City Research Online

City, University of London Institutional Repository

1

2

3

4

5

6

7

8

9 **Human value learning and representation reflects rational adaptation to task demands**

10 Keno Juechems[1,2], Tugba Altun[3], Rita Hira[3], Andreas Jarvstad[3*]

11

12 [1] Department of Experimental Psychology, University of Oxford, UK

13 [2] St John's College, University of Oxford, UK

14 [3] Department of Psychology, City University of London, UK

15

16

17

18

19

20 Main text - 4957 words

21 Methods - 2374 words

22 7 figures.

23

24 Additional text and figures in Supplementary Methods and Supplementary Results

25

26 Andreas Jarvstad, andreas.jarvstad@city.ac.uk, College Building, City, University of London,

27 Northampton Square, London EC1V 0HB, United Kingdom.

28

29   **Abstract**

30   Humans and other animals routinely make choices between goods of different value. Choices are

31   often made within identifiable contexts, such that an efficient learner may represent values relative to

32   their local context. However, if goods occur across multiple contexts, a relative value code can lead to

33   irrational choice. In this case, an absolute context-independent value is preferable to a relative code.

34   Here, we test the hypothesis that value representation is not fixed, but rationally adapted to context

35   expectations. In two experiments, we manipulated participants' expectations about whether item

36   values learned within local contexts would need to be subsequently compared across contexts. Despite

37   identical learning experiences, the group whose expectations included choices across local contexts,

38   went on to learn more absolute-like representation than the group whose expectations only covered

39   fixed local contexts. Thus, human value representation is neither relative nor absolute, but efficiently

40   and rationally tuned to task demands.

41

42

43

44

45

46

47

48

49

50

51

52

Introduction

54   Humans and other animals often behave "as if" they calculated the value of goods, arranged

55   goods according to their preferences in a rational manner, and chose the good with highest value. One

56   way to achieve rational decision-making is to represent all items on an absolute scale, where an item's

57   value is expressed as the amount of fixed units of measurement it provides. Units of measurement

58   might be the number of food items in a foraging patch, money, or the subjective utility of consumer

59   products. Such an absolute value code is assumed in normative theories of decision-making[1], optimal

60   foraging theory[2], computational models of learning[3] , and in key descriptive theories of choice[4].

61   Whilst an absolute code would equip the agent to make decisions across all contexts in which

62   this unit of measurement is relevant, there are many reasons why biologically constrained systems

63   may utilise different coding regimes. For example, absolute codes maintaining a constant unit may

64   reserve precious coding range for values that occur with low frequency. Moreover, absolute codes

65   may be more prone to deleterious noise if values cluster within a small range in a given context

66   (leading to easily confusable items).

67   From the olfactory system in the fruitfly[5], to visual systems[6], through to value coding in

68   humans[7], neural systems can overcome such problems by encoding input relative to the local context

69   (and/or state[8,9]). The value of one foraging patch can, for example, be encoded relative to other nearby

70   patches. Such context-dependent encoding has been formalised in computational models, for instance

71   by ensuring that coding covers the entire range of values ('range adaptation'[10]) or by ensuring that

72   values are normalised by concurrent inputs ('divisive normalisation'[11]).

73   The key advantage of relative value codes is that they enable even small populations of

74   neurons to efficiently represent items within a local context[11]. For the perceptual system, for example,

75   adapting to local brightness levels (e.g., dark adaptation[12]) is likely close to optimal given the

76   temporal and spatial autocorrelation in brightness in natural scenes (e.g., day-night light cycle). For

77   value-based decisions, agents can boost discriminability of items of similar value using relative codes,

78   which may be of particular importance if the agent aims to choose "correctly" (i.e., choose the highest

79   valued item). This means that a foraging animal employing a relative value code may discriminate

80   between patches of values A = 5 and B = 6 with equal precision to when choosing between patches of

81   values C = 20 and D = 21.

82   There is now ample evidence from psychology, behavioural ecology, primate

83   neurophysiology and cognitive neuroscience that humans and other animals learn, and/or make

84   choices consistent with such context-dependent value codes ([9,13–19] but see[20]). A relative context-

85   dependent code also describes the firing pattern of neurons in value-related areas of the prefrontal

86   cortex[21] and explains human errors of judgment across many domains[17]. Relative codes have also

87   been shown to be efficient in the sense that they maximize mutual information between stimulus and

88   neural code under certain conditions[22]. In this latter sense, context-dependent codes can be locally

89   optimal and resource efficient – allowing animals to choose the best option with the use of minimal

90   resources[22,23].

91   However, as can easily be seen, relative value encoding can lead to inferior decision-making

92   if the local contexts in which values were encoded are intermixed. In the above example, for instance,

93   foraging patch B = 6 is the locally superior option to A = 5, which means that a pure relative encoder

94   may prefer it to the globally superior option from a different context - provided it is inferior in its

95   local context (e.g., prefer B=6 to C=20, where C is from [C=20, D=21]). Such 'irrational' decision-

96   making has been observed across species in many laboratory tasks[10,15,17,24].

97   Thus, one is faced with an additional problem: How to arbitrate the costs and benefits of

98   absolute and relative encoding to optimize decision-making. This problem can be recast as one of

99   expectation about context: If contexts are stable and distinct, relative encoding will be sufficient and

100   maximizes discriminability, but if contexts are either volatile or overlapping in time, a coding regime

101   approximating absolute encoding will be better. Here, we take a first step towards this question by

102   implicitly manipulating human participants' expectations about contexts in two experiments. In spirit,

103   our work is similar to efforts in reinforcement learning to delineate under what circumstances, and

104   under what cost, humans switch from a habitual (model-free) representation to a more costly

105   representation that allows planning (model-based)[25,26].

106   In particular, we propose that humans do not use a single fixed representation of value, but

107   flexibly tune value codes based on their expectations what the codes are for[27]. Further, we propose

108   that the selection of which code to learn, is rational and efficient[28]. Thus, we do not ask whether

109   human value learning is absolute or relative overall [13,15], but whether humans flexibly adapt[29,30] their

110   value representation in a manner that can be explained by expectation.

111   We tested the hypothesis that value representation rationally adapts to task demands in two

112   value-learning experiments, in which human participants learned values of pair-wise presented items.

113   We implicitly manipulated task expectations, such that one group expected to make decisions within

114   fixed local contexts ('Uncrossed'), and another group expected to make decisions across local

115   contexts ('Crossed'). If value learning is fixed, the learnt value representations should be identical

116   across groups. If value learning is rationally and flexibly adapted to task demands, people in the

117   'Crossed' group should go on to learn more absolute-like representations (because they expect these

118   to be task-relevant).

119   Despite identical learning experiences, learnt value codes differed: participants learned more

120   complex (absolute) representations only when they expected it to be necessary, thus highlighting the

121   rational and dynamic nature of value representation.

122

123   **Results**

124   **Tasks and design**

125   We conducted two value learning experiments. The first experiment used real-valued items,

126   akin to studies in economic decision-making[31], whereas the second used binomial outcomes akin to

127   many reinforcement learning paradigms in this domain[15]. In both experiments, participants went

128   through two independent phases of learning and decision-making.

129   In the learning phases, participants learned the value of items through trial-by-trial feedback.

130   As our key experimental manipulation, we implicitly altered participants' expectations about the local

131   contexts in which items had to be compared. After the initial learning phase, one group ('Uncrossed')

132   was presented with choices between fixed pairs of items (within contexts), whereas the other group

133   ('Crossed') encountered items also in intermixed pairs (across contexts).

134   We expected the Crossed group to use the experience of intermixed contexts to alter their value

135   encoding for the subsequent independent item set. Value representations in both groups were

5

136   measured with two surprise tasks at the end of each experiment (see below). In the following, we first

137   report on Experiment 1, which used real-valued items.

138   Participants took on the role of consultants to manufacturers of reproduction items (replicas of

139   historical items). There were two separate manufacturers (of cars & antiques) in two separate Phases

140   (Fig 1A). Participants' goal was to learn market prices to advise on which items to manufacture. In

141   the Learning Phases, participants learned item values through trial-by-trial feedback, after which they

142   advised the manufacturer in separate Decision phases - without feedback. At the end, there were two

143   surprise tasks (All-possible pairs, Value judgment) designed to measure value encoding in the last

144   Learning phase.

145   Participants were randomly and blindly assigned to either the Uncrossed or Crossed group

146   (colour-coded green and blue respectively, Fig 1A). In Experiment 1, each Phase began with a

147   Learning stage, in which participants sampled market values (Fig 1B). A single mouse-click on an

148   item returned a single sale price (superimposed on the clicked item, Fig 1B). Participants were free to

149   sample in any order and as much as they wished. Sampling for each pair was terminated by a selling

150   decision (Fig 1B), after which the next pair was shown. In each Phase, participants learned the values

151   of 6 items arranged into 3 pairs with normally distributed market prices (Fig 1A).

152   In the Decision phases (Fig 1C), the Uncrossed group made decisions about the pairs they had

153   previously experienced. The Crossed group additionally made decisions within novel pairings, thus

154   breaking their learning contexts (Methods). Participants might, for example, decide between $Item_2$

155   and $Item_4$ which had previously formed part of the first and second pair respectively. Participants'

156   choices in the Decision phases and surprise tasks were incentive-compatible (Methods).

157   We hypothesized that people do not use a fixed value-learning mechanism, but flexibly adapt

158   their value-learning mechanisms to learn useful value representations. Given double-blind assignment

159   to groups, both groups should start Learning 1 with the same expectations. However, the first

160   Decision phase, Decision 1 (Fig 1C), provides very different implicit signals for the two groups.

161   The Uncrossed group should have no problem performing in this task given successful learning

162   (Fig 1C). This would even be the case if participants used extreme context-dependent encoding: a

163   binary Valence code. Using this mechanism, one learns, for each pair, that one item is 'good' and that

164   one item is 'bad'. That is, one learns the following (separate) sets of orderings: [$Item_1 < Item_2$],

165   [$Item_3 < Item_4$], and [$Item_5 < Item_6$].

166   In the Crossed group (Fig 1C), however, even participants who used less extreme relative

167   encoding strategies may struggle to compare items across contexts, such as $Item_5$ (locally inferior,

168   value of 320) and $Item_2$ (locally superior, value 280). These unexpected and potentially more difficult

169   experiences led participants to be slower in responding (Mann-Whitney-U test, $U = 116$, $p < .001$;

170   Supplementary Results III, V, Fig. S13).

171

172   **Decision-making performance**

173   If people adapt to expected task demands, and the implicit manipulation is sufficient to induce

174   different expectations, the two groups should go on to learn different representations for the

175   subsequent set of items – Learning 2 and Decision 2. Immediately after these task phases, we tested

176   participants' learned representations using two surprise tasks.

177   First, we tested their performance in an All-Pairs task, where all possible pairs of items were

178   presented to both groups (without choice feedback). We found that the Crossed group's choice

179   accuracy was significantly better than the Uncrossed group's despite identical learning Phases ($t(44) =$

180   2.61, $p = .012$, $CI = .026–.199$, $d = .77$, independent t-test) and above chance performance in both

181   groups (Fig. 2A, CIs do not overlap .5, see also Supplementary Results II). The performance

182   difference is consistent with the Crossed group having encoded a more absolute-like value

183   representation (Supplementary Methods I, Fig. S1).

184   Next, we turned to a feature of our experimental design which allowed us to dissociate

185   absolute-like encoding from any relative encoding using 'diagnostic' item pairs. The intuition is that

186   any relative encoding will result in a fraction of choices that are globally inferior, but locally superior

187   within the learning context, whereas an absolute code would not result in the same mistakes. Our task

188   items were chosen to optimize for this (Supplementary Methods I).

189   Specifically, in Phase 2, $Item_2 \sim N(280,28)$ was paired with $Item_1 \sim N(250, 25)$. On the one

190   hand, a relative learner would learn that $Item_2$ is 'good' within its local context. On the other hand,

191 they would learn that both $Item_3 \sim N(300,30)$ and $Item_5 \sim N(320,32)$ are 'bad' – because they were

192 paired with higher-value items. Thus, a relative-value learner would prefer the locally 'good' (but

193 globally inferior) $Item_2$, to the globally superior (but locally 'bad') $Item_{3,5}$: exhibiting irrational

194 choice [see also e.g., 15].

195   In line with these predictions, we found that the Uncrossed group preferred the globally inferior

196 option, choosing it instead of the globally superior options (preferring $Item_2$ to $Item_3$, and to $Item_5$),

197 whereas the Crossed group expressed a weak preference for the globally superior items. The

198 difference between groups was marginal for the first pair ($U = 183$, $p = .055$, $r = .31$), and highly

199 significant for the second pair ($U = 145$, $p = .003$, $r = .45$), by Mann-Whitney U tests.

200   In summary, participants choice behaviour shows that the groups learned different value

201 representations despite identical learning Phases, and that the Crossed group's choices were more

202 consistent with an absolute-like code than the Uncrossed group's (with statistical contrasts

203 specifically selected to discriminate absolute from relative encoding, Supplementary Methods I).

204

205 **Value representation**

206   While the above analyses provide tentative evidence that the groups learned different value

207 representations, we next set out to address value representation more directly. For this purpose,

208 participants were asked to directly indicate their learned value for each item in a Value Judgment task.

209 Items were presented sequentially (in random order), and participants indicated the value using a

210 slider. To test value representation, we applied representational similarity analysis (RSA)[32,33] to this

211 final judgement task (Fig 1A). Note that, although RSA was developed mainly as a multivariate

212 analysis technique for neural data, it is increasingly deployed to characterize brain representations

213 given behavioural data (e.g.,[34–36]) and can be used whenever the measure of interest is pair-wise

214 differences on a univariate or multivariate space.

215   We computed Representational dissimilarity matrices (RDMs) separately for each participant

216 and averaged them to form group-wise RDMs. These RDMs, shown in Fig 3A-D, depict each group's

217 value representation in the form of a dissimilarity structure (rank-transformed and scaled, see

218   Methods and Supplementary Results IV, Fig. S16, for the overall judgments). On this scale, a

219   dissimilarity of 0 implies that item values are represented identically (item pairs along the diagonal),

220   and a dissimilarity of 1 implies that item values are highly dissimilar.

221   Empirical RDMs are most readily interpreted when compared to model RDMs. As noted, the

222   experiment was designed to allow absolute-like codes to be dissociated from relative-like context-

223   dependent codes – irrespective of the precise context-dependent encoding. However, the RSA

224   analysis allows contrasts between different kinds of relative value representation. Thus, to corroborate

225   our results, we contrasted two of the most common relative value encoding models: range adaptation

226   and divisive normalisation. For completeness, we also include a fully contrastive, binary valence

227   model. The higher the correlation between participants' RDMs and the model RDMs – the better the

228   model RDMs describe participants' representation of value.

229   The first relative model ('Valence', Fig 3E) formalizes the extreme 'good vs bad' encoding

230   mentioned in the introduction. The better option in each local context is encoded as 'good' and the

231   worse option as 'bad'. The second relative model ('Range adaptation', Fig 3F) formalizes range-

232   adaptation encoding, a highly successful class of context-dependent encoding schemes[10,16].

233   Accordingly, the value of the left item equals $\frac{item^{left}}{max(item^{left}, item^{right})}$ (and vice versa for the right item).

234   Note that this model scales values within local contexts to the interval $[\frac{min(V)}{max(V)}, 1]$, rather than the

235   interval [0,1]. This is necessary here as with only two items, the full range adaptation model (e.g.[16])

236   would otherwise reduce to the valence model. The third relative model ('Divisive normalisation', Fig

237   3G) formalizes the divisive normalisation encoding highlighted in the Introduction. Here the value of

238   the left item equals $\frac{item^{left}}{1+item^{left}+item^{right}}$ (and vice versa for the right item). We formalize absolute-like

239   context-independent encoding, as the expected value for items. For example, $Item_2$ is encoded as 180

240   because $Item_2 \sim N(180,18)$.

241   As can be seen in Figure 3, the three relative RDMs (E-G) have clusters of items that are

242   objectively similar in value but are nonetheless encoded as highly dissimilar. For example, all the

243   relative models capture the 'irrational' value encoding, by which 280 ($Item_2$) is encoded as more

244   similar to 330 (Item$_4$) than to 300 (Item$_3$). The 'irrational' dissimilarity structure follows from the

245   context-dependent encoding of value formalized in the relative value models.

246   In Figure 3, we first highlight qualitative similarities between the Uncrossed group's RDM

247   and the relative model RDMs (E-G), and between the Crossed group (D) and the Absolute model

248   RDM (H). For example, the items with values 250 and 280 are encoded as highly dissimilar in the

249   Uncrossed RDM (C) - as it is in the relative models (E-G). The Absolute RDM (H), on the other hand

250   correctly encodes this pair as similar, as does the Crossed group RDM (D). This pattern contrasts with

251   the gradient of increasing dissimilarity between 390 and the other items in the Crossed RDM (D). The

252   Uncrossed RDM does not exhibit this gradient (C). Finally, it is clear that both groups encode value in

253   a format that goes beyond mere valence encoding (c.f., Fig 3E, and A-B). Thus, participants in both

254   groups encode and retain at least some value magnitude information.

255   Next, we turned to the key quantitative comparison. We contrasted the correlations between

256   each model and the two groups which takes individual differences into account. As per standard

257   practice[33], model RDMs were compared to the RDMs derived from participants' behaviour using

258   rank-correlations (Methods). A large positive correlation between a participant's RDM and a given

259   model RDM, shows that their representation of value is well accounted for by the model in question.

260   For presentation purposes, we focus on the two relative models that capture key aspects of

261   participants value representation: range-adaptation and divisive normalisation.

262   The Crossed group learned a more absolute value representation than the Uncrossed group:

263   both compared to the Range-adaptation model ($t(44) = 2.97$, $p = .005$, $CI = .15 - .77$, $d = .88$) and the

264   Divisive normalisation model ($t(44) = 2.57$, $p = .014$, $CI = .10 - .85$, $d = .76$).

265   Fig 4A-B plots model-participant RDM similarities expressed as partial Spearman Correlation

266   Coefficients (thus discounting shared variance between models, Methods). Because the Range-

267   adaptation and the Divisive normalisation RDMs were highly correlated, we ran separate analyses

268   contrasting each with the Absolute RDM. Symbols in Fig. 4 reflect group averages, and grey lines

269   reflect individual participants.

270   For the Uncrossed group (A), no model consistently outperforms another, indicated by the

271   mix of slopes. In the Crossed group, however, most participants are substantially better accounted for

272   by absolute encoding (upward sloping lines), indicating that most participants changed their encoding

273   strategy towards an absolute code.

274   Fig 4C shows the within-group contrast between the Absolute model and the two relative

275   models from Fig 4A-B. Positive Δr indicate evidence in favour of the absolute model, and negative

276   indicate evidence in favour of the relative model. As can be seen, no model is consistently favoured in

277   the Uncrossed group (CI's overlap 0). However, in the Crossed group, the absolute model is favoured

278   (CIs do not overlap 0).

279   The previous analyses additionally used a partial correlation approach to rule out the

280   contribution of any shared variance. Fig 4D plots identical analyses, except that they were carried out

281   on independently run correlations. That is, model-by-participant correlations were evaluated

282   independently for the relative encoding models. As can be seen, the results persist with independent

283   correlations. The Crossed group learnt a more absolute value representation than the Uncrossed

284   group: whether one considers the Range-adaptation RDM ($t(44) = 3.23$, $p = .002$, $CI = .21 - .92$, $d = $

285   .95), or the Divisive normalisation RDM ($t(44) = 2.88$, $p = .006$, $CI = .13 - .74$, $d = .85$).

286   Jointly, the results so far show that people 1) adapt their learning to expected task demands

287   (difference between groups despite identical learning Phases), and 2) only learn absolute-like value

288   representations when a relative representation is expected to be insufficient for the task at hand (i.e.,

289   in the 'Crossed' group).

290

291   **Choice and value representation in a binomial task**

292   Next, we turned our focus to a binomial decision task akin to many decision-making tasks in

293   the field of reinforcement learning. Although economic values often come from continuous

294   distributions as in Experiment 1 (e.g., market prices, food quantities, etc.), laboratory tasks often

295   involve binomial outcome distributions[15,37–39]. Next, we therefore sought to establish whether people

296   can also flexibly tune their value-learning mechanism for binomial outcome distributions.

297   As can be seen in Fig 5A, key design features were kept identical to Experiment 1: learning

298   experiences were identical across conditions, Phase 1 was designed to set participants' expectations

299   for Phase 2 in a condition-dependent manner (Crossed vs Uncrossed), and learnt values were assessed

300   in separate surprise tasks (All possible pairs, Value judgement) as before, with the notable exceptions

301   that value distributions were binomial, the number of 'samples' from each distribution was fixed

302   across participants, and the experiment was run online (Methods).

303   Based on Experiment 1, we predicted that the Crossed group would show 1) better All-pairs

304   task performance, 2) improved choice for the single diagnostic item pair in this experiment and 3)

305   more absolute-like value representations – compared to the Uncrossed group. We ran an initial

306   Experiment, which broadly confirmed these predictions, but which was underpowered to find a

307   between-group effect of moderate size. We therefore ran a better powered pre-registered replication

308   on which we report next (see Supplementary Results I for the results of the initial experiment).

309   As can be seen in Fig 5B, choice performance was significantly above chance (CIs do not

310   overlap .5, see also Supplementary Results II) in both groups. As in Experiment 1, the Crossed group

311   made significantly better decisions when choosing between All-pairs following learning (Fig. 5B,

312    $t(222) = 2.30$, $p = .011$, $CI = .011$ - inf, $d = .31$, one-tailed unpaired t-test). Next, we further

313   constrained our comparison to those item pairs for which a divisive normalisation model would make

314   opposing predictions to an absolute value code (see Supplementary Methods II, Fig. S3). Figure 4C

315   shows choice accuracy only for those stimulus-pairs. Even for this restricted analysis, for which

316   choosing is more difficult (differences between values are smaller, Fig. S3A) choice performance was

317   significantly above chance in both groups (non-overlapping CIs, Fig 5C). However, for this sub-

318   selection, the Crossed group again made better decisions than the Uncrossed group ($t(222) = 3.56$, $p <$

319   .001, $CI = .073 – $ inf, $d = .48$). Restricting the analysis further to the single diagnostic stimulus pair

320   (Supplementary Methods II) replicates Experiment 1 (Fig 5D): Crossed group participants chose the

321   higher-value option more frequently than Uncrossed (Fig. 5C; $U = 4722$, $p < .001$, $r = .25$, one-tailed

322   Mann-U Whitney test).

323   Next, we turned our attention again to the RSA analyses. Fig 6A,B show the group-wise

324   average RDMs for Experiment 2. As in Experiment 1, Fig 6C-D highlight similarities between the

325   empirical average RDMs and the model RDMs. As can be seen, and as in Experiment 1, participants'

326   value representation was not consistent with a Valence code (Fig 6E, see Supplementary Results IV,

327   Fig. S17, for the overall judgments).

328   However, as in Experiment 1, the Crossed group learned a more absolute value representation

329   than the Uncrossed group: whether one considers the Range-adaptation RDM ($t(222) = 3.25$, $p < .001$,

330   lower $CI = .17$, upper $CI = $ inf, $d = .43$), or the Divisive normalisation RDM ($t(222) = 3.09$, $p = .001$,

331   lower $CI = .15$, upper $CI = $ inf, $d = .41$).

332   Comparing the empirical RDMs to the finer-grained model RDMs, the 'cross-type' pattern in

333   the two-remaining relative RDMs (Fig 6F-H) is evident in Uncrossed group (Fig. 6C), but largely

334   absent in the Crossed group (Fig. 6D). The latter instead seems to reflect a gradient of dissimilarity

335   approximating the underlying outcome probabilities as in the Absolute Model (.1 vs the remaining

336   item values, Fig. 6C).

337   Next, we turned to our partial correlation analyses, plotted in Fig 7A-B. For the Uncrossed

338   group (A), there was a trend towards the relative models performing better than the absolute model.

339   However, as in Experiment 1, no model consistently outperformed another (mix of sloped lines). In

340   the Crossed group, however, participants were substantially better accounted for by absolute encoding

341   (upward sloping lines); regardless of whether the comparison is a Range adaptation or a Divisive

342   normalisation RDM.

343   Fig. 7C shows the within-group contrast between the Absolute model and the two relative

344   models from the data in Fig. 7A-B. Positive Δr indicate evidence in favour of the absolute model, and

345   negative indicate evidence in favour of the relative model. As can be seen, no model is consistently

346   favoured in the Uncrossed group (though the Range adaptation RDM is close to significant, Fig. 7C, p

347   $= .071$). However, in the Crossed group, the absolute model is clearly favoured (CIs do not overlap 0).

348   Fig. 7D shows an analysis identical to that in Fig. 7C except that it has been carried out on

349   independently run correlations. As can be seen, the results persist with independent correlations. The

350   Crossed group learned a more absolute value representation than the Uncrossed group: whether one

351   considers the Range-adaptation RDM ($t(222) = 3.01$, $p = .002$, lower $CI = .05$, upper $CI = $ inf, $d = $

352   .40), or the Divisive normalisation RDM ($t(222) = 3.15$, $p < .001$, lower $CI = .02$, upper $CI = $ inf, $d = $

353   .42).

354   In summary Experiment 2 replicated and generalised the results of Experiment 1, using an

355   online study with binomial outcome distributions. Choice task data showed that the Crossed group

356 learned a different value representation than the Uncrossed group – despite identical learning

357 experiences. The choices in the Crossed group were on average better than those in the Uncrossed

358 group and were better specifically for item pairs for which an absolute-like representation will result

359 in improved choice. The RSA analyses further show that the Crossed group learned an absolute-like

360 representation, and that they learned a more absolute-like representation than the Uncrossed group.

361

362 **Discussion**

363 We sought to reconcile the theoretical and empirical tension between two diametrically

364 opposing accounts of value learning and encoding: a context-independent but potentially

365 computationally costly absolute value representation[1,2,4], and an efficient local, but potentially

366 irrational, relative value representation [7,11,13,15]. We proposed that humans (and possibly other animals)

367 do not use a single fixed mechanism – learning either absolute or relative value codes – but adapt their

368 learning to expected task demands in an efficient and rational manner: learning sufficient and

369 necessary value representations.

370 We tested this hypothesis in two human value-learning experiments: one involving normally

371 distributed values and the other involving binomial outcomes. In each study, the first Phase was

372 equivalent to the full experience of participants in many experimental paradigms[(e.g., 15,37)]. The second

373 Phase gave participants the chance to use their prior experience with the task to tune their learning

374 mechanism to optimise task performance. Phase 2 thus mimicked the opportunity to adapt and tune

375 learning mechanisms that arise in many real-life tasks (and which are performed more than once).

376 Despite identical learning experiences, the two groups learned different value codes.

377 Specifically, across the two studies, the Crossed group made decisions that are consistent with a

378 higher-fidelity representation (Figs. 3,6), made fewer irrational choices (Fig. 2,5), and learned value

379 representations that were more absolute-like than the Uncrossed group (Fig. 4,7). Importantly,

380 participants consistently learned more absolute representations only when it was expected to be

381 useful. Thus, people do not learn either absolute or relative value codes but adapt their learning to

382 what they expect to use the code for.

383   Nevertheless, the reliable group differences were not always reflected at the individual level.

384   In the Uncrossed condition, many participants appeared to have learnt absolute-like codes. This may

385   be driven by the fact that both absolute and relative codes yield good results for the Uncrossed group.

386   Thus, whichever code participants favour as their "default" would be expected to persist. In the

387   Crossed condition some participants appeared to have learnt relative codes. This may be driven by

388   different factors beyond the scope of our current study: by cognitive capacity limitations[40], intrinsic

389   computational noise[41,42], or by mechanisms relating to working memory or attention[43,44]. Future work

390   might manipulate task demands and difficulty[38,45, (c.f. 46)] to address individual differences in value

391   encoding.

392   A second outstanding question is the learning mechanisms that give rise to the flexible and

393   adaptive value representations we observe. Our studies were designed for well-controlled

394   measurement of value representation following learning. The trade-off is that the design is not

395   effective in characterizing learning mechanisms themselves - as opposed to the codes they give rise

396   to. Nevertheless, our design allowed us to successfully recover the relative and absolute models in

397   simulations, thus supporting our key RDM contrasts (Supplementary Results VII).

398   It is possible that a single mechanism underlies the observed flexibility in value encoding.

399   Such a mechanism could be implemented with a free parameter governing the extent to which

400   learning is relative in the Divisive normalisation model, such that $item^{left} = \frac{item^{left}}{1+w*(item^{left}+item^{right})}$,

401   where w is a free parameter between 0 (for wholly absolute encoding) and 1 (for wholly relative

402   encoding). However, it is also possible that mechanisms rely at least in part on different cognitive

403   substrates as in, for example, model-based and model-free learning[47–49]. Future work is needed to

404   address the question of mechanism, and perhaps more importantly mechanism selection, which likely

405   requires higher-level cognition and monitoring of expectations. Our experiments were designed to

406   tests for coarse between-group differences in encoding, allowing us to ask: Is value encoding

407   adaptive, and if so – is it rationally adaptive? Thus, further work is needed to allow more precise

408   classification of learning and encoding mechanisms.

409   Finally, extrapolating beyond the behavioural data at hand, one might reasonably expect that

410   relative values behave in a way similar to "cached" values in Reinforcement Learning, in the sense

411   that they incorporate context into their code (without later being able to retrieve context values),

412   whereas absolute-like encoding may rely on memory systems that separate item and context

413   representations, allowing the system to flexibly combine them at decision time. In this sense, one

414   might expect the absolute-like representation to preferentially recruit hippocampal-medial prefrontal

415   circuits, whereas relative encoding may rely more heavily on striatal-prefrontal circuits, as

416   approximately in the model-free / model-based distinction in RL[49]. However, further research is

417   needed to identify the neural mechanisms arbitrating between the two encodings.

418   In summary, our results highlight the highly dynamic and rational nature of value

419   representation: humans do not simply have a single, fixed form of representation, but rather adjust

420   their code in a rational[50–52] manner according to expected task demands. Further, our findings

421   highlight that both absolute and relative codes previously found can potentially be explained by the

422   fact that participants infer which code would be sufficient for the current task.

423

424   **Methods**

425   **Experiment 1 - Participants**

426   The study complied with all relevant ethical regulations and was approved by the local ethics

427   committee at City, University of London. Sixty participants (37 female) were recruited via the local

428   participation panel. Participants provided written informed consent and were debriefed. Participants

429   had normal, or corrected-to-normal, vision, were fluent in English, healthy (no known physical or

430   psychological conditions), and between 18-45 years old. No statistical methods were used to pre-

431   determine sample sizes, but our sample sizes are similar to those in previous work[10,15].

432   Participants were reimbursed for their time and were paid a performance-related bonus: a base

433   pay of £5 and an additional bonus between £0 and £6. The average bonus was for a total of £2.78

434   (range £0-6). The performance bonus was determined by choice performance across all Decision

435   Phases as well as during the final two tasks. The greater the number of high-value choices, the greater

436   the bonus, and the closer the judgement to the true item value the higher the bonus.

437   We excluded participants who did not fulfil minimal task requirements. Criteria apply to the

438   Learning phases only (Fig 1A,B), and are therefore orthogonal to the target behaviour in the final

439   tasks (Fig 1A). Exclusion criteria were based on 1) sampling behaviour and 2) below-chance

440   performance for the preliminary decisions in the first sampling phase. Participants who only sampled

441   once (or fewer times), per item per item-pair sampling opportunity, were excluded (Learning 1-2, Fig

442   1). This cut-off represents <= 18 samples per Context and is far lower than the median of 123

443   (IQR=118) and 143 (IQR=91) for Phase 1 and 2 respectively. There were 9 preliminary decisions in

444   the first Phase (3 pairs presented three times each, Fig 1A,B). Someone who responded randomly

445   when making these decisions, would be expected to achieve a choice accuracy between .22 and .78

446   (with a mean choice accuracy of .5). This range reflects the lower and upper 95% confidence interval

447   on a hypothetical agent who responds randomly (i.e., selects each option with p = .5). Participants

448   who performed worse than the upper confidence interval (i.e., did not achieve at a greater choice

449   accuracy than expected by chance) were excluded.

450   In summary, we excluded participants who showed no or little evidence of learning – a pre-

451   condition for encoding value (whether in an absolute or relative form). In total, fourteen participants

452   met one or both exclusion criteria for a final sample size of n=46: 24 of which had been assigned to

453   the Uncrossed condition, and 22 of which had been assigned to the Crossed condition.

454   **Experiment 1 - Materials**

455   Participants took on the role of a consultant to a manufacturer of reproduction items in two

456   different contexts (antiques/cars, Fig 1). The item-values and item-pairs were Phase-specific (Fig 1A).

457   However, the mapping of item type (antiques/cars) to Phase, the mapping of specific items (e.g.,

458   typewriter) to item-values (e.g., $N(180,18)$), and the side on which items were presented during

459   sampling, were all randomized across participants.

460   Item-values (Fig 1A) were selected primarily so that absolute-value and relative-value

461   representations dissociate (Supplementary Methods I-II, Fig. S1-3), and secondarily to achieve a

462   balance between task-difficulties in the Learning and Decision phases (Supplementary Methods III,

463   Fig. S4-7). A single sample from one item resulted in a draw from the corresponding normal value

464   distribution (truncated at ±2 SD). The Learning phases (Fig 1A,B,D) were self-paced, and participants

465 had a wide range of different strategies as evidenced by the wide range of the number of samples

466 drawn (range Phase 1: min=48, max=478: range Phase 2; min=32, max=509).

467   The Decision phases (Fig 1A,C,D) involved 18 decisions per Phase. The Uncrossed group

468 decided between the pairs they had experienced during sampling (repeated 6 times = 18 decisions).

469 The Crossed group made decisions between novel pairs (6 novel pairs x 2 = 12 decisions, see Fig 1

470 for examples), in addition to learnt pairs (3 pairs x 2 = 6 decisions).

471   The two final tasks (Fig 1A,E) were identical across groups. The All-possible pairs task

472 involved 15 pairs, representing a full factorial combination of all possible pairs from Phase 2

473 (excluding identical pairs), repeated three times for a total of 45 pairs. The Value judgment task

474 involved the 6 items in Phase 2, presented one at a time along with a slider-interface (min=100,

475 max=450). For all tasks, the presentation order and presentation side (where applicable) were

476 randomized across participants.

477 **Experiment 1 Procedure**

478   Participants read the information sheet, provided written informed consent, and completed the

479 tasks. After completing the behavioural tasks, participants completed three questionnaires. These

480 formed parts of one author's MSc dissertation project and are not reported on here.

481 **Experiment 1 Apparatus**

482   Stimuli were displayed on a touchscreen (Ilyama T2245MSC) and code was written in

483 MATLAB (Mathworks) using PsychToolbox[53] on Linux (Xubuntu 18.04) with a soft real-time kernel.

484 **Experiment 2 Participants**

485   The study was approved by the local ethics committee at City, University of London.

486 Participants were recruited via Prolific Academic, fully informed, provided written informed consent

487 and were debriefed. Participants were between the age of 18 and 40, were UK residents, were healthy

488 (no ongoing mental health conditions, dementia/mild cognitive impairment, no daily impact of mental

489 illness), had not participated in similar studies of ours, had a minimum approval rate on Prolific of 99

490 and minimum of 10 submissions. We sought to include a minimum of 280 participants, conditional on

491 having at least 100 participants in each condition passing post-completion exclusion criteria. The

492 sample size was determined based on power calculations, which in turn were based on the pilot study

493   (*Supplementary Results I*). Power calculation, exclusion criteria, and sampling strategy were all pre-

494   registered (https://osf.io/xjsmh).

495   Online panels provide little experimental control and the potential for poor participant

496   engagement (see also discussion in Supplementary Results VI). To minimise this issue, we employed

497   an initial check that participants had read and understood task instructions. To be eligible, potential

498   participants had to answer 8 multiple-choice questions correctly. In addition, participants were

499   allowed to make only one error in the first Decision block for the stimuli they had just learnt about.

500   Specifically, if after experiencing 10 learning trials per item-pair, participants were unable to choose

501   the higher value items 2 out of the first 3 presentations the study ended prematurely, and participants

502   pay was pro-rated. We chose to allow 1 error as even engaged participants might be expected to make

503   mistakes especially for the more difficult stimulus pair (.8 vs .9). In total 888 participants expressed

504   interest and 352 completed the full study. Most non-completers (92%) failed the initial knowledge

505   test.

506   Participants were reimbursed for their time and were paid a performance-related bonus.

507   Participants were paid a base pay of £2.92 for participation (the experiment took ~35 mins) and an

508   additional bonus between £0 and £2.92. The average bonus was £1.46 (range £0.50–£2). The

509   performance bonus was determined by choice performance across all Learning, Decision Phases and

510   the final two tasks. Correct choices in the Decision phases and the All-pairs task were weighted x10

511   compared to Learning. This was done to encourage participant engagement for the tasks which did not

512   involve feedback. In general, the reward structure was as in Experiment 1 in that the greater the

513   number of high-value choices, the higher the bonus, and the closer the judged value to the true item

514   value, the higher the bonus.

515   In addition to the pre-registered *a priori* exclusions, we also employed pre-registered exclusion

516   criteria based on participants' not fulfilling minimal task performance criteria after completing the full

517   study. Because each participant experienced the same number of trials, sampling behaviour cannot be

518   used for excluding disengaged participants (unlike in Experiment 1). Instead, we excluded

519   participants who did not learn to choose among the pairs experienced during Learning. All

520   participants were trained on the following binomial probability pairs: [.1 vs .6], [4 vs .7] and [.8 vs

521 .9.] - irrespective of condition. We excluded participants who made more than two errors in two

522 repeats of these three pairs (i.e., more than 2/6 errors) at the end of the experiment (in the All-pairs

523 task). In other words, we include only participants who showed evidence of encoding these learning

524 phases for later recall. Note that these exclusion criteria are orthogonal to the question of absolute and

525 relative value codes. Both absolute and relative models of learning will allow participants to learn to

526 choose between the items in the Learning phase. In other words, choices between items of pairs that

527 participants directly learned about - unlike novel combinations of the component items - are not

528 diagnostic with regards to value representation.

529 Applying these exclusion criteria, which are orthogonal to which model participants may use to

530 encode value, leaves N=224 participants of which n=119 participants were from the Uncrossed group

531 and n=105 were from the Crossed group. That is, it resulted in the exclusion of ~36% of participants.

532 We report analyses also including these excluded participants in Supplementary Results VI and note

533 that these analyses replicate those reported in the main text.

534 **Experiment 2 Materials and Procedure**

535 Experiment 2 was a pre-registered version of a previous study (https://osf.io/xjsmh). As in

536 Experiment 1, participants took on the role of a consultant to a manufacturer of reproduction items in

537 two different contexts (antiques/cars, Fig 4). Key design features were identical to Experiment 1.

538 However, outcomes were binomial (successful sale/unsuccessful sale), the task was not self-paced,

539 and the learning experience was not 'blocked' by item-pairs (item-pairs were randomly intermixed

540 during learning) and involved a relatively rapid stimulus display sequence.

541 In the Learning Phases, participants saw each item pair presented side-by-side (~1 sec),

542 followed by a response phase in which participants had ~1.5 second to make a choice, followed by

543 sequential feedback, in which the chosen item was presented first followed by the unchosen item.

544 Outcome feedback was in the form of a green double-rectangle image outline (successful sale) or a

545 single-rectangle red image outline (unsuccessful sale).

546 Experienced outcomes matched the expected outcome of the binomial distributions (Fig 4A).

547 This was achieved by pre-allocating and shuffling an outcome vector (of 1's and 0's) for each item.

548 This design minimizes the impact of sampling error[54] on differences between participants and/or

549   conditions. There were two learning blocks per Learning Phase. In each block each pair was presented

550   10 times, for a total of 30 trials per block and 60 trials per Phase. The presentation order was

551   randomized.

552   Each of the two blocks of Decision trials (two for each Learning Phase), involved 12 decisions

553   without feedback. The Uncrossed group made decisions between pairs experienced during learning (3

554   pairs x 4). In addition to experienced pairs (3 pairs presented once), the Crossed group made decisions

555   also between novel pairs composed of items from different learning pairs (9 novel pairs, randomized

556   across participants). Thus, each group experienced 24 Decision trials per Learning Phase.

557   The two final tasks (Fig 1A,E) were identical across groups. The All-possible pairs task

558   involved 15 pairs, representing a full factorial combination of all possible pairs from Phase 2

559   (excluding identical pairs), repeated twice (controlling for presentation side) for a total of 30 pairs.

560   The Value judgment task involved the 6 items in Phase 2, presented one at a time along with a slider-

561   interface (min=0%, max=100%) representing the probability of an item selling. The presentation

562   order and presentation side (where applicable) were randomized across participants for all tasks.

563   **Design and Statistical Analyses – Experiment 1 & 2**

564   Both experiments used a between-subject design with participants assigned randomly and

565   blindly to one of two conditions: Uncrossed and Crossed. Our analyses focus on differences between

566   the two groups for the two final tasks and within-task contrasts against reference magnitudes.

567   The primary inferential statistic was the t-test. T-tests are relatively robust and were used

568   whenever feasible. For data with clear deviations from parametric assumptions (e.g., Fig 2B), less

569   powerful non-parametric tests were used. To rule out potential limits to t-test robustness affecting

570   inferences we also ran all our t-tests reported here using non-parametric tests (all resulting in the same

571   conclusion as the t-test). We also report 95% CIs (parametric or bootstrapped) for all descriptive

572   statistics here. CIs can be used for inference by comparing them to reference magnitudes. For

573   example, if the mean choice accuracy is above .5, and the 95% CI of that mean does not overlap .5,

574   choice performance was significantly greater than chance.

575   All reported tests for Experiment 1 are two-tailed. Predictions for Experiment 2 were pre-

576   registered (https://osf.io/xjsmh) and derived from results from Experiment 1, and the initial pilot

577   version of Experiment 2 (Supplementary Results I), and all between-group contrasts were one-tailed.

578   Reported effect sizes are Cohen's d for t-tests (>= .2 small, >= .5 medium, >= .8 large) and rank-

579   biserial correlation r for non-parametric tests (>= .1 small, >= .3 medium, >= .5 large).

580   Standard RSA protocols[33] were followed. Empirical value RDMs were computed as the

581   Euclidean distance between each participant's value judgements. Average RDMs were computed by

582   averaging (arithmetic mean) over participants' RDMs separately for each group. Model RDMs were

583   computed as the Euclidean distance between item values defined by the relevant model equations

584   (Main Text). For display purposes RDMs were rank-transformed (equal stays equal) and scaled to 0-

585   1, where 0 implies identical item-values and 1 means maximally dissimilar item-values.

586   We computed the similarity between model RDMs and participant RDMs by partial correlation

587   (Spearman). Partial correlation accounts only for unique variance. This means that a correlation

588   between one model RDM and a participant's RDM cannot be explained by the second model RDM.

589   Because our interest lay in dissociating absolute from relative encoding (not distinguishing between

590   various relative models), and because relative models were highly correlated, we performed these

591   analyses separately for each contrasting relative model (Fig 4 & 7). We also performed analyses with

592   independent correlations (i.e., any shared variance between models is not taken into account). Like the

593   partial correlation analyses, these used the Spearman correlation coefficient. For all correlational

594   analyses, large positive r's imply a high degree of similarity between participants value representation

595   and model RDMs (and $r = 0$ implies no relationship).

596   For the key statistical analysis, to establish whether the evidence in favour of the absolute

597   model over the relative model was greater in the Crossed group than the Uncrossed group, we

598   computed the difference between the absolute and the relative models independently for each group

599   and contrasted those differences with t-tests. A positive difference in r indicates evidence in favour of

600   absolute encoding and a negative difference in r indicates evidence in favour of relative encoding.

601   These differences can also be used to infer whether there was a tendency to favour relative or absolute

602   encoding within each group by contrasting the 95% CIs of those averages differences to 0.

603   All statistical analyses were performed in MATLAB 2021b (MathWorks).

604

605

Data availability

Data is available online on OSF (https://osf.io/h32u6/).

608

Code availability

Analysis code is available on OSF (https://osf.io/h32u6/).

611

References

1. Morgenstern, O. & Von Neumann, J. *Theory of games and economic behavior*. (Princeton university press, 1953).

2. Stephens, D. W. & Krebs, J. R. *Foraging theory*. vol. 1 (Princeton University Press, 1986).

3. Sutton, R. S., Barto, A. G., & others. *Introduction to reinforcement learning*. vol. 135 (MIT press Cambridge, 1998).

4. Tversky, A. & Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* **5**, 297–323 (1992).

5. Olsen, S. R., Bhandawat, V. & Wilson, R. I. Divisive Normalization in Olfactory Population Codes. *Neuron* **66**, 287–299 (2010).

6. Heeger, D. J. Normalization of cell responses in cat striate cortex. *Vis Neurosci* **9**, 181–197 (1992).

7. Khaw, M. W., Glimcher, P. W. & Louie, K. Normalized value coding explains dynamic adaptation in the human valuation process. *Proc Natl Acad Sci USA* **114**, 12696–12701 (2017).

8. Karni, E., Schmeidler, D. & Vind, K. On State Dependent Preferences and Subjective Probabilities. *Econometrica* **51**, 1021 (1983).

9. Pompilio, L., Kacelnik, A. & Behmer, S. T. State-Dependent Learned Valuation Drives Choice in an Invertebrate. *Science* **311**, 1613–1615 (2006).

10. Bavard, S., Rustichini, A. & Palminteri, S. Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.* **7**, eabe0340 (2021).

632 11. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat Rev*

633   *Neurosci* **13**, 51–62 (2012).

634 12. Normann, R. A. & Werblin, F. S. Control of retinal sensitivity: I. Light and dark adaptation of

635   vertebrate rods and cones. *The Journal of general physiology* **63**, 37–61 (1974).

636 13. Stewart, N., Chater, N. & Brown, G. D. A. Decision by sampling. *Cognitive Psychology* **53**, 1–26

637   (2006).

638 14. Yamada, H., Louie, K., Tymula, A. & Glimcher, P. W. Free choice shapes normalized value

639   signals in medial orbitofrontal cortex. *Nat Commun* **9**, 162 (2018).

640 15. Klein, T. A., Ullsperger, M. & Jocham, G. Learning relative values in the striatum induces

641   violations of normative decision making. *Nat Commun* **8**, 16033 (2017).

642 16. Palminteri, S. & Lebreton, M. *Context-dependent outcome encoding in human reinforcement*

643   *learning*. https://osf.io/4qh2d (2021) doi:10.31234/osf.io/4qh2d.

644 17. Rigoli, F. Reference effects on decision-making elicited by previous rewards. *Cognition* **192**,

645   104034 (2019).

646 18. Rigoli, F., Mathys, C., Friston, K. J. & Dolan, R. J. A unifying Bayesian account of contextual

647   effects in value-based choice. *PLoS Comput Biol* **13**, e1005769 (2017).

648 19. Ciranka, S. *et al. Asymmetric learning facilitates human inference of transitive relations*.

649   http://biorxiv.org/lookup/doi/10.1101/2021.04.03.437766 (2021)

650   doi:10.1101/2021.04.03.437766.

651 20. Gluth, S., Kern, N., Kortmann, M. & Vitali, C. L. Value-based attention but not divisive

652   normalization influences decisions with multiple alternatives. *Nature Human Behaviour* **4**, 634–

653   645 (2020).

654 21. Rustichini, A., Conen, K. E., Cai, X. & Padoa-Schioppa, C. Optimal coding and neuronal

655   adaptation in economic decisions. *Nat Commun* **8**, 1208 (2017).

656 22. Bhui, R. & Gershman, S. J. Decision by sampling implements efficient coding of

657   psychoeconomic functions. *Psychological Review* **125**, 985 (2018).

658 23. Polanía, R., Woodford, M. & Ruff, C. Efficient coding of subjective value. *Nature Neuroscience*

659   **22**, 134–142 (2019).

660 24. Pompilio, L. State-Dependent Learned Valuation Drives Choice in an Invertebrate. *Science* **311**,

661   1613–1615 (2006).

662 25. Kool, W., Gershman, S. J. & Cushman, F. A. Cost-Benefit Arbitration Between Multiple

663   Reinforcement-Learning Systems. *Psychol Sci* **28**, 1321–1333 (2017).

664 26. Griffiths, T. L. *et al.* Doing more with less: meta-reasoning and meta-learning in humans and

665   machines. *Current Opinion in Behavioral Sciences* **29**, 24–30 (2019).

666 27. James, W. *The principles of psychology*. vol. 1 (Henry Holt & Co, 1890).

667 28. Anderson, J. R. *The adaptive character of thought*. (Psychology Press, 2013).

668 29. Payne, J. W., Bettman, J. R. & Johnson, E. J. *The adaptive decision maker*. (Cambridge

669   university press, 1993).

670 30. Anderson, J. (1990). The adaptive character of thought. *Hillsdale, NJ: Erl-baum*.

671 31. Kahneman, D. & Tversky, A. Prospect Theory: An Analysis of Decision under Risk.

672   *Econometrica* **47**, 263–292 (1979).

673 32. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping.

674   *Proceedings of the National Academy of Sciences* **103**, 3863–3868 (2006).

675 33. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol* **10**,

676   e1003553 (2014).

677 34. Luyckx, F., Nili, H., Spitzer, B. & Summerfield, C. Neural structure mapping in human

678   probabilistic reward learning. 1–16 (2018) doi:10.1101/366757.

679 35. Sheahan, H., Luyckx, F., Nelli, S., Teupe, C. & Summerfield, C. *Neural state space alignment for*

680   *magnitude generalisation in humans and recurrent networks*.

681   http://biorxiv.org/lookup/doi/10.1101/2020.07.22.215541 (2020)

682   doi:10.1101/2020.07.22.215541.

683 36. Hunt, L. T. *et al.* Triple dissociation of attention and decision computations across prefrontal

684   cortex. *Nature neuroscience* **21**, 1471–1481 (2018).

685 37. Hertwig, R., Barron, G., Weber, E. U. & Erev, I. Decisions From Experience and the Effect of

686   Rare Events in Risky Choice. *Psychological Science* **15**, 6 (2004).

687   38. Bavard, S., Rustichini, A. & Palminteri, S. The construction and deconstruction of sub-optimal

688   preferences through range-adapting reinforcement learning.

689   39. Hotaling, J. M., Jarvstad, A., Donkin, C. & Newell, B. R. How to Change the Weight of Rare

690   Events in Decisions From Experience. *Psychol Sci* **30**, 1767–1779 (2019).

691   40. Shenhav, A. *et al.* Toward a rational and mechanistic account of mental effort. *Annual review of*

692   *neuroscience* **40**, 99–124 (2017).

693   41. Prat-Carrabin, A. & Woodford, M. *Efficient coding of numbers explains decision bias and noise.*

694   http://biorxiv.org/lookup/doi/10.1101/2020.02.18.942938 (2020)

695   doi:10.1101/2020.02.18.942938.

696   42. Juechems, K., Balaguer, J., Spitzer, B. & Summerfield, C. Optimal utility and probability

697   functions for agents with finite computational precision. *Proceedings of the National Academy of*

698   *Sciences* **118**, (2021).

699   43. Spektor, M. S., Gluth, S., Fontanesi, L. & Rieskamp, J. How similarity between choice options

700   affects decisions from experience: The accentuation-of-differences model. *Psychological Review*

701   **126**, 52–88 (2019).

702   44. Collins, A. G. E. & Frank, M. J. How much of reinforcement learning is working memory, not

703   reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working

704   memory in reinforcement learning. *European Journal of Neuroscience* **35**, 1024–1035 (2012).

705   45. Hayes, W. M. & Wedell, D. H. Regret in experience-based decisions: The effects of expected

706   value differences and mixed gains and losses. *Decision* **8**, 277–294 (2021).

707   46. Edwards, D. J., Pothos, E. M. & Perlman, A. Relational versus absolute representation in

708   categorization. *Am J Psychol* **125**, 481–497 (2012).

709   47. Collins, A. G. E. & Cockburn, J. Beyond dichotomies in reinforcement learning. *Nat Rev*

710   *Neurosci* **21**, 576–586 (2020).

711   48. Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J. & Daw, N. D. Predictive

712   representations can link model-based reinforcement learning to model-free mechanisms. *PLoS*

713   *Comput Biol* **13**, e1005768 (2017).

714 49. Koechlin, E. Prefrontal executive function and adaptive behavior in complex environments.

715 *Current Opinion in Neurobiology* **37**, 1–6 (2016).

716 50. Gershman, S. J., Horvitz, E. J. & Tenenbaum, J. B. Computational rationality: A converging

717 paradigm for intelligence in brains, minds, and machines. *Science* **349**, 273–278 (2015).

718 51. Hunter, L. E. & Gershman, S. J. *Reference-dependent preferences arise from structure learning*.

719 http://biorxiv.org/lookup/doi/10.1101/252692 (2018) doi:10.1101/252692.

720 52. Lieder, F., Shenhav, A., Musslick, S. & Griffiths, T. L. Rational metareasoning and the plasticity

721 of cognitive control. *PLoS Comput Biol* **14**, e1006043 (2018).

722 53. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).

723 54. Fox, C. R. & Hadar, L. "Decisions from experience" = sampling error + prospect theory:

724 Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making* **1**, 3

725 (2006).

726

727   **Acknowledgements**

733

734   **Author Contributions**

735 KJ, AJ designed the research; TA, RH conducted the research; AJ analysed the data; KJ, AJ

736 contributed materials/analysis tools and wrote the paper.

737

738   **Competing Interests**

739 The authors declare no competing interests.

740

741   **Figure legends**

Figure 1. Experiment 1 Design and Tasks. (A) Each participant was double-blindly assigned to either the Uncrossed (green) or the Crossed (blue) group. There were two Phases, which were structurally identical, but with different market values and item types. The mapping between item types and context was randomized across participants, as was the item-value mapping, and item type was counterbalanced. In each Phase, participants first learnt market values of 6 items (antiques or vintage cars) arranged into 3 pairs. The notation in the panel indicates the normal value distributions from which experienced samples were drawn: $N(M,SD)$ where M is the mean and SD the standard deviation. Samples were truncated at ±2SD to avoid potential extreme outlying values (A, see also Supplementary Methods III). Participants learnt by sampling (B). A click on an item returned a single sample. Participants were free to sample as much as they wished. Sampling for a given pair ended once a preliminary selling decision was made. There were three sampling phases for each item-pair (three preliminary decision/item). Learning was followed by Decision (C), in which participants made decisions without feedback. The Uncrossed groups made decisions about previously sampled item-pairs. The Crossed group also made decisions between novel item-pairings, composed of items from different item-pairs. We predicted that the expectations induced by Decision in Phase 1 would cause value learning mechanisms to diverge across groups in Phase 2 (D). Phase 2 learnt values were assessed in two 'surprise' final tasks: In All-possible pairs (E) participants made decisions between all possible pairs from Phase 2 (N=15, repeated thrice for N=45). In Value judgment (F), participants judged the value the value of the six stimuli in Phase 2 presented in a random order by adjusting a slider (min=100, max=450, in integer steps) until it matched the perceived item value.

Figure 2. Experiment 1 All-pairs choice accuracy. (A) Choice accuracy as a function of group. Coloured symbols represent group means (green square = Uncrossed; blue triangle = Crossed). Grey discs represent individual participants (Uncrossed N=24; Crossed N=22). Error bars are 95% CIs. Statistics reflect the group-wise contrast $t(44) = 2.61$, $p = .012$, $CI = .026–.199$, $d = .77$, independent t-test. (B) Choice accuracy for a sub-selection of highly diagnostic pairs, in which a local high-value item ($Item_2$) was globally inferior to other local low-value items ($Item_3$, $Item_5$). Error bars are bootstrapped 95% CIs. P-values reflect Mann-Whitney U tests: $U = 183$, $p = .055$, $r = .31$ and $U = 145$, $p = .003$, $r = .45$ respectively. X-axis coordinates of participants' data have been jittered for presentation purposes.
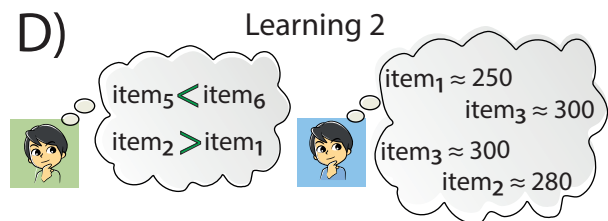
Figure 3. Experiment 1 Value RDMs. Average RDMs for the Uncrossed group (A) and the Crossed group (B). Note that items are ordered by the underlying value (not item number). Average RDMs (C,D) but with pair-wise similarities matching those of different models (E-H) highlighted. Model RDMs (E-H). The colour scale indicates rank-transformed and rescaled dissimilarity (see Methods, 0=minimal dissimilarity, 1=maximal dissimilarity).

Figure 4. Experiment 1 RDM correlations. Partial Spearman participant x model correlations for the Uncrossed (A, green squares, N=24) and Crossed group (B, blue triangles, N=22) respectively. Each panel (A,B) shows two analyses: one in which range-adaptation is pitted against absolute encoding, and another in which divisive normalisation is pitted against absolute encoding. The larger the r the better the model accounts for participants' value representation. Symbols indicate group means and error bars reflect 95% CIs. Grey lines represent

individual participants. Downwards sloping lines (from left to right) indicate that participants' representation of value is better modelled as relative. Upward sloping lines (from left to right) indicate that the participants' value code is better accounted for by an absolute code. (C) Mean participant x Model correlation differences (participant x Absolute *r* – participant x Relative *r*). Positive *r*'s indicate that the absolute model fits better and negative *r*'s that the relative model fits better. Symbols reflect means and error bars reflect 95% CIs. The reported p-values reflect group-wise contrasts, which assess whether the evidence in favour of the absolute model over the relative model was stronger in the Crossed group: $t(44) = 2.97$, $p = .005$, $CI = .15 - .77$, $d = .88$) and $t(44) = 2.57$, $p = .014$, $CI = .10 - .85$, $d = .76$ respectively. (D) As (C) but for independent correlations. The p-values reflect key across-group contrasts $t(44) = 3.23$, $p = .002$, $CI = .21 - .92$, $d = .95$ and ($t(44) = 2.88$, $p = .006$, $CI = .13 - .74$, $d = .85$).

Figure 5. Experiment 2 Design and All-pairs choice accuracy. Key design features of Experiment 2 were identical to Experiment 1. Each participant was assigned (double-blind) to either the Uncrossed (green colour) or the Crossed (blue colour) groups. There were two Phases, which were structurally identical, but with different market values and item types. In each phase, participants first learnt the likelihood that an item would sell of 6 items (antiques or vintage cars, order counterbalanced across participants) arranged into 3 pairs. Values were matched to the expected outcomes of binomial distributions ($B(N, p)$), where p is the probability of observing a sale on a single trial ($N=1$). Values were matched such that with $p=.1$, for example, participants would observe a successful sale on 2 out of 20 trials (Methods, see also Supplementary Methods III). Learning was followed by Decision, in which participants made consequential decisions without feedback. The Uncrossed groups made decisions about previously sampled item-pairs. The Crossed group made decisions between novel item-pairings, composed of items from different previously sampled item-pairs. (B) All-pairs choice accuracy as a function of group. Coloured symbols represent group means (Uncrossed=green square, N=119; Crossed=blue triangle, N=105). Error bars are 95% CIs. Gray dots represent individual participants. The p-value reflects a one-tailed independent t-test $t(222) = 2.30$, $p = .011$, $CI = .011 - inf$, $d = .31$. (C) Sub-set of All-pairs trials for which Divisive normalisation and Absolute encoding make different predictions (see Supplementary Methods III). Coloured symbols represent group means (Uncrossed=green square, Crossed=blue triangle). Error bars are 95% CIs. Gray dots represent individual participants. The p-value reflects a one-tailed independent t-test $t(222) = 3.56$, $p < .001$, $CI = .073 – inf$, $d = .48$ (D) The single All-pairs stimulus-pair for which strong context-dependent encoding would result in different choices compared to absolute value encoding. Error bars are bootstrapped 95% CIs. Gray dots represent individual participants. For (D) participants could either make 0, 1 or 2 errors. The p-value reflects a one-tailed Mann-Whitney U, $U = 4722$, $p < .001$, $r = .25$. X-axis coordinates of participants' data have been jittered for presentation purposes.
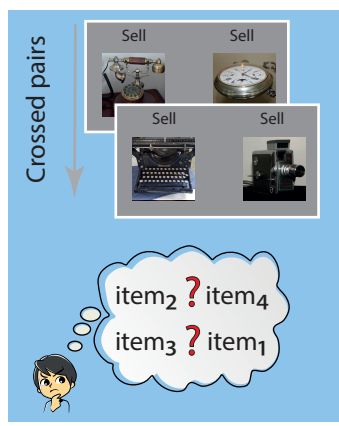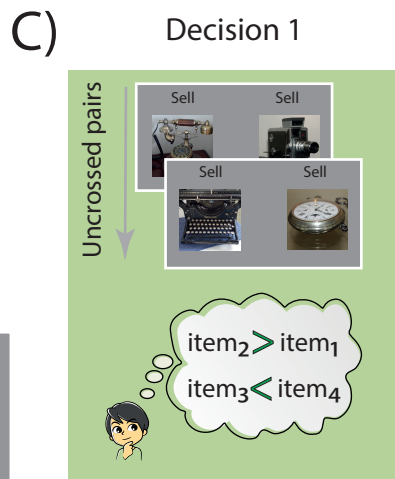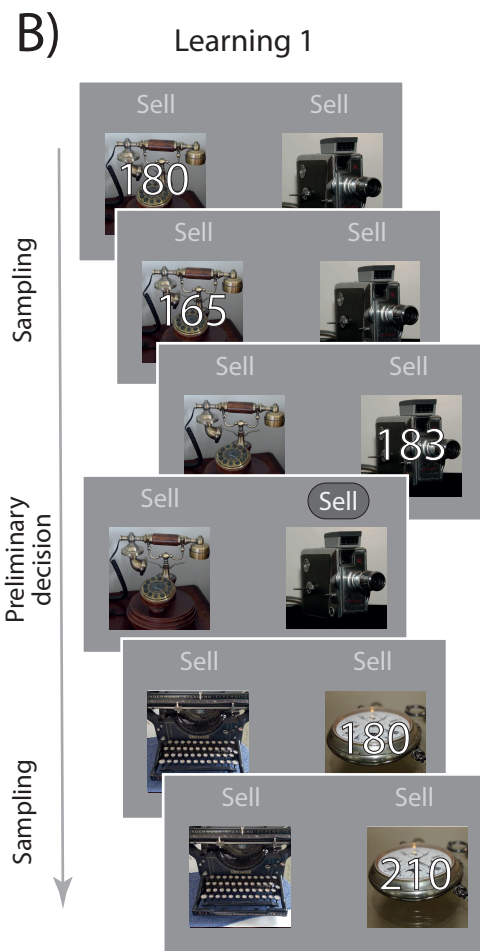
Figure 6. Experiment 2 Value RDMs. Average RDMs for the Uncrossed group (A) and the Crossed group (B). Note that items are ordered by the underlying value (not item number). Average RDMs (C,D) but with pair-wise similarities matching those of different models (E-H) highlighted. Model RDMs (E-H). The colour scale indicates rank-transformed and rescaled dissimilarity (see Methods, 0=minimal dissimilarity, 1=maximal dissimilarity).
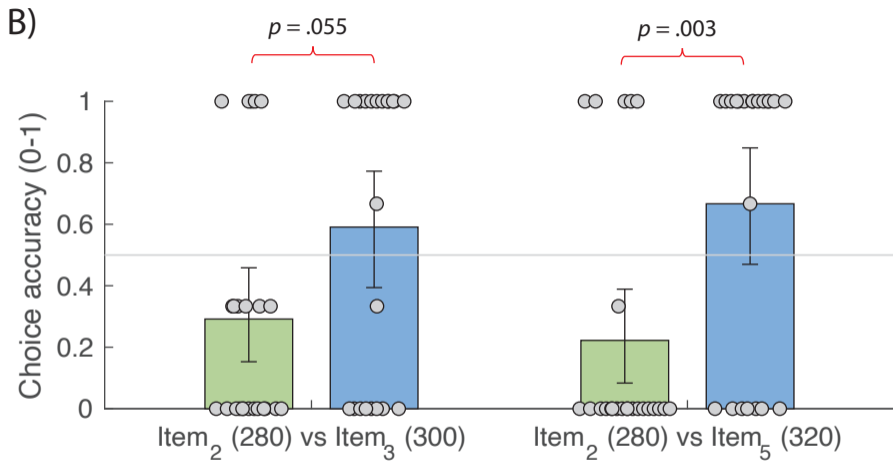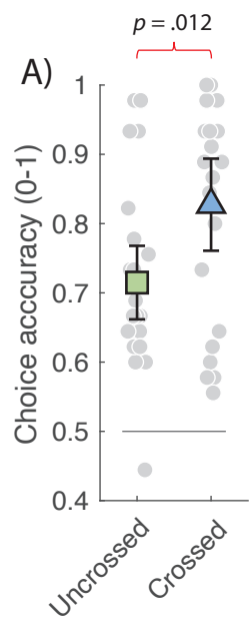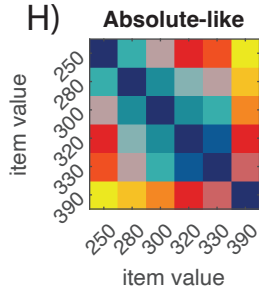
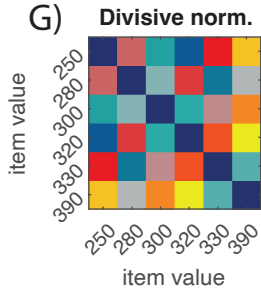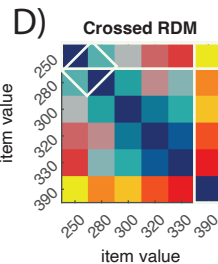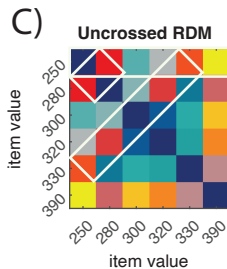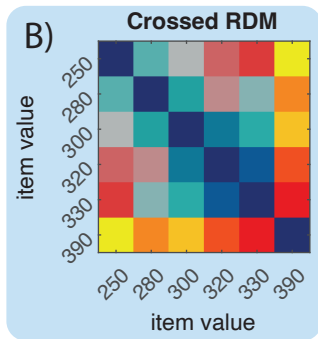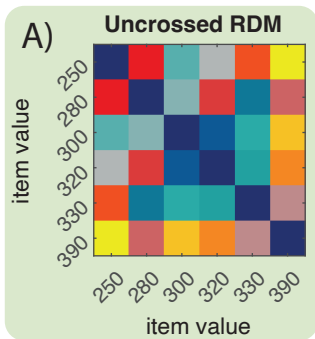822   **Figure 7.** Experiment 2 Model RDM correlations. Partial Spearman participant x model correlations for the
823   Uncrossed group (A, green squares, N=119) and Crossed group (B, blue triangles, N=105). Each plot shows two
824   analyses: one in which range-adaptation is pitted against absolute encoding, and another in which divisive
825   normalisation is pitted against absolute encoding. The larger the r the better the model accounts for participants'
826   value representation. Symbols indicate group means and error bars reflect 95% CIs. Grey lines represent
827   individual participants. Downwards sloping lines (from left to right) indicate that participants' representation of
828   value is better modelled as relative. Upward sloping lines (from left to right) indicate that the participants' value
829   code is better accounted for by an absolute code. (C) Mean participant x Model correlation differences
830   (participant x Absolute r – participant x Relative r). Positive r's indicate that the absolute model fits better and
831   negative r's that the relative model fits better. Symbols reflect means and error bars reflect 95% CIs. The
832   reported p-values reflect key Crossed-Uncrossed group-wise contrasts assessing whether the evidence in favour
833   of the absolute model over the relative model was stronger in the Crossed group: $t(222) = 3.25$, $p < .001$, lower
834   CI = .17, upper CI = inf, $d = .43$); $t(222) = 3.09$, $p = .001$, lower CI = .15, upper CI = inf, $d = .41$). (D) As (C)
835   but for independent correlations. The p-values reflect key across-group contrasts: $(t(222) = 3.01$, $p = .002$,
836   lower CI = .05, upper CI = inf, $d = .40$); $(t(222) = 3.15$, $p < .001$, lower CI = .02, upper CI = inf, $d = .42$).
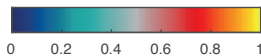
A)

| Condition | Phase 1 (antiques/cars) | | | Phase 2 (cars/antiques) | | | |
| | Learning 1 | Decision 1 | | Learning 2 | Decision 2 | | All-pairs | Judgement |

A)

Phase 1 (antiques/cars) — **Phase 2 (cars/antiques)**

Condition	Learning 1	Decision 1	Learning 2	Decision 2	All-pairs	Judgement
Uncrossed	item$_1$ ~ $N(150,15)$ vs. item$_2$ ~ $N(180,18)$ item$_3$ ~ $N(200,20)$ vs. item$_4$ ~ $N(230,23)$	item$_1$ vs. item$_2$ item$_3$ vs. item$_4$ item$_5$ vs. item$_6$	item$_1$ ~ $N(250,25)$ vs. item$_2$ ~ $N(280,28)$ item$_3$ ~ $N(300,30)$ vs. item$_4$ ~ $N(330,33)$	item$_1$ vs. item$_2$ item$_3$ vs. item$_4$ item$_5$ vs. item$_6$	item$_1$ vs. item$_3$ item$_6$ vs. item$_1$ item$_2$ vs. ...	item$_3$ item$_6$ ...
Crossed	item$_5$ ~ $N(220,22)$ vs. item$_6$ ~ $N(290,29)$	item$_1$ vs. item$_4$ item$_2$ vs. item$_6$ item$_3$ vs. ...	item$_5$ ~ $N(320,32)$ vs. item$_6$ ~ $N(390,39)$	item$_1$ vs. item$_4$ item$_2$ vs. item$_6$ item$_3$ vs. ...		

B) Learning 1

Sampling

Sell — Sell — 180
Sell — Sell — 165
Sell — Sell — 183

Preliminary decision

Sell — Sell

Sampling

Sell — Sell — 180
Sell — Sell — 210

C) Decision 1

Uncrossed pairs

Sell — Sell
Sell — Sell

item$_2$ > item$_1$
item$_3$ < item$_4$

Crossed pairs

Sell — Sell
Sell — Sell

item$_2$? item$_4$
item$_3$? item$_1$

D) Learning 2

item$_5$ < item$_6$
item$_2$ > item$_1$

item$_1$ ≈ 250
item$_3$ ≈ 300
item$_3$ ≈ 300
item$_2$ ≈ 280

Decision 2

item$_5$ < item$_6$
item$_2$ > item$_1$

250 < 300
300 > 280

E) All-pairs

item$_2$? item$_4$
item$_3$? item$_1$

280 < 330
300 > 250

F) Judgement

item$_1$ = ?

item$_1$ ≈ 250

A) Uncrossed RDM

B) Crossed RDM

C) Uncrossed RDM

D) Crossed RDM

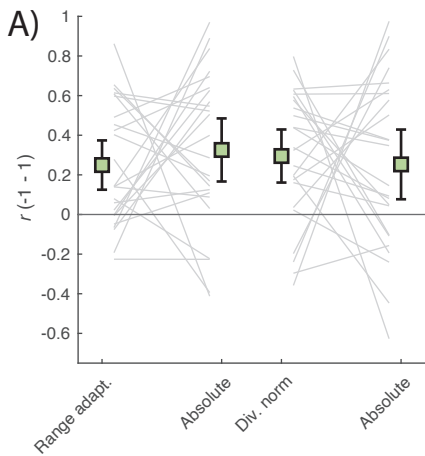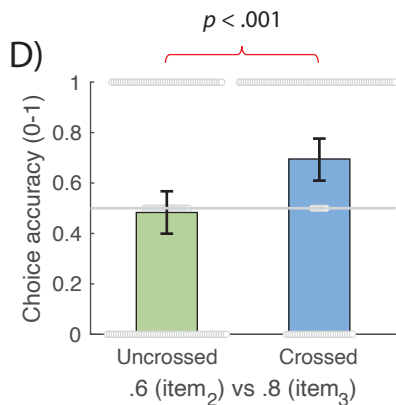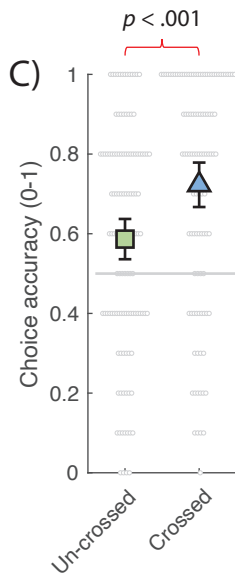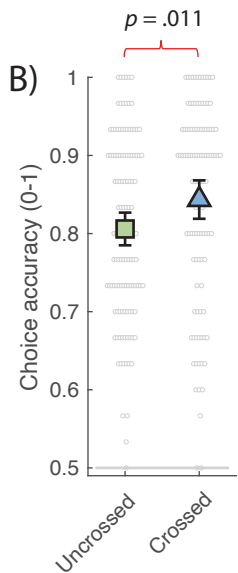**E)** Valence

F) Range adapt.

G) Divisive norm.

H) Absolute-like

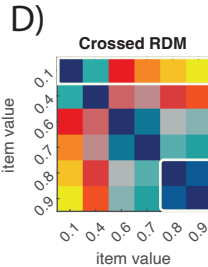Rank-transformed scaled similarity:

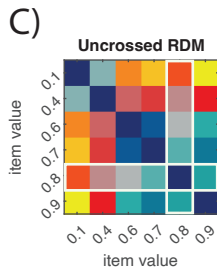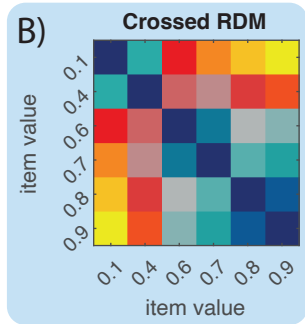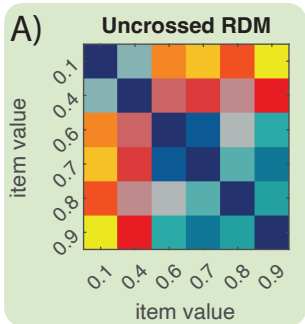0   0.2   0.4   0.6   0.8   1

A)

Condition	Phase 1 (antiques/cars)			Phase 2 (cars/antiques)			
	Learning 1	Decision 1	Learning 2	Decision 2	All-pairs	Judgement
Uncrossed	$item_1 \sim B(.1, 1)$ vs. $item_2 \sim B(.6, 1)$ $item_3 \sim B(.4, 1)$ vs. $item_4 \sim B(.7, 1)$	$item_1$ vs. $item_2$ $item_3$ vs. $item_4$ $item_5$ vs. $item_6$	$item_1 \sim B(.1, 1)$ vs. $item_2 \sim B(.6, 1)$ $item_3 \sim B(.4, 1)$ vs. $item_4 \sim B(.7, 1)$	$item_1$ vs. $item_2$ $item_3$ vs. $item_4$ $item_5$ vs. $item_6$	$item_1$ vs. $item_3$ $item_6$ vs. $item_1$ $item_2$ vs. ...	$item_3$ $item_6$ ...
Crossed	$item_5 \sim B(.8, 1)$ vs. $item_6 \sim B(.9, 1)$	$item_1$ vs. $item_4$ $item_2$ vs. $item_6$ $item_3$ vs. ...	$item_5 \sim B(.8, 1)$ vs. $item_6 \sim B(.9, 1)$	$item_1$ vs. $item_4$ $item_2$ vs. $item_6$ $item_3$ vs. ...		

Rank-transformed scaled similarity:
0   0.2   0.4   0.6   0.8   1