

1 **Individual differences in learning behaviours in humans: Asocial**
2 **exploration tendency does not predict reliance on social learning**

3 **Author names and affiliations:**

4 Wataru Toyokawa^{a, b, 1, *}, Yoshimatsu Saito^{b, c}, Tatsuya Kameda^c

5 a: Department of Behavioural Science, Hokkaido University, N10W7, Kita-ku,
6 Sapporo, Japan 060-0810

7 b: Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku
8 Tokyo, Japan 102-0083

9 c: Department of Social Psychology, The University of Tokyo, 7-3-1 Hongo,
10 Bunkyo-ku, Tokyo, Japan 113-0033

11 1: Present address: School of Biology, University of St Andrews, St Andrews, Fife,
12 UK, KY16 9TH

13 **E-mail addresses:**

14 WT (wt25@st-andrews.ac.uk), YS (middle158725071@gmail.com), and TK
15 (tkameda@l.u-tokyo.ac.jp)

16 ***Corresponding author:** Wataru Toyokawa (wt25@st-andrews.ac.uk)

17 **Word count:** 5834 (excluding references), 7720 (including references)

18

19 **Abstract**

20 A number of empirical studies have suggested that individual differences in asocial
21 exploration tendencies in animals may be related to those in social information use.
22 However, because the ‘exploration tendency’ in most previous studies has been
23 measured without considering the exploration-exploitation trade-off, it is yet hard to
24 conclude that the animal asocial ‘exploration-exploitation’ tendency may be tied to
25 social information use. Here, we studied human learning behaviour in both asocial
26 and social multi-armed bandit tasks. By fitting reinforcement learning models
27 including asocial and/or social decision processes, we measured each individual’s (1)
28 asocial exploration tendency and (2) social information use. We found consistent
29 individual differences in the exploration tendency in the asocial tasks. We also found
30 substantive heterogeneity in the adopted learning strategies in the social task: One-
31 third of participants were most likely to have used the copy-when-uncertain strategy,
32 while the remaining two-thirds were most likely to have relied only on asocial learning.
33 However, we found no significant individual association between the exploration
34 frequency in the asocial task and the use of the social learning strategy in the social
35 task. Our results suggest that the social learning strategies may be independent from
36 the asocial search strategies in humans.

37

38 **Keywords:** Social learning; individual differences; reinforcement learning;
39 multi-armed bandit; exploration-exploitation; copy-when-uncertain

40

41 **1. Introduction**

42 To find better behavioural options in foraging, mate choice, nest search, etc., group
43 living animals can benefit from asocial information-gathering strategy (e.g.,
44 reinforcement learning rules (Sutton and Barto, 1998; Trimmer et al., 2012)) and from
45 strategic use of social information (Boyd and Richerson, 1985; Laland, 2004).
46 Although there has been much recent interest in the inter-individual variation in both
47 asocial and social learning behaviour (Mesoudi et al., 2016; Reader, 2015), little is
48 known about whether and (if so) how they associate with each other.

49 The adaptive reason of the individual differences in asocial exploration tendency
50 might be the trade-off between exploration and exploitation. Given the limited
51 time/energy budget, a single animal must strike the right balance between trying
52 unfamiliar behaviours to sample information (i.e., ‘exploration’) versus choosing
53 known best behaviour (i.e., ‘exploitation’) so as to improve the long-term net decision
54 performance (Cohen et al., 2007; Hills et al., 2014). The optimal balance of
55 exploration-exploitation depends on the costs and benefits of information gathering,
56 which may differ between individuals. For example, an individual with poor
57 information processing performance may have large costs of exploration, an
58 individual with shorter expected life-span may benefit less from sampling more
59 information, while an individual experiencing a temporary volatile environment may
60 be forced to explore so as to update their knowledge (Reader, 2015).

61 On the other hand, the individual variation in reliance on social information might
62 come from the balance of cost and benefit of copying others (Mesoudi et al., 2016).
63 For instance, an individual possessing inaccurate private information will potentially
64 incur a large cost if relying solely on the private knowledge and hence may tend to
65 copy others more (e.g., ‘copy-when-uncertain’ (Laland, 2004; Rendell et al., 2011)),
66 an individual living in a large group may benefit more from following the majority

67 (King and Cowlshaw, 2007), while an individual faced with a highly volatile
68 environment may rely more on private information due to the potentially large cost
69 from copying an out-of-date behaviour (Aoki and Feldman, 2014).

70 The increasing body of empirical studies has suggested that the individual
71 differences in the asocial exploration tendency might associate with those in the
72 social information use (but see Webster and Laland (2015)). For instance, the
73 individual exploration propensity negatively correlates with individual tendency of
74 copying conspecifics in barnacle geese *Branta leucopsis* (Kурvers et al., 2010a,
75 2010b) and zebra finches *Taeniopygia guttata* (Rosa et al., 2012), while the opposite
76 is true in three-spined sticklebacks *Gasterosteus aculeatus* (Nomakuchi et al., 2009)
77 and great tits *Parus major* (Marchetti and Drent, 2000). The interesting question here
78 would be why each of these species shows such individual correlations between
79 asocial and social search behaviour. Several possible mechanisms may be
80 conceivable to them. For example, environmental volatility may increase asocial
81 exploration tendency while also decreasing copying tendency. On the other hand, a
82 common cognitive ability underlying both asocial and social learning may generate a
83 positive correlation between them (Mesoudi et al., 2016).

84 However, the term 'exploration' has been used rather loosely in the previous
85 literature, and has been confounded with other personality traits (reviewed in Réale
86 et al. (2007)), which might have contributed to somewhat incoherent previous
87 findings on the relation between asocial exploration and social information use.
88 Broadly speaking, more active, neophilic, or bolder individuals tend to be labelled as
89 'explorative' while more inactive, neophobic, or shyer individuals tend to be labelled
90 as 'unexplorative' (Réale et al., 2007). However, it was untested whether more active
91 individuals actually gather information more (i.e., explore more) during the learning
92 process compared to inactive individuals. Also, from the viewpoint of the exploration-

93 exploitation trade-off, more explorative (exploitative) individuals are not necessarily
94 bolder (shyer): Especially in a changing environment, individuals who seldom explore
95 (i.e., who exploit the same option for a long time) may also be seen as very 'bold'
96 (e.g., Carere and Locurto, 2011; Groothuis and Carere, 2005; Koolhaas et al., 1999).
97 Therefore, it is yet hard to conclude that animal asocial 'exploration-exploitation'
98 tendency may be tied to the social information use. A more clear-cut measurement of
99 exploration behaviour is needed.

100 In this study, we focused on human learning behaviour in a multi-armed bandit (MAB)
101 problem, and saw whether the individual differences in asocial exploration tendency
102 might predict the reliance on social learning. In the MAB task, individuals have
103 multiple choice options, but at the outset they do not have exact knowledge of which
104 option is the most profitable (Figure 1a). In every round, each individual has to make
105 a decision whether to exploit (i.e., choosing the option that has higher estimated
106 reward value as of that round; see Methods section) or to explore (i.e., choosing the
107 other option with lower estimated reward value). Because the MAB problem embeds
108 the exploration-exploitation trade-off in its heart (Sutton and Barto, 1998), it is a
109 suitable test bed for unambiguously measuring exploration behaviour (Daw et al.,
110 2006; Keasar et al., 2002; Racey et al., 2011; Toyokawa et al., 2014). Fitting a
111 reinforcement learning model to each participant's decision data (O'Doherty et al.,
112 2003), we quantified each participant's asocial exploration tendency.

113 In addition to the asocial situation where participants engaged in the MAB task alone
114 (hereafter, 'solitary task'), participants also played the MAB task in a pairwise
115 situation ('paired task') in which they were able to observe the other participant's
116 choice (but not the peer's earned payoff) displayed on the monitor. To examine
117 whether the participants adopted social learning strategy in the paired task, we fitted
118 several asocial- and social-learning models to each participant's decision data, and

119 then selected the most likely learning model individually. Also, we analysed each
120 participant's gaze movement measured by an eye-tracker in order to confirm the
121 participant's information use during the task. Finally, we examined whether the
122 exploration tendency in the solitary task (i.e., asocial exploration) might predict the
123 use of social learning strategy in the paired task.

124

125 **2. Material and methods**

126 ***2-1 Participants***

127 Fifty-six right-handed undergraduate students were randomly selected from a subject
128 pool at Hokkaido University in Japan to participate in the experiment. Of these 56
129 participants, 8 participants failed at eye tracker calibration, leaving us with 48
130 participants (24 females; Mean Age \pm S.D. = 19.0 \pm 0.90) to be included in data
131 analysis. After the experimental session, participants received monetary rewards
132 based on their performance in the experimental tasks as compensation for their
133 participation (mean \pm S.D. = 1253 \pm 18.5 JPY).

134 ***2-2 Task overview: the restless 2-armed bandit***

135 Participants performed a restless 2-armed bandit task on a computer screen (Figure
136 1a). We used a '2-armed' task as the simplest case of a MAB. Each participant had
137 to repeatedly choose between two slot machines. The participant's goal was to
138 maximise the total reward earned over a sequence of plays by deciding which
139 machine to play for how many times.

140 Each round started with a 1-second interval during which a crossbar was shown at
141 the centre of the grey background. After this interval, two boxes (i.e., slot machines)
142 appeared on the left and right side of the screen. Participants chose the left box or

143 right box by pressing the 'left' or 'right' key (respectively) on a keyboard with their
144 right hand. Participants had a maximum of 2 seconds to make their choices (decision
145 interval): If no choice was made during the decision interval, a 'TIME OUT' message
146 appeared in the centre of the screen for 2.5 seconds to signal a missed round
147 (average number of missed rounds per participant was 0.44 out of 80 rounds in the
148 'first solitary task', 0.52 out of 180 rounds in the 'paired task', and 0.15 out of 80
149 rounds in the 'second solitary task'). If participants responded within 2 seconds, the
150 frame of the chosen option turned to be bold for $1.5 + (2 - \text{response time})$ seconds
151 so as to confirm their choice, followed by a 1-second display of earned points in the
152 chosen box. After showing the rewards (or showing 'TIME OUT'), i.e., after 4.5
153 seconds from the outset of the decision interval, the next round started with a
154 crossbar.

155 Each option yielded random points (50 points = 1 JPY) from a normal probability
156 distribution unique to each box, rounded up to the next integer, or truncated to zero if
157 it would have been a negative value (although this never happened). The mean
158 values of the probabilistic payoff were different between the two options. Additionally,
159 the mean values of payoff were changing during the task (Figure 1b). The standard
160 deviations of the probabilistic payoff distributions were identical for both boxes and
161 did not change during the task (S.D. = 10). Before each task started, participants
162 were informed about the total number of rounds (80 rounds in the first and second
163 solitary tasks, respectively; 180 rounds in the paired task) as well as about the
164 possibility that the mean payoff from the options might change at some points during
165 the task. However, they were not informed of the actual value of mean payoff from
166 each slot, when or how it would actually change, or the exact rate at which the payoff
167 points would be transformed into JPY after the experiment. To confirm that our
168 results would not change in different bandits' settings (e.g., mean payoff, pattern or
169 schedule of payoff changing), we used four different settings (hereafter, conditions)

170 for the paired task as a between-pair design, randomly assigned for each pair (Online
171 Supporting Material Figure S2).

172 The computer-based task was constructed using Python with PsychoPy package
173 (Peirce, 2007, 2009) with Tobii SDK 3.0 for Python. Python code used for the task is
174 available from the corresponding author. Further details of the settings of each task
175 are available in Online Supporting Material S1-1.

176

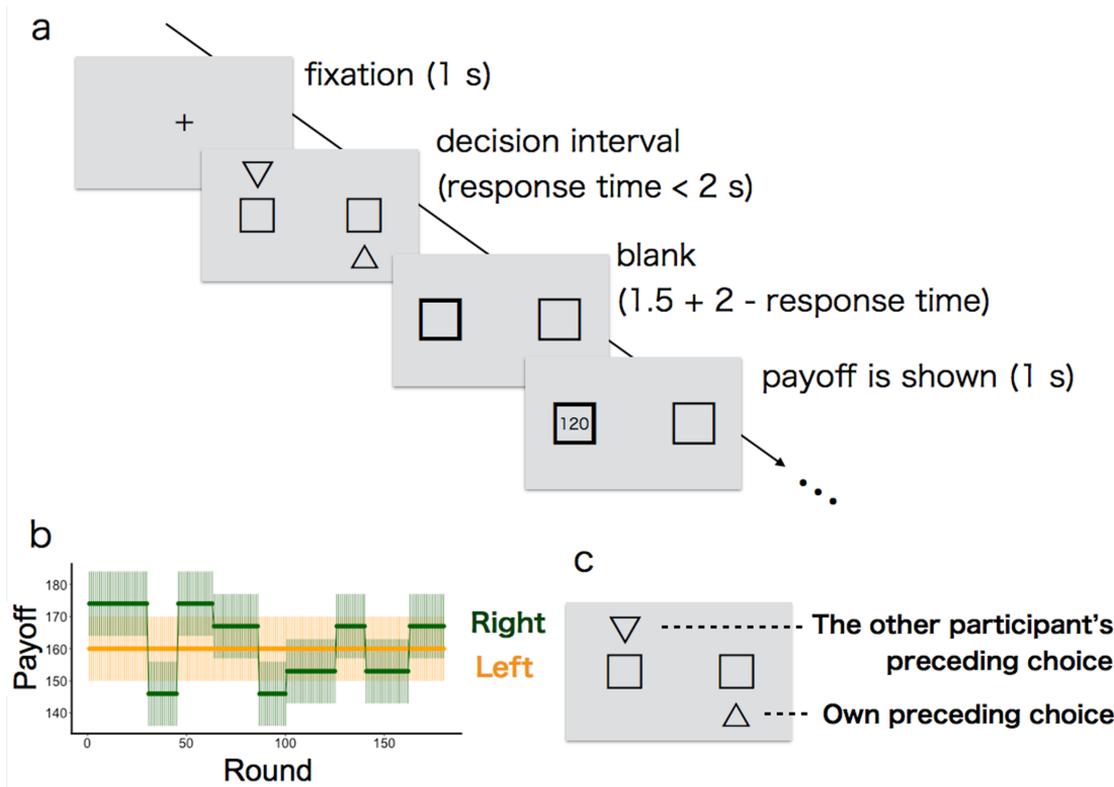
177

178

179

180

181



182

183 **Figure 1. The restless 2-armed bandit task.** (a) Illustration of the time line within a
184 round. After fixation with a crossbar for one second, two slots (boxes) were
185 presented. The participant had to choose one within 2 seconds (decision interval).
186 When choosing one, the frame of the chosen option turned to be bold. Up to 3.5
187 seconds later, the number of payoff points earned was revealed to the participant
188 (120 points in this example). After a further second, the next round started with a
189 crossbar. (b) Example of mean payoffs for each option in paired task (bold lines). The
190 payoff received for a particular choice is drawn from a Gaussian noise around each
191 mean (shaded areas show 1 S.D. of this noise). Note that the most profitable slot
192 ('optimal option') was switched several times due to the volatility; only one box was
193 volatile and the other box's mean was fixed. The left-right location of the volatile box
194 was counterbalanced across pairs in the paired task (the right box was volatile in this
195 example). (c) Example of social information in the paired task. The other participant's
196 choice in the preceding round was marked by a downward triangle (left in this
197 example) while their own choice in the preceding round was marked by an upward
198 triangle (right in this example).

199

200 **2-3 Experimental procedure**

201 For each experimental session, two participants of the same sex, randomly selected
202 from the participant pool, were called to the laboratory. Upon arrival, each participant
203 was seated in a separate soundproof chamber equipped with a computer terminal
204 and received general instructions about the experiment. After filling in the consent
205 form (see Ethics section), they read further paper-based instructions individually.
206 Participants remained strictly anonymous to each other throughout the experiment.

207 A single experimental session was composed of three tasks: (1) first solitary task, (2)
208 paired task, and (3) second solitary task. Although participants were informed that
209 there would be three tasks in total, they did not know any details of each task until
210 reading the instruction about the next task during the 5-minute break between the
211 tasks. The three tasks differed in both the mean payoffs of slots and their changing
212 pattern over time (see Online Supporting Material S1-1 for details). Furthermore, only
213 in the paired task were participants able to see social information—the other
214 participant's choice made in the preceding round—in addition to their own preceding
215 choice (Figure 1c). They were explicitly informed that both participants played the
216 identical task so that they were able to understand that the social information could
217 be informative. The session lasted for about 70 minutes in total.

218 **2-4 Acquisition and processing of gaze data**

219 We recorded each participant's gaze movement during the task at 300 Hz using Tobii
220 TX300 eye trackers connected to 23-inch monitors (Tobii Technology, Stockholm,
221 Sweden). Participants were seated with their head positioned on a chinrest, 70 cm in
222 front of the screen. Before each task, the calibration of the eye-tracker was validated
223 with five fixation points and re-calibrated if needed.

224 We used Tobii's default setting of noise reduction and fixation classification to
225 process the gaze data. Because we were interested in participants' information use
226 before making each decision, we focused upon the gaze positions in the 1-second
227 time window before each choice was made. Note that, because there was a 1-
228 second interval between rounds, the 1-second time window did not include gaze
229 positions in the previous round. The fixation positions of gazes were classified into
230 the following five categories: *gazing-at-social-information*, *gazing-at-own-preceding-*
231 *choice*, *gazing-at-the-left-box*, *gazing-at-the-right-box*, and *others* (see Online
232 Supporting Material S1-2 for the gaze data processing). All eye-tracking data and
233 Python code used in the gaze data processing are available from the corresponding
234 author upon request.

235 **2-5 Testing exploration tendency**

236 In order to classify each choice in the solitary tasks as either exploration or
237 exploitation, we fitted a standard reinforcement learning model to each participant's
238 choice data using maximal likelihood estimation method (Daw et al., 2006; O'Doherty
239 et al., 2003). Following the previous empirical studies on social learning strategies in
240 humans (McElreath et al., 2005, 2008), our learning model consists of two parts. First,
241 the values for choosing the options (i.e., Q-values (Sutton and Barto, 1998)) are
242 recursively updated by experienced reward according to the Rescorla-Wagner rule
243 (Trimmer et al., 2012) that is commonly used in the animal learning literature. Second,
244 the Q-values are transformed to the choice probability for each option by the 'softmax'
245 choice policy (Daw et al., 2006). The mathematical expression of this asocial learning
246 model is shown in the next sub-section (Eq.1 and Eq.2; see Online Supporting
247 Material S1-3 for full details of the model fitting procedure).

248 After obtaining the most likely parameters of the asocial learning model separately for
249 each participant, we classified each choice of each participant as either exploitative

250 or explorative according to whether the chosen box had the larger Q-value
251 (exploitation) or the smaller Q-value (exploration) between the two options. Then, we
252 summed up the number of explorations for each participant for each of the solitary
253 tasks. We calculated Pearson's correlation coefficient between the numbers of
254 explorations in the first and second solitary tasks so as to examine the individual
255 consistency in exploration tendency.

256 **2-6 Determining learning strategies**

257 We examined four different learning models as candidates of participants' strategies
258 in the paired task: *asocial learning model* (AL), *unconditional-copying model* (UNC),
259 *copy-when-uncertain model* (CWU), and *random choice model* (Random). AL and
260 Random are asocial, while UNC and CWU are social learning models. Since each
261 participant interacted with only one anonymous peer without seeing the peer's
262 earned payoff, we could not consider any 'frequency based' strategies (e.g., copy the
263 majority) and 'model based' strategies (e.g., copy the prestigious individual or payoff-
264 based copying) (Rendell et al., 2011). We fitted each of the models to each
265 participant's choice data individually.

266 All models have the same updating rule for Q-values, called the Rescorla-Wagner
267 rule. Q-value for the left option is updated as follows:

$$Q_{t+1,l} \leftarrow (1 - \alpha)Q_{t,l} + \alpha\pi_{t,l}, \quad (\text{Eq.1})$$

268 where α ($0 \leq \alpha \leq 1$) is a parameter determining the weight given to new experience
269 (learning rate) and $\pi_{t,l}$ is the amount of payoff obtained from choosing the left option
270 (l) in round t . When the left option was not chosen in round t , the left option's Q-
271 value was not updated ($Q_{t+1,l} = Q_{t,l}$). The same updating rule applied to the right
272 option. We set $Q_{0,l} = Q_{0,r} = 0$ because there was no reason to expect participants to
273 have any prior preference for either option at the outset (McElreath et al., 2008).

274 Of course, we could have considered any other updating rules (e.g., Bayesian
275 updating model (Payzan-LeNestour and Bossaerts, 2011)). However, because our
276 scope here is examining whether the participants might conduct social learning or not,
277 rather than quantifying detailed computational learning algorithms, we decided to
278 focus only on the simple Rescorla-Wagner rule which has been shown to be
279 evolutionary adaptive across a broad range of environmental conditions (Trimmer et
280 al. 2012).

281 The Q-values were transformed into the choice probability in different ways between
282 the four models, shown as follows.

283 **2-6-1 Asocial learning model (AL)**

284 In the asocial learning model, the probability of choosing the left option in round $t + 1$
285 is given by the following softmax rule:

$$P_{t+1,l} = \frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta) + \exp(Q_{t,r}/\beta)}, \quad (\text{Eq.2})$$

286 where β ($\beta > 0$) is a parameter that measures the influence of the difference
287 between Q-values on choice ('temperature' parameter). As $\beta \rightarrow \infty$, choice is
288 completely random (i.e., $P_{t+1,l} = 1/2$; highly explorative). As $\beta \rightarrow 0$, choice becomes
289 deterministic, in favour of the option with the higher Q-value (i.e., highly exploitative).
290 Therefore, β regulates the individual's asocial exploration tendency.

291 This asocial learning model was also used in the last subsection (section 2-5) to
292 estimate the exploration tendencies in the solitary tasks.

293 **2-6-2 Unconditional-copying model (UNC)**

294 Next we considered unconditional/unselective copying, the simplest case of social
295 information use (Laland, 2004). The individuals copied the other participant's choice

296 at the fixed rate λ ($0 \leq \lambda \leq 1$), otherwise their decisions were determined by the
297 softmax rule (Eq. 2). In the case that the peer chose the left option in round t , the
298 focal participant's probability of choosing the left option at round $t + 1$ is given by:

$$P_{t+1,l} = (1 - \lambda) \frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta) + \exp(Q_{t,r}/\beta)} + \lambda, \quad (\text{Eq.3-1})$$

299 whereas when the peer chose the right option in round t , the participant's probability
300 of choosing the left option at round $t + 1$ is given by:

$$P_{t+1,l} = (1 - \lambda) \frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta) + \exp(Q_{t,r}/\beta)}. \quad (\text{Eq.3-2})$$

301 When $\lambda = 1$, the individuals always copy the peer's choice. When $\lambda = 0$, the
302 individuals rely only on asocial learning. We assumed that the individuals also relied
303 only on the asocial learning rule when the peer missed the preceding round (i.e.,
304 when no choice was made by the peer).

305 **2-6-3 Copy-when-uncertain model (CWU)**

306 A number of studies have suggested that animals are selective in timing to use social
307 information, depending on the degree of uncertainty that they are experiencing (e.g.,
308 Coolen et al., 2003; Galef, 2009; van Bergen et al., 2004). To quantify the uncertainty
309 level concerning which of the two slots is more rewarding at a given round, we used
310 absolute difference in Q-values: The closer the Q-values between the two options,
311 the higher the uncertainty. In the copy-when-uncertain model, when the peer chose
312 the left option in round t , the focal participant's choice probability for the left option at
313 $t + 1$ is given by:

$$P_{t+1,l} = (1 - \gamma H_t) \frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta) + \exp(Q_{t,r}/\beta)} + \gamma H_t, \quad (\text{Eq.4-1})$$

314 whereas when the peer chose the right option in round t ,

$$P_{t+1,l} = (1 - \gamma H_t) \frac{\exp(Q_{t,l}/\beta)}{\exp(Q_{t,l}/\beta) + \exp(Q_{t,r}/\beta)}, \quad (\text{Eq.4-2})$$

315 where

$$H_t = - \sum_k \left(\frac{Q_{t,k}}{Q_{t,l} + Q_{t,r}} \right) \log_2 \left(\frac{Q_{t,k}}{Q_{t,l} + Q_{t,r}} \right). \quad (\text{Eq.4-3})$$

316 γ ($0 \leq \gamma \leq 1$) is a parameter that determines the upper limit of copying probability,
317 and H_t (i.e., information entropy, $0 \leq H_t \leq 1$) determines the actual copying rate at
318 round t , where $k \in \{l, r\}$. When both Q-values are equal ($Q_l = Q_r$), the uncertainty
319 becomes $\max(H = 1)$, leading the copying probability to be maximal (i.e., the
320 individual copies the peer at the probability γ when uncertainty is the highest). As the
321 difference between Q-values becomes larger, H approaches to 0, with the result that
322 the choice is mostly determined by the asocial learning.

323 **2-6-4 Random choice model**

324 We also considered the case that the choices are made randomly at fixed rate
325 regardless of the Q-values, so as to verify that the participants did not behave just
326 randomly in the experiment. The choice probability for the left option is always a fixed
327 rate ε ($0 \leq \varepsilon \leq 1$); hence, the probability of choosing the right option is $1 - \varepsilon$.

328 **2-7 Model selection**

329 We used Bayesian Model Selection (BMS) (Rigoux et al., 2014; Stephan et al., 2009)
330 that estimates how likely it is that each learning model generated the data of a
331 *randomly* chosen participant (see Online Supporting Material S1-4). Using BMS, we
332 were also able to calculate the probability that each model had generated a *given*
333 participant's data (Stephan et al., 2009). Because here we were interested in inter-
334 individual variation of social information use, we mainly focused on the latter
335 'individual-level' probabilities for each model, rather than the former 'group-level'

336 probability that BMS was originally aimed at. To verify this BMS estimation for each
337 participant's learning model, we also conducted the standard AIC comparison (see
338 Online Supporting Material S1-5).

339 ***2-8 The relation between gaze movement and learning strategy***

340 We examined the gaze movement so as to confirm the participants' information use
341 during the paired task. We focused on '*gazing-at-social-information before choice*'
342 (yes = 1/no = 0) for each round as a binary response variable in a binomial
343 generalised linear mixed model (GLMM), including the following three random
344 effects: individuals, pairs, and bandit's conditions. We considered estimated learning
345 strategy (social = 1/asocial = 0), information uncertainty (i.e., similarity between Q-
346 values [Eq. 4-3]), rounds, and possible 2-way interactions as fixed effects. Model
347 selection was done based on each AIC value (Online Supporting Material S1-6).

348 ***2-9 The relations between asocial exploration and social learning***

349 To examine whether the asocial exploration tendency might predict the use of social
350 learning strategy, we analysed a binomial GLMM with social = 1/asocial = 0 learning
351 as a binary response variable, with random effects of pairs and bandit's conditions.
352 We considered the asocial exploration tendency (frequency of explorative choices in
353 the first solitary task), asocial learning performance (average payoff earned per
354 choice in the first solitary task), and possible 2-way interactions as fixed effects.
355 Model selection was done based on each AIC value (Online Supporting Material S1-
356 7).

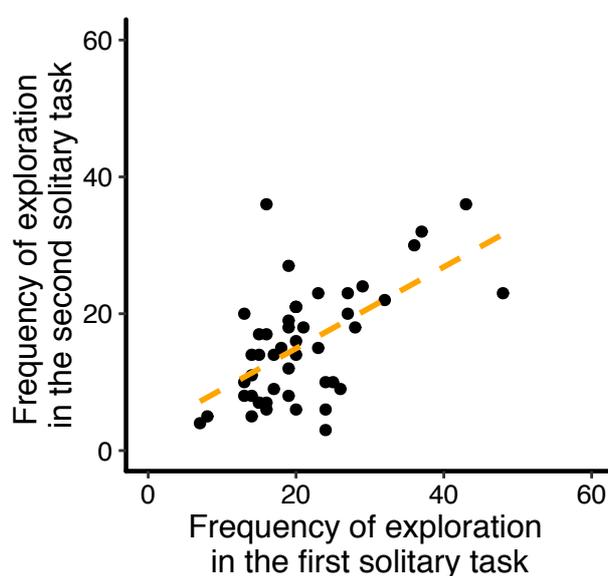
357

358 **3. Results**

359 **3-1 Individual consistency in the asocial exploration tendency**

360 The frequency of explorative choices in the first solitary task was positively correlated
361 with that in the second solitary task ($r = 0.58$, $p < 0.001$; Figure 2). This result
362 indicates that participants exhibited a stable asocial exploration tendency across the
363 two solitary tasks. Results of the fitting of the asocial learning model are shown in
364 Online Supporting Result S2-1.

365



366

367 **Figure 2. Individual consistency in the asocial exploration frequency.** The x-axis
368 shows the frequency of explorative choices in the first solitary task. The y-axis shows
369 the frequency of explorative choice in the second solitary task. The dotted line is a
370 linear regression.

371

372 **3-2 Detecting social learning strategies**

373 The Bayesian Model Selection (BMS) resulted in the heterogeneous distribution of
374 learning strategies among the participants (Figure 3a). The estimated likelihood of

375 each model for the whole data was 0.47 for the Asocial Learning model (AL), 0.10 for
376 the Unconditional Copying model (UNC), 0.41 for the Copy-when-uncertain model
377 (CWU), and 0.02 for the Random model.

378 Focusing on each individual's probabilities for each model, the BMS revealed that 32
379 participants were most likely to have adopted AL, while the remaining 16 participants
380 were most likely to have adopted CWU (Figure 3b). No individuals were most likely to
381 have used UNC or Random. We hereafter call the former 32 individuals 'asocial
382 learners' and the latter 16 individuals 'social learners' for simplicity. We also checked
383 each model's AIC values for each participant and confirmed that the result was not
384 qualitatively changed (although three participants were most likely to have adopted
385 the UNC model instead; see Online Supporting Material S2-2 and Figure S6).

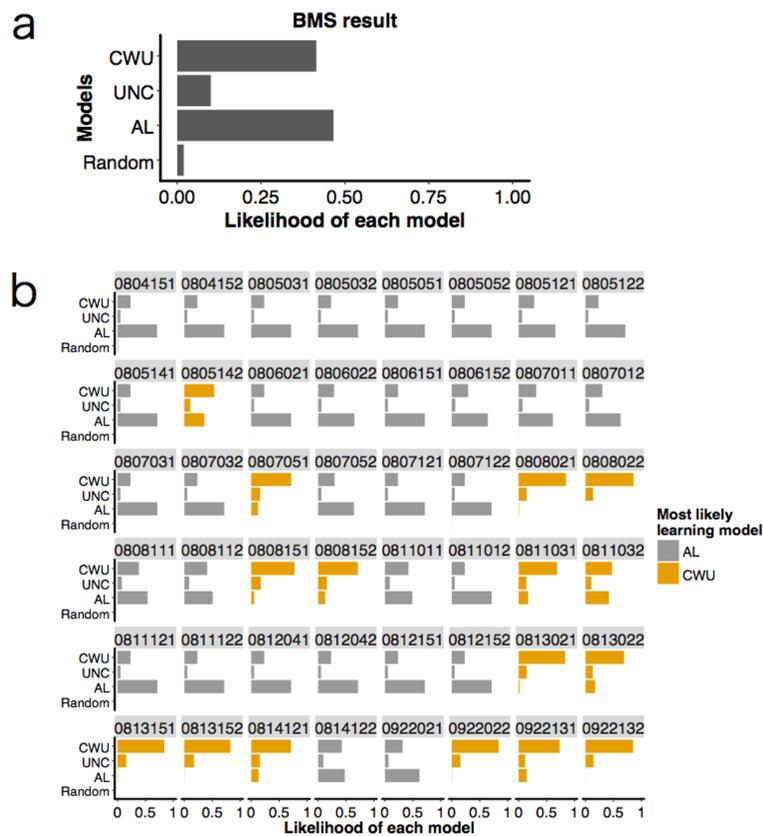
386 Table 1 shows means (± 1 SDs) of best-fitted parameters for asocial and social
387 learners. The mean value of γ (i.e., maximum copying probability) for social learners
388 (CWU) was 0.162 (± 0.072), which means that, when uncertainty level was max (i.e.,
389 setting $Q_l = Q_r$ so that $H = 1$ and $\exp(Q_t/\beta)/\sum \exp(Q_t/\beta) = 0.5$, see Eq. 4), there
390 was about 58% chance of choosing the same option as chosen by the other
391 participant in the preceding round. Note that asocial learners should choose
392 whichever option by 50% chance (i.e., randomly) when they have the maximum
393 uncertainty. Therefore, there was at most 8% increase in the probability of "copying
394 others' choice" by social learners as compared to asocial learners. Together with the
395 smaller frequency of social learners ($N = 16$) than asocial learners ($N = 32$), we will
396 revisit this rarity of social learning in Discussion.

	<i>Asocial Learning (AL)</i>	<i>Copy-when-uncertain (CWU)</i>
α (learning rate)	0.842 \pm 0.187	0.787 \pm 0.192
β (temperature)	11.31 \pm 10.39	6.03 \pm 2.03
γ (maximum copying probability)	N.A.	0.162 \pm 0.0720
N	32	16

397

398 **Table 1.** Fitted parameter values (mean \pm 1 S.D.) that provide the maximum
 399 likelihood for asocial learners (N = 32) and for social (copy-when-uncertain) learners
 400 (N = 16).

401



402

403 **Figure 3. Results of Bayesian Model Selection.** We considered the following four
404 models: random choice model, asocial learning model (AL), unconditional copying
405 model (UNC), and copy-when-uncertain model (CWU). (a) Distributions of
406 probabilities concerning how likely it is that each model was adopted by a randomly
407 chosen participant (i.e., group-level model likelihood). (b) Likelihood of each model
408 for each participant (i.e., individual-level model likelihood). The numbers shown
409 above each cell indicate participants' IDs. Thirty-two participants were most likely to
410 have used the asocial learning strategy (shown in grey); the remaining 16
411 participants were most likely to have used the copy-when-uncertain strategy (shown
412 in orange).

413

414 **3-3 Social learning did not improve the performance**

415 Given the result that some participants used social learning, an interesting question
416 is whether social learners performed better than asocial learners in the paired task.
417 We compared each individual's average payoff earned per choice using a Gaussian-
418 GLMM with a fixed effect of learning type (asocial = 0/social = 1), and with two
419 random effects of pairs and bandit's conditions. The fixed effect was not significant,
420 implying that social learners performed no better than asocial learners (*coefficient*
421 [95% confidence interval] = 0.38 [-0.33, 1.12], $p = 0.28$; see Online Supporting Figure
422 S8).

423 **3-4 The relation between gaze movement and learning strategy**

424 The result of BMS could be argued as an artefact of the limited number of models
425 considered. To confirm that estimated 'social learners' actually saw the social
426 information, we also examined the participants' gaze patterns. We used the
427 estimated learning strategy as a dummy fixed effect (i.e., asocial learner = 0; social
428 learner = 1). The binomial-GLMM for the probability of looking at social information
429 revealed the significant effects of uncertainty (coefficient [odds ratio] = -0.63 [0.53], p
430 = 0.036), round (coefficient [odds ratio] = 0.24 [1.25], $p < 0.001$), the interaction

431 between uncertainty and learning type (coefficient [odds ratio] = 1.45 [4.25], p =
432 0.0028), and the intercept (coefficient [odds ratio] = -3.25 [0.039], p < 0.001).
433 However, the fixed effect of learning strategy itself was not significant (coefficient
434 [odds ratio] = -0.65 [0.52], p = 0.37). The confidence intervals are shown in Online
435 Supporting Table S2.

436 The significant negative effect of uncertainty suggests that the participants tended
437 not to look at the social information when uncertainty (here defined as the closeness
438 of the two options in terms of Q-values) was high. However, the significant positive
439 interaction between uncertainty and learning strategies suggests that, when
440 uncertainty was high, the social learners tended to look at the social information more
441 frequently than asocial learners. This interaction effect is thus consistent with the
442 behavioural-choice pattern predicted by the CWU model as compared to the AL
443 model.

444 ***3-5 The relation between asocial exploration tendency and social learning strategy***

445 Our results showed that (1) there were consistent individual differences in asocial
446 exploration tendency across the two solitary tasks and that (2) use of social learning
447 was heterogeneous between individuals. Given this, we finally examined whether the
448 asocial exploration tendency in the solitary task might predict the use of social
449 learning strategy in the paired task. We analysed a binomial GLMM that predicts the
450 use of social learning strategy in the paired task (Online Supporting Material S1-7).

451 The selected GLMM contains both fixed effects of exploration frequency at the first
452 solitary task and performance at the first solitary task. However, none of them were
453 statistically significant (exploration tendency: coefficient [odds ratio] = -1.21 [0.30], p
454 = 0.16, performance: coefficient [odds ratio] = -1.05 [0.35], p = 0.07). Hence, we
455 cannot say either of them predicts a participant's reliance on social learning. Looking

456 at the scatter plots (Online Supporting Material Figure S8) rather than just
457 considering p-values of the GLMM parameters, however, the average performance of
458 the social learners in the first solitary task was not higher than the expected
459 performance from completely random choices (Student's t-test: $t_{15} = 0.64$, $p = 0.53$),
460 while that of asocial learners was better than the chance-level (Students' t-test: $t_{31} =$
461 6.15 , $p < 0.001$; Figure S8a). It might suggest that the asocial learning performance
462 in the first solitary task has a negative effect on the use of social learning in the
463 paired task. On the other hand, the frequencies of explorations in the first solitary
464 task were not different between asocial and social learners (Figure S9).

465

466 **4. Discussion**

467 In this study, we investigated human search strategies in the asocial/social 2-armed
468 bandit tasks, respectively, and tested whether the individual differences in asocial
469 exploration tendency in isolated settings might predict the use of social learning in
470 group settings.

471 Across the first and second solitary tasks, our results showed the consistent
472 individual differences in asocial exploration tendency (Figure 2). Since participants
473 were not informed how the environmental change would occur in advance, it was
474 virtually impossible for them to calculate the optimal exploration schedule (Gittins et
475 al., 2011). Therefore, individual differences in calculation abilities are unlikely to
476 explain this individual variation. Instead, it is known that human exploration tendency
477 in a learning/decision task may be dependent on dopaminergic functions in the
478 prefrontal cortical region of the brain (Daw et al., 2006; Frank et al., 2009), which is
479 associated with tracking informational uncertainty (Yoshida and Ishii, 2006). Sensing
480 decision uncertainty accurately is important in deciding when to explore in multi-

481 armed bandit problems (Cohen et al., 2007). The individual differences in asocial
482 exploration tendency shown in our result may be related to such individual
483 differences in neural activity for sensing uncertainty, although we did not investigate
484 direct neural mechanisms here.

485 As for the social task, we found a heterogeneous distribution of learning strategies:
486 Two-thirds of the participants seems to have used the *asocial learning* strategy while
487 the rest seems to have used the *copy-when-uncertain* strategy (Figure 3b), even
488 though a few of them might have adopted the *unconditional-copying* strategy instead
489 (Online Supporting Figure S6). The copy-when-uncertain type of social information
490 use is predicted by adaptive evolutionary models (e.g., Boyd and Richerson, 1988),
491 and has been repeatedly reported in empirical studies about non-human animals
492 (e.g., Galef et al., 2008; Kendal et al., 2009) as well as humans (e.g., Kameda and
493 Nakanishi, 2002; Morgan et al., 2011; Muthukrishna et al., 2015). Our result
494 replicated those findings. However, in the previous empirical studies, uncertainties
495 were externally manipulated by, for example, varying the task difficulty (Morgan et al.,
496 2011), changing the option numbers (Muthukrishna et al., 2015), or changing costs
497 required for getting asocial information (Kameda and Nakanishi, 2002). Instead of
498 fixing such parameters via experimental manipulations, our copy-when-uncertain
499 model tracks each individual's internal information uncertainty (i.e., entropy) that can
500 dynamically change during the course of learning. Our result thus provides a finer
501 picture about possible plasticity in human learning strategies, shedding light on the
502 individual flexibility of social information use during the course of problem solving.

503 It is noteworthy that, however, there were other forms of uncertainty that we did not
504 consider here. For example, there were uncertainties concerning how noisily the
505 payoff would be generated from a slot machine (i.e., variance of the payoff
506 distribution), how often the environment would change (i.e., 'unexpected uncertainty'),

507 and what kind of distribution payoff was generated from (i.e., ‘structural uncertainty’
508 or ambiguity) (Cohen et al., 2007; Payzan-LeNestour and Bossaerts, 2011; Payzan-
509 LeNestour et al., 2013). Further study is needed to investigate how those different
510 types of uncertainty may affect the timing of social information use.

511 Technically, we used AIC as an approximation to the log-evidence for the Bayesian
512 Model Selection (see Online Supporting Method). Compared to other forms of
513 calculation (e.g., Free Energy approximation), AIC tends to prefer complex models
514 because of the weak penalty for having more parameters (Penny, 2012; Stephan et
515 al., 2009). It might thus have caused over-estimation of the frequency of social
516 learners because our social learning models have more parameters than do asocial
517 models. Nevertheless, the gaze movement patterns were consistent with the
518 behavioural-choice pattern from the BMS result—individuals categorised as copy-
519 when-uncertain learners saw the social information more often than asocial learning
520 individuals when they faced high uncertainty (Result 3-4). Importantly, the BMS only
521 considered the behavioural (i.e., choice) data, which was measured independently
522 from the eye-tracking. Therefore, the result confirmed that we successfully captured
523 the significant pattern of individual differences in learning strategies.

524 Although evolutionary theory generally tends to suggest heavy reliance on social
525 learning in a broader range of situations (Rendell et al., 2011), only one-third of
526 participants seemed to have used social information in our experiment. In addition to
527 the low prevalence of social learners among participants, the maximum copying
528 probability of social learners was also low (at most 8% higher than that of asocial
529 learners). This rarity of social learning might be because social learning was not so
530 useful in the paired task (Result 3-3). Social information is useful if and only if others’
531 behaviours can filter better options (Rendell et al., 2010). In our paired task, however,
532 there was only one other individual playing the same task. Thus, the filtering effect

533 from social learning was minimal. Also, there might be no ‘worth-copying’ peer (e.g.,
534 expert or veteran) because both participants started the task at the same time: When
535 a focal participant was naïve to the task, so was the other participant. Therefore,
536 information about the peer’s choice might not be so accurate compared to the own
537 (asocial) learning experiences. Additionally, a recent empirical study suggests that
538 the reliance on social information becomes stronger as the number of choice options
539 increases (Muthukrishna et al., 2015). Having only two options, therefore, our
540 current study might underestimate the potential use of social learning. Further studies
541 are needed to investigate whether social information use may change with group size
542 and/or number of options.

543 We also explored the possible associations between behaviour in the solitary task
544 and social learning in the paired task. Different from the previous empirical findings
545 about non-human animals, we found no relation between asocial exploration
546 tendency and the reliance on social learning. One possible reason for this difference
547 might come from the inconsistent definition of ‘exploration’ in the previous literature.
548 As described in Introduction, asocial exploration tendency has often been
549 confounded with other personality traits that could relate to more general asocial
550 learning ability (Réale et al., 2007). Better asocial learners may show less social
551 information use because they possess more accurate private information (e.g.,
552 Kurvers et al., 2010a, 2010b), while the opposite (better asocial learners rely more on
553 social learning) could also be plausible if both asocial and social learning reflect a
554 common basic cognitive ability (Mesoudi et al., 2016). Indeed, a number of studies
555 have shown that asocial learning ability correlates with the use of social information
556 (Bouchard et al., 2007; Katsnelson et al., 2011; Mesoudi, 2011). Although it was not
557 statistically significant, our result might also suggest that better asocial learners tend
558 to ignore social information (Result 3-5; Figure S8a). Importantly, in our task, the
559 asocial exploration tendencies were not correlated with the asocial learning

560 performance (first solitary task: $r = -0.20$, $p = 0.17$; second solitary task: $r = -0.14$, $p =$
561 0.35 ; Online Supporting Figure S7). Our results that the asocial learning performance
562 might have a negative effect on the use of social learning (but not exploration
563 tendency) may suggest the importance of drawing a distinction between information-
564 gathering behaviour (e.g., exploration tendency) and learning performance.

565 Overall, our study supports that humans are very selective about when to use social
566 information. On the other hand, we should also acknowledge that our experimental
567 set-up did not allow considering any 'copy-from-whom' (Heyes, 2015; Laland, 2004;
568 Rendell et al., 2011) types of strategies, which might cause overlooking the potential
569 social information use. Comparing 'who' strategies with 'when' strategies in the same
570 framework will provide more comprehensive understanding on social learning for
571 both human and non-human animals. We also believe that the computational
572 learning model can be a strong tool for quantitative empirical investigations on animal
573 social learning strategies (Daw et al., 2006; McElreath et al., 2005, 2008; O'Doherty
574 et al., 2003; Sutton and Barto, 1998).

575

576 **5. Ethics**

577 This study was approved by the Institutional Review Board of the Centre for Experimental
578 Research in Social Science at Hokkaido University (No. H26-01). Written informed consent
579 was obtained from all participants before beginning the task.

580

581 **6. Data accessibility**

582 Behavioural data are available in Online Supporting Materials

583

584 **7. Competing interest**

585 We have no competing interest.

586

587 **8. Authors' contributions**

588 WT designed the study, performed the experiment, carried out the data analysis and
589 drafted the manuscript; YS carried out the acquisition of gaze data and participated in
590 the data analysis; TK participated in the design of the study, supervised the study
591 and participated in writing the manuscript. All authors gave final approval for
592 publication.

593

594 **9. Acknowledgments**

595 We are grateful to Shinsuke Suzuki for valuable discussions related to the
596 computational learning models, and Mike Webster for helpful comments on earlier
597 draft of this manuscript.

598

599 **10. Funding**

600 JSPS KAKENHI Grant Number 25245063 and 25118004 to Tatsuya Kameda, and
601 JSPS KAKENHI grant-in-aid for JSPS fellows 24004583 to Wataru Toyokawa. The
602 funders had no role in design study, data collection and analysis, decision to publish,
603 or preparation of the manuscript.

604

605 **11. References**

606 Aoki K, Feldman MW. 2014 Evolution of learning strategies in temporally and
607 spatially variable environments: A review of theory. *Theor. Popul. Biol.* 91, 3-
608 19. (doi:10.1016/j.tpb.2013.10.004)

609 Bouchard J, Goodyer W, Lefebvre L. 2007 Social learning and innovation are
610 positively correlated in pigeons (*Columba livia*). *Anim. Cogn.* 10, 259–266.
611 (doi: 10.1007/s10071-006-0064-1)

612 Boyd R, Richerson PJ. 1985 *Culture and the evolutionary process*. Chicago, IL:
613 University of Chicago Press.

614 Boyd R, Richerson PJ. 1988 An evolutionary model of social learning: The effects of
615 spatial and temporal variation. In T. Zentall & B. G. Galef (Eds.), *Social*
616 *learning: A psychological and biological approach* (pp. 29-48). Hillsdale, NJ:
617 Erlbaum.

618 Carere C, Locurto C. 2011. Interaction between animal personality and animal
619 cognition. *Curr. Zool.* 57(4), 491-498.

620 Cohen JD, McClure SM, Yu AJ. 2007 Should I stay or should I go? How the human
621 brain manages the trade-off between exploitation and exploration. *Phil. Trans.*
622 *Roy. Soc. B.* 362, 933–942. (doi: 10.1098/rstb.2007.2098)

623 Coolen I, van Bergen Y, Day RL, Laland KN. 2003 Species difference in adaptive use
624 of public information in sticklebacks. *Proc. Roy. Soc. B.* 270, 2413-2419. (doi:
625 10.1098/rspb.2003.2525)

- 626 Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates
627 for exploratory decisions in humans. *Nature* 441, 876–879.
628 (doi:10.1038/nature04766)
- 629 Frank MJ, Doll BB, Oas-Terpstra J, Moreno F. 2009 Prefrontal and striatal
630 dopaminergic genes predict individual differences in exploration and
631 exploitation. *Nat. Neurosci.* 12, 1062–1068. (doi:10.1038/nn.2342)
- 632 Galef, B. G. 2009 Strategies for Social Learning: Testing Predictions from Formal
633 Theory. *Adv. Study Behav.* 39, 117-151. (doi: 10.1016/S0065-
634 3454(09)39004-X)
- 635 Galef BG, Dudley KE, Whiskin EE. 2008 Social learning of food preferences in
636 'dissatisfied' and 'uncertain' Norway rats. *Anim. Behav.*, 75(2), 631-637.
637 (doi:10.1016/j.anbehav.2007.06.024)
- 638 Gittins J, Glazebrook K, Weber R. 2011 *Multi-armed Bandit Allocation Indices 2nd*
639 *Edition*. Sussex: John Wiley & Sons, Inc.
- 640 Groothuis TG, Carere C. 2005 Avian personalities: characterization and epigenesis.
641 *Neuroscience & Biobehavioral Reviews*, 29(1), 137-150.
- 642 Heyes C. 2015 Who Knows? Metacognitive Social Learning Strategies. *Trends. Cog.*
643 *Sci.* 20, 204–213. (doi: 10.1016/j.tics.2015.12.007)
- 644 Hills TT, Todd, PM, Lazer D, Redish AD, Couzin ID, CSR Group. 2014 Exploration
645 versus exploitation in space, mind, and society. *Trends Cog. Sci.* 19, 46-54.
646 (doi: <http://dx.doi.org/10.1016/j.tics.2014.10.004>)
- 647 Kameda T, Nakanishi D. 2002 Cost-benefit analysis of social/cultural learning in a
648 nonstationary uncertain environment: an evolutionary simulation and an

- 649 experiment with human subjects. *Evol. Hum. Behav.* 23, 373–393.
650 (doi:10.1016/S1090-5138(02)00101-0)
- 651 Katsnelson E, Motro U, Feldman MW, Lotem A. 2011. Individual-learning ability
652 predicts social- foraging strategy in house sparrows. *Proc. Roy. Soc. B* 278,
653 582–589. (doi: 10.1098/rspb.2010.1151)
- 654 Keasar T, Rashkovich E, Cohen D, Shmida A. 2002 Bees in two-armed bandit
655 situations: foraging choices and possible decision mechanisms. *Behav. Ecol.*
656 13, 757–765. (doi: 10.1093/beheco/13.6.757)
- 657 Kendal RL, Kendal JR, Hoppitt W, Laland KN. 2009 Identifying Social Learning in
658 Animal Populations: A New ‘Option-Bias’ Method. *PLoS ONE* 4, e6541.
659 (doi:10.1371/journal.pone.0006541)
- 660 King AJ, Cowlshaw G. 2007 When to use social information: the advantage of large
661 group size in individual decision making. *Bio. Lett.* 3, 137-139. (doi:
662 10.1098/rsbl.2007.0017)
- 663 Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H,
664 De Jong IC, Ruis MAW, Blokhuis HJ. 1999 Coping styles in animals: current
665 status in behavior and stress-physiology. *Neuroscience & Biobehavioral*
666 *Reviews*, 23(7), 925-935.
- 667 Kurvers RHJM, Prins HHT, van Wieren SE, van Oers K, Nolet BA, Ydenberg RC.
668 2010a The effect of personality on social foraging: shy barnacle geese
669 scrounge more. *Proc. R. Soc. B.* 277: 601–608 (doi: 10.1098/rspb.2009.1474)
- 670 Kurvers RHJM, van Oers K, Nolet BA, Jonker RM, van Wieren SE, Prins HHT,
671 Ydenberg RC. 2010b Personality predicts the use of social information. *Ecol.*
672 *Lett.* 13: 829–837 (doi: 10.1111/j.1461-0248.2010.01473.x.)

- 673 Laland KN. 2004 Social learning strategies. *Learn. Behav.* 32, 4–14.
674 (doi:10.3758/BF03196002)
- 675 Marchetti C, Drent PJ. 2000 Individual differences in the use of social information in
676 foraging by captive great tits. *Anim. Behav.* 60, 131–140.
677 (doi:10.1006/anbe.2000.1443)
- 678 McElreath R, Lubell M, Richerson PJ, Waring TM, Baum W, Edsten E, Efferson C,
679 Paciotti B. 2005 Applying evolutionary models to the laboratory study of social
680 learning. *Evol. Hum. Behav.* 26, 483–508.
681 (doi:10.1016/j.evolhumbehav.2005.04.003)
- 682 McElreath R, Bell AV, Efferson C, Lubell M, Richerson PJ, Waring T. 2008 Beyond
683 existence and aiming outside the laboratory: Estimating frequency-dependent
684 and pay-off-biased social learning strategies. *Phil. Trans. Roy. Soc. B* 363,
685 3515–3528. (doi: 10.1098/rstb.2008.0131)
- 686 Mesoudi A. 2011 An experimental comparison of human social learning strategies:
687 payoff-biased social learning is adaptive but underused. *Evol. Hum. Behav.*
688 32, 334–342. (doi: 10.1016/j.evolhumbehav.2010.12.001)
- 689 Mesoudi A, Chang L, Dall SRX, Thornton A. 2016 The Evolution of Individual and
690 Cultural Variation in Social Learning. *Trends. Ecol. Evol.* (doi:
691 <http://dx.doi.org/10.1016/j.tree.2015.12.012>)
- 692 Muthukrishna M, Morgan TJH, Henrich J. 2015 The when and who of social
693 learning and conformist transmission. *Evol. Hum. Behav.* 37, 10–20.
694 (doi:10.1016/j.evolhumbehav.2015.05.004)

- 695 Morgan TJH, Rendell LE, Ehn M, Hoppitt W, Laland KN. 2011 The evolutionary basis
696 of human social learning. *Proc. R. Soc. B* 279, 653–662. (doi:10.
697 1098/rspb.2011.1172)
- 698 Nomakuchi S, Park PJ, Bell MA. 2009 Correlation between exploration activity and
699 use of social information in three-spined sticklebacks. *Behav. Ecol.* 20, 340–
700 345. (doi: 10.1093/beheco/arp001)
- 701 O’Doherty JP, Dayan P, Friston K, Critchley H, Dolan, RJ. 2003 Temporal difference
702 models and reward-related learning in the human brain. *Neuron* 38, 329–
703 337. (doi:10.1016/S0896-6273(03)00169-7)
- 704 Payzan-LeNestour E, Bossaerts P. 2011 Risk, Unexpected Uncertainty, and
705 Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLoS*
706 *Comput. Biol.* 7, e1001048. (doi:10.1371/journal.pcbi.1001048)
- 707 Payzan-LeNestour E, Dunne S, Bossaerts P, O’Doherty, JP. 2013 The Neural
708 Representation of Unexpected Uncertainty during Value-Based Decision
709 Making. *Neuron* 79, 191-201 (doi:
710 <http://dx.doi.org/10.1016/j.neuron.2013.04.037>)
- 711 Peirce, JW. 2007 PsychoPy - Psychophysics software in Python. *J. Neurosci.*
712 *Methods* 162, 8-13. (doi: 10.1016/j.jneumeth.2006.11.017)
- 713 Peirce JW. 2009 Generating stimuli for neuroscience using PsychoPy. *Front.*
714 *Neuroinform.* 2:10. (doi:10.3389/neuro.11.010.2008)
- 715 Penny WD. 2012 Comparing dynamic causal models using AIC, BIC and free energy.
716 *NeuroImage*, 59, 319–330. (doi: 10.1016/j.neuroimage.2011.07.039)

- 717 Racey D, Young ME, Garlick D, Pham JNM, Blaisdell AP. 2011 Pigeon and human
718 performance in a multi-armed bandit task in response to changes in variable
719 interval schedules. *Learn. Behav.* 39, 245–258. (doi: 10.3758/s13420-011-
720 0025-7)
- 721 Reader SM. 2015 Causes of Individual Differences in Animal Exploration and
722 Search, *Topics in Cognitive Science* 7, 451–468.
- 723 Réale D, Reader SM, Sol D, McDougall PT, Dingemanse, NJ. 2007 Integrating
724 animal temperament within ecology and evolution. *Biological Reviews*, 82,
725 291–318.
- 726 Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, Forgaty L,
727 Ghirlanda S, Lillicrap T, Laland KN. 2010 Why copy others? Insights from the
728 social learning strategies tournament. *Science* 328, 208–213. (doi:
729 10.1126/science.1184719)
- 730 Rendell L, Fogarty L, Hoppitt WJE, Morgan TJH, Webster MM, Laland KN. 2011
731 Cognitive culture: theoretical and empirical insights into social learning
732 strategies. *Trends. Cog. Sci.* 15, 68-76. (doi:10.1016/j.tics.2010.12.002)
- 733 Rigoux L, Stephan KE, Friston KJ, Daunizeau J. 2014 Bayesian model selection for
734 group studies — Revisited. *NeuroImage* 84, 971-985. (doi:
735 10.1016/j.neuroimage.2013.08.065)
- 736 Rosa P, Nguyen V, Dubois F. 2012 Individual differences in sampling behaviour
737 predict social information use in zebra finches. *Behav. Ecol. Socio. Biol.* 66,
738 1259-1265. (doi: 10.1007/s00265-012-1379-3)

- 739 Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009 Bayesian model
740 selection for group studies. *NeuroImage* 46, 1004–1017. (doi:
741 10.1016/j.neuroimage.2013.08.065)
- 742 Sutton RS, Barto AG. 1998 *Reinforcement Learning: An Introduction*. Cambridge,
743 MA: MIT press.
- 744 Toyokawa W, Kim H, Kameda T. 2014 Human Collective Intelligence under Dual
745 Exploration-Exploitation Dilemmas. *PLoS ONE* 9, e95789. (doi:10.
746 1371/journal.pone.0095789)
- 747 Trimmer PC, McNamara JM, Houston AI, Marshall JAR. 2012 Does natural selection
748 favour the Rescorla-Wagner rule? *Journal of Theoretical Biology*, 302,39-52.
749 (<http://dx.doi.org/10.1016/j.jtbi.2012.02.014>)
- 750 van Bergen Y, Coolen I, Laland KN. 2004 Nine-spined sticklebacks exploit the most
751 reliable source when public and private information conflict. *Proc. Roy. Soc. B*.
752 271, 957-962. (doi: 10.1098/rspb.2004.2684)
- 753 Webster M, Laland K. 2015 Space-use and sociability are not related to public-
754 information use in ninespine sticklebacks. *Behav. Ecol. Scio. Biol.* 69, 895-
755 907. (doi: 10.1007/s00265-015-1901-5)
- 756 Yoshida W, Ishii S. 2006 Resolution of uncertainty in prefrontal cortex. *Neuron* 50,
757 781–789.