appelton-Century-Crofts.
- 850 Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of
851 explanatory depth. *Cognitive Science*, 26(5), 521-562.
- 852 Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental*
853 *Psychology. A, Human Experimental Psychology*, 48A(2), 257-279.
- 854 Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H.
855 Bower (Ed.), *The psychology of learning and motivation: Advances in research and*
856 *theory* (Vol. Vol. 21, pp. pp. 229-261). San Diego, CA, US: Academic Press, Inc.
- 857 Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative
858 psychology. *Trends in Cognitive Sciences*, 14(11), 477-481.
- 859 Wagner, A. R. (1981). SOP. A model of autonomic memory processing in animal behavior.
860 In N. E. Spear & R. Miller (Eds.), *Information processing in animals: Memory*
861 *mechanisms* (pp. 5-47). Hillsdale, NJ: Lawrence Erlbaum, Associates, Inc.
- 862 Waldmann, M. R. (2000). Competition among causes but not effects in predictive and
863 diagnostic learning. *Journal of Experimental Psychology-Learning Memory and*
864 *Cognition*, 26(1), 53-76.
- 865 Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in
866 rats and humans: a minimal rational model. In N. Chater & M. Oaksford (Eds.), *The*
867 *probabilistic mind. Prospects for Bayesian Cognitive Science* (pp. 453-484). Oxford:
868 Oxford University Press.
- 869 Waldmann, M. R., & Hagmayer, Y. (2013). *Causal reasoning*. New York, NY: Oxford
870 University Press; US.

- 871 Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given:
872 Causal models in learning and reasoning. *Current Directions in Psychological Science*,
873 *15*(6), 307-311.
- 874 Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal-
875 models: Asymmetries in cue competition. *Journal of Experimental Psychology-*
876 *General*, *121*(2), 222-236.
- 877 Waldmann, M. R., Schmid, M., Wong, J., & Blaisdell, A. P. (2012). Rats distinguish between
878 absence of events and lack of evidence in contingency learning. *Animal Cognition*,
879 *15*(5), 979-990.

Table 1 – Dipper Manipulation

Condition	Train	Extinction	Test	Uncertainty & Causal Model	Response Prevention	Renewal	Secondary Reinforcement
Dipper Cover	Light to sucrose filled dipper	Light alone & dipper magazine covered	Light to empty dipper	Status of reward uncertain in extinction phase – this protects <i>light causes sucrose</i> model. Expression of causal model at test supports responding to light.	Cover prevents magazine response, therefore extinction effect of light alone presentation reduced for this response.	Extinction and test phases differ in presence of the cover and dipper operation. This is a large difference between extinction and test phases, so expect renewal. I.e. Test phase responding high.	Primary reward (sucrose) removed. Secondary reward properties of dipper preserved as the dipper is not experienced without sucrose in extinction. Secondary reward can support responding at test. I.e. Test phase responding moderate.
No-Dipper Cover	Light to sucrose filled dipper	Light alone & dipper magazine covered	Light alone	I.e. Test phase responding moderate to high (depending on degree of protection by uncertainty).	I.e. Test phase responding high.	Extinction and test phases differ with in presence of cover, but are the same in the non-operation of the dipper. This is a smaller difference between extinction and test phases than in the Dipper Cover condition. So expect some renewal, but not as much as in Dipper Cover condition. I.e. Test phase responding moderate.	Primary reward (sucrose) and secondary (dipper) removed. Neither primary nor secondary reward can support responding at test. I.e. Test phase responding low.

Note 1: These predictions assume the cover completely blocks all access to the operation of the dipper. As an operational means to ensure this assumption is accurate, in the both the Dipper Cover, and No-Dipper Cover conditions, the dipper would not be operated at all in the extinction phase.

Note 2: Cells have been merged to highlight where predictions are not affected by the key manipulation.

Note 3: Additional control conditions where the extinction phase takes place without a magazine cover (e.g. Dipper No-Cover and No-Dipper No-Cover) would be needed in order to establish the baseline level of responding, these have not been illustrated here as all accounts predict experimental extinction and negligible responding at test.

Table 2 – Sign- vs Goal-tracking

Condition	Train	Extinction	Test	Uncertainty & Causal Model	Response Prevention	Renewal	Secondary Reinforcement
Dipper Cover Measure sign-tracking (lever press)	Lever insertion to sucrose filled dipper	Lever alone & Dipper magazine covered	Lever to empty dipper	<p>Status of reward uncertain in extinction phase – this protects <i>light causes sucrose</i> model. Expression of causal model at test supports responding.</p> <p>I.e. Test phase responding moderate to high for lever and magazine entry (depending on degree of protection by uncertainty).</p>	<p>Cover does not prevent lever response, therefore extinction from lever alone presentation expected.</p> <p>I.e. Test phase lever responding low.</p>	<p>Local context for sign tracking response is lever, which is unchanged between extinction and test phase. Unchanged local context attenuates renewal effect based on global context change due to extinction and test phases differing in presence of the cover and dipper operation.</p> <p>I.e. Test phase responding to the lever moderate.</p>	<p>Primary reward (sucrose) removed. Secondary reward properties of dipper protected by covering but high levels of orienting to lever may reduce experience of dipper as secondary reward. Secondary reward can support responding at test to the extent it is experienced.</p> <p>I.e. Test phase responding to the lever moderate to low.</p>
Dipper Cover Measure goal-tracking (magazine response)					<p>Cover prevents magazine response, therefore extinction effect of lever alone presentation reduced for this response.</p> <p>I.e. Test phase magazine responding high.</p>	<p>Local context for goal tracking response is the magazine. Extinction and test phases differences (magazine cover and dipper operation) focused on magazine. This is a large difference between extinction and test phases so expect renewal.</p> <p>I.e. Test phase magazine responding high.</p>	<p>Primary reward (sucrose) removed. Secondary reward properties of dipper protected by covering. Secondary reward can support responding at test.</p> <p>I.e. Test phase magazine responding moderate.</p>

Note1: This is a within-subject experiment with sign- and goal-tracking responses measured in all animals – however, the panels have been split to illustrate where different predictions are made for different response types.

Note 2: As with the previous experiment, these predictions assume the cover completely blocks all access to the operation of the dipper. As an operational means to ensure this assumption is accurate, in the Dipper Cover condition, the dipper would not be operated at all in the extinction phase.

Note 3: Again, additional control conditions where the extinction phase takes place without a magazine cover would be needed in order to establish the baseline level of responding, these have not been illustrated here as all accounts predict experimental extinction and negligible sign or goal tracking responding at test.

