Title:

Human value learning and representation reflects rational adaptation to task demands

Author list:

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

Affiliations:

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

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

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

*Andreas Jarvstad, andreas.jarvstad@city.ac.uk, College Building, City, University of London, Northampton Square, London EC1V 0HB, United Kingdom.

Abstract:

Humans and other animals routinely make choices between goods of different value. Choices are often made within identifiable contexts, such that an efficient learner may represent values relative to their local context. However, if goods occur across multiple contexts, a relative value code can lead to irrational choice. In this case, an absolute context-independent value is preferable to a relative code. Here, we test the hypothesis that value representation is not fixed, but rationally adapted to context expectations. In two experiments, we manipulated participants' expectations about whether item values learned within local contexts would need to be subsequently compared across contexts. Despite identical learning experiences, the group whose expectations included choices across local contexts, went on to learn more absolute-like representation than the group whose expectations only covered fixed local contexts. Thus, human value representation is neither relative nor absolute, but efficiently and rationally tuned to task demands.

Main text:

Humans and other animals often behave "as if" they calculated the value of goods, arranged goods according to their preferences in a rational manner, and chose the good with highest value. One way to achieve rational decision-making is to represent all items on an absolute scale, where an item's value is expressed as the amount of fixed units of measurement it provides. Units of measurement might be food items in a foraging patch, money, or the subjective utility of consumer products. Such an absolute value code is assumed in normative theories of decision-making[1], optimal foraging theory[2], many computational models of learning[3] , and in key descriptive theories of choice[4].

Whilst an absolute code would equip the agent to make decisions across all contexts in which this unit of measurement is relevant, there are many reasons why biologically constrained systems may utilise different coding regimes. Absolute codes that maintain a constant unit may, for example, reserve precious coding range for values that occur with low frequency. Moreover, absolute codes may be more prone to deleterious noise if values cluster within a small range in each context (leading to easily confusable items).

From the olfactory system in the fruitfly[5], to visual systems[6], through to value coding in humans[7], neural systems can overcome such problems by encoding input relative to the local context (and/or state[8,9]). The value of one foraging patch can, for example, be encoded relative to other nearby patches. Such context-dependent encoding has been formalised in computational models, for instance by ensuring that coding covers the entire range of values ('range adaptation'[10]) or by ensuring that values are normalised by concurrent inputs ('divisive normalisation'[11]).

The key advantage of relative codes is that they enable even small populations of neurons to efficiently represent items within a local context[11]. For the perceptual system, for example, adapting to local brightness levels (dark adaptation[12]) is likely close to optimal given the temporal and spatial autocorrelation in brightness in natural scenes (day-night light cycle). For value-based decisions, agents can boost discriminability using relative codes, which may be of particular importance if the agent aims to choose "correctly" (i.e., choose the highest valued item). This means that a foraging animal employing a relative value code may discriminate between patches of values A=5 and B=6 with equal precision to when choosing between patches of values C=20 and D=21.

There is now ample evidence from psychology, behavioural ecology, primate neurophysiology and cognitive neuroscience that humans and other animals learn, and/or make choices consistent with such context-dependent value codes ([9,13–19] but see[20]). A relative context-dependent code also describes the firing pattern of neurons in value-related areas of the prefrontal cortex[21] and explains human errors of judgment across many domains[17]. Relative codes have also been shown to be efficient in the sense that they maximize mutual information between stimulus and neural code under certain conditions[22]. In this latter sense, context-dependent codes can be locally optimal and resource efficient – allowing animals to choose the best option with the use of minimal resources[22,23].

However, as can easily be seen, relative value encoding can lead to inferior decision-making if the local contexts in which values were encoded are intermixed. In the above example, for instance, foraging patch B=6 is the locally superior option to A=5, which means that a pure relative encoder may prefer it to the globally superior option from a different context - provided it is inferior in its local context (e.g., prefer B=6 to C=20, where C is from [C=20, D=21]). Such 'irrational' decision-making has been observed across species in many laboratory tasks[10,15,17,24].

Thus, one is faced with an additional problem: How to arbitrate the costs and benefits of absolute and relative encoding to optimize decision-making. This problem can be recast as one of expectation about context: If contexts are stable and distinct, relative encoding will be sufficient and maximizes discriminability, but if contexts are volatile and/or overlapping in time, a coding regime approximating absolute encoding will be better. Here, we take a first step towards this question by implicitly manipulating human participants' expectations about contexts in two experiments. In spirit, our work is similar to efforts in reinforcement learning to delineate under what circumstances, and under what cost, humans switch from a habitual (model-free) representation to a more costly representation that allows planning (model-based)[25,26].

In particular, we propose that humans do not use a single fixed representation of value, but flexibly tune value codes based on their expectations what the codes are for[27]. Further, we propose that the selection of which code to learn, is rational and efficient[28]. Thus, we do not ask whether human value learning is absolute or relative overall [13,15], but rather whether humans flexibly adapt[29,30] their value representation in a manner that can be explained by expectation.

We tested the hypothesis that value representation rationally adapts to task demands in two value-learning experiments, in which human participants learned values of pair-wise presented items. We implicitly manipulated task expectations, such that one group expected to make decisions within fixed local contexts (‘Uncrossed’), and another group expected to make decisions across local contexts (‘Crossed’). If value learning is fixed, the learnt value representations should be identical across groups. If value learning is rationally and flexibly adapted to task demands, people in the ‘Crossed’ group should go on to learn more absolute-like representations (because they expect these to be task-relevant).

Despite identical learning experiences, learnt value codes differed: participants learned more complex (absolute) representations only when they expected it to be necessary, thus highlighting the rational and dynamic nature of value representation.

Results

Design

We conducted two value learning experiments. The first experiment used real-valued items, akin to studies in economic decision-making[31], whereas the second used binomial outcomes akin to many reinforcement learning paradigms in this domain[15]. In both experiments, participants went through two independent phases of learning (with feedback) and decision-making (without feedback), with the stimuli optimised to allow reliable distinction between absolute and relative value encoding (Supplementary Methods: I; II).

In the learning phases, participants learned the value of items through trial-by-trial feedback. As our key experimental manipulation, we implicitly altered participants’ expectations about the local contexts in which items had to be compared. After the initial learning phase, one group (Uncrossed) was presented with choices between fixed pairs of items (within contexts), whereas the other group (Crossed) encountered items also in intermixed pairs (across contexts).

We expected the Crossed group to use the experience of intermixed contexts to alter their value encoding for the subsequent independent learning phase. Value representations in both groups were measured with two surprise tasks at the end of each experiment. We first report on Experiment 1.

Experiment 1 – real-valued items

Participants took on the role of consultants to manufacturers of reproduction items (replicas of historical items). There were two separate manufacturers (of cars & antiques) in two separate Phases (Figure 1A). Participants' goal was to learn market prices in order to 'consult' on which items to manufacture.

In the Learning Phases, participants learned item values through trial-by-trial feedback, after which they advised the manufacturer in separate Decision phases. At the end, there were two surprise tasks (All-pairs, Value judgment) designed to measure value encoding in the last Learning phase. Participants were randomly and blindly assigned to either the Uncrossed or Crossed group (colour-coded green and blue respectively, Figure 1A).

Each Phase began with a Learning stage, in which participants sampled market values (Figure 1B). A single mouse-click on an item returned a single sale price (superimposed on the clicked item). Participants were free to sample in any order and as much as they wished. Sampling for each pair was terminated by a selling decision, after which the next pair was shown. In each Phase, participants learned the values of 6 items arranged into 3 pairs with normally distributed market prices (Figure 1A).

In the Decision stage (Figure 1C), the Uncrossed group made decisions about the pairs they had previously experienced. The Crossed group additionally made decisions within novel pairings, thus breaking their learning contexts. Participants might, for example, decide between Item_2 and Item_4 which had previously formed part of the first and second pair respectively. Participants' choices in the Decision stages and surprise tasks were incentive compatible.

We hypothesized that people do not use a fixed value-learning mechanism, but flexibly adapt their value-learning mechanisms to learn useful value representations. Given double-blind assignment to groups, both groups start ‘Learning 1’ with the same expectations. However, the first Decision phase (Figure 1C) provides very different implicit signals for the two groups.

The Uncrossed group should have no problem performing in this task given successful learning (Figure 1C). This would even be the case if participants used extreme context-dependent encoding: a binary Valence code. Using this mechanism, one learns, for each pair, that one item is ‘good’ and that one item is ‘bad’. That is, one learns the following (separate) sets of orderings: [Item_1 < Item_2], [Item_3 < Item_4], and [Item_5 < Item_6].

In the Crossed group (Figure 1C), however, even participants who used less extreme relative encoding strategies may struggle to compare items across contexts: comparing, for example, Item_5 ~ 320 (low-value in its context) and Item_2~280 (high-value in its context). These unexpected and potentially more difficult experiences led participants to respond more slowly (Supplementary Results III).

If people adapt to expected task demands as hypothesized, and the implicit manipulation is sufficient to induce different expectations, the two groups should go on to learn different representations for the subsequent set of items – Learning 2 and Decision 2. Immediately after these tasks, we tested participants’ learned representations using two independent surprise tasks.

Experiment 1 – real-valued items - Decision-making performance

First, we tested participants performance in an All-pairs task, in which all possible pairs of items were presented to both groups (without feedback). We found that the Crossed group’s choice accuracy was statistically significantly better than the Uncrossed group’s despite identical learning Phases ($t(44) = 2.61$, $p = .012$, $CI = .026–.199$, $d = .77$, two-tailed independent t-test) and observed above-chance performance in both groups (Figure 2A, CIs do not overlap .5, see also Supplementary Results II). The difference in performance is consistent with the Crossed group having encoded a more absolute-like value representation than the Uncrossed group (Supplementary Methods I).

The worse performance for Uncrossed was accompanied by slower decision times, suggesting greater processing demands, and shows that worse performance was not simply due to spending less time on decisions (i.e., a speed-accuracy trade-off, Supplementary Results VI).

Next, we turned to a feature of our experimental design which allowed us to dissociate absolute-like encoding from any relative encoding using ‘diagnostic’ item pairs. The intuition is that any relative encoding will result in a fraction of choices that are globally inferior, but locally superior within the learning context, whereas an absolute code would not result in the same mistakes. The items in our task were chosen to optimize for this (Supplementary Methods I).

Specifically, in Phase 2, $\text{Item}_2 \sim N(280,28)$ was paired with $\text{Item}_1 \sim N(250, 25)$. On the one hand, a relative learner would learn that Item_2 is ‘good’ within its local context. On the other hand, they would learn that both $\text{Item}_3 \sim N(300,30)$ and $\text{Item}_5 \sim N(320,32)$ are ‘bad’, because they were paired with higher-value items in their respective context (Figure 1A). Thus, a relative-value learner should prefer the locally ‘good’ (but globally inferior) Item_2, to the locally ‘bad’ (but globally superior) $\text{Item}_{3,5}$: thus exhibiting irrational choice [see also e.g., 15].

In line with these predictions, we found that the Uncrossed group preferred the globally inferior option, choosing it instead of the globally superior options (preferring Item_2 to $\text{Item}_{3,5}$), whereas the Crossed group expressed a weak preference for the globally superior items. The difference between groups was statistically marginally significant for the first pair ($U(46) = 183$, $p = .055$, $r = .31$), and statistically significant for the second pair ($U(46) = 145$, $p = .003$, $r = .45$), by two-tailed Mann-Whitney U’s.

In summary, participants choice behaviour shows that the groups learned different value representations despite identical learning Phases, and that, compared to Uncrossed, choices in Crossed were more consistent with an absolute code.

Decoding value representation

While the above analyses provide tentative evidence that the groups learned different value representations, we next set out to address this more directly. For this purpose, participants were asked to directly indicate their learned value for each item in a Value Judgment task (Figure 1A). Items were presented sequentially (in random order), and participants indicated perceived value using a slider.

We applied representational similarity analysis (RSA)[32,33] to these data. Although RSA was developed mainly as a multivariate analysis technique for neural data, it is increasingly deployed to characterize brain representations given behavioural data (e.g.,[34–36]) and can be used whenever the measure of interest is pair-wise distances on a univariate or multivariate space.

We computed representational dissimilarity matrices (RDMs) separately for each participant and averaged them to form group-wise RDMs (for raw judgement data see Supplementary Results IV). Shown in Figure 3A-D, these RDMs depict each group's value representation in the form of a dissimilarity structure that is rank-transformed and scaled (Methods). On this scale, a dissimilarity of 0 implies that item values are represented identically (item pairs along the diagonal), and a dissimilarity of 1 implies that item values are highly dissimilar.

Empirical RDMs are most readily interpreted when compared to model RDMs. We compared participants RDMs to four model RDMs, three relative value RDMs: 'Valence', 'Range-adaptation', 'Divisive normalisation' and an 'Absolute' value RDM. We included different classes of relative models to ensure that results do not only hold for a single type of relative encoding.

Note, however, that whilst we can readily contrast different relative models with the absolute model - and ask which best explains people's representation of value - we cannot reliably determine if a relative code was generated by range-adaptation or by divisive normalisation (Supplementary Results VIII, IX). Indeed, our study was designed specifically to discriminate absolute from relative encoding, *regardless* of the precise implementation of the relative value encoding (Supplementary Methods I; II).

The first implemented relative model ('Valence', Figure 3E) formalizes the extreme 'good vs bad' encoding discussed above. The better option in each local context is encoded as 'good' and the worse option as 'bad'. Thus, this model does not retain any magnitude information. The second relative model ('Range adaptation', Figure 3F) formalizes range-adaptation encoding, a highly successful class of context-dependent encoding schemes[10,16]. Accordingly, the value of the left item equals $\frac{\text{item}_{left}}{\max\left(\text{item}_{left}, \text{item}_{right}\right)}$ (and vice versa for the right item). Note that this model scales values within local contexts to the interval $[\frac{\min(V)}{\max(V)}, 1]$, rather than the interval [0,1]. This is necessary here as with only two items, the full range adaptation model (e.g.[16]) would otherwise reduce to the valence model. The third relative model ('Divisive normalisation', Figure 3G) formalizes the divisive normalisation encoding highlighted in the Introduction. Here the value of the left item equals $\frac{\mathrm{item}_{left}}{1+\mathrm{item}_{left}+\mathrm{item}_{right}}$ (and vice versa for the right item). Finally, we formalize absolute context-independent encoding, as the expected value for items. For example, Item_2 is encoded as 180 because $\mathrm{Item}_2 \sim N(180,18)$.

Experiment 1 – real-valued items – Value representation

First, we highlight qualitative differences and similarities between participants value representation and those predicted by the different models. As can be seen in Figure 3, the three relative RDMs (E-G) have clusters of items that are objectively similar in value but are nonetheless encoded as highly dissimilar. For example, all the relative models capture the 'irrational' value encoding, by which $\mathrm{Item}_2 \sim N(280,28)$ is encoded as more like $\mathrm{Item}_4 \sim N(330,33)$, than $\mathrm{Item}_3 \sim N(300,30)$. This 'irrational' dissimilarity structure follows from the context-dependent encoding of value formalized in the relative value models (see also Supplementary Methods I; II).

In Figure 3 Panel C-D, qualitative similarities are highlighted by white outlines. Items with values 250 and 280, for example, are encoded as dissimilar in the Uncrossed RDM - as they are in the relative models (E-G). The Crossed RDM, on the other hand, encodes this item pair as similar: in keeping with the Absolute RDM (H). The Crossed RDM also reflects a gradient of increasing dissimilarity between 390 and the other lower-value items. The Uncrossed RDM does not seem to exhibit this gradient. Finally, both Uncrossed and Crossed groups encode value in a format that goes beyond mere valence encoding (Figure 3A-B, E). Thus, participants in both groups encode value magnitude information (Supplementary Results V for formal tests).

Next, we turned to the key quantitative comparisons. We contrasted the correlations between each model and the two groups. As per standard practice[33], model RDMs were compared to the RDMs derived from participants' behaviour using rank-correlations (Methods). A large positive correlation between a participant's RDM and a given model RDM, shows that their representation of value is well accounted for by the model. For presentation purposes, we focus on the two relative models that capture key aspects of participants value representation: range-adaptation and divisive normalisation.

The Crossed group learned a more absolute value representation than the Uncrossed group: both compared to the Range-adaptation model ($t(44) = 2.97$, $p = .005$, $CI = .15$-$.77$, $d = .88$) and the Divisive normalisation model ($t(44) = 2.57$, $p = .014$, $CI = .10$-$.85$, $d = .76$), both two-tailed independent t-tests.

Figure 4A-B plots model-participant RDM similarities expressed as partial Spearman Correlation Coefficients (thus discounting shared variance between models). Because the Range-adaptation and the Divisive normalisation RDMs were highly correlated, we ran separate analyses contrasting each with the Absolute RDM. Symbols in Figure 4 reflect group averages, and grey lines reflect individual participants.

For the Uncrossed group (A), no model consistently outperforms another, indicated by the mix of slopes. In the Crossed group, however, most participants are substantially better accounted for by absolute encoding (upward sloping lines), indicating that most participants shifted their encoding strategy towards an absolute code.

Figure 4C shows the within-group contrast between the Absolute model and the two relative models from Figure 4A-B. Positive Δr indicate evidence in favour of the absolute model, and negative indicate evidence in favour of the relative model. As can be seen, no model is consistently favoured in the Uncrossed group (CI's overlap 0). However, in the Crossed group, the absolute model is favoured (CIs do not overlap 0).

The previous analyses used a partial correlation approach to rule out the contribution of any shared variance. To ensure that these results do not depend on removing shared variance, we also ran the same analysis using independently run correlations (Methods). As can be seen in Figure 4D, the results replicate with independent correlations; the Crossed group learnt a more absolute value representation than the Uncrossed group: whether one considers the Range-adaptation RDM ($t(44) = 3.23$, $p = .002$, $CI = .21 - .92$, $d = .95$), or the Divisive normalisation RDM ($t(44) = 2.88$, $p = .006$, $CI = .13 - .74$, $d = .85$).

Jointly, the results show that 1) people adapt their learning to expected task demands (difference between groups despite identical learning Phases), and 2) people only learn absolute-like value representations when a relative representation is expected to be insufficient for the task at hand (in the Crossed group).

Experiment 2 – Binomial outcomes

Next, we turned to a binomial decision task similar to many decision-making tasks in the field of reinforcement learning. Although economic values often come from continuous distributions as in Experiment 1 (e.g., market prices, food quantities), laboratory tasks often involve binomial distributions[15,37–39]. We therefore sought to establish whether people can also flexibly tune their value-learning mechanism(s) for binomial outcome distributions.

Key design features were kept identical to Experiment 1 (Figure 5A): learning experiences were identical across conditions, Phase 1 was designed to set participants' expectations for Phase 2 in a condition-dependent manner (Crossed vs. Uncrossed), and learnt values were assessed in separate surprise tasks (All-pairs, Value judgement). Notable exceptions include using binomial value distributions, the number of 'samples' being fixed, and the experiment being run online.

Based on Experiment 1, we predicted that, compared to the Uncrossed group, the Crossed group would show 1) better All-pairs performance, 2) improved choice for the single diagnostic item pair and 3) more absolute-like value representations. An initial experiment (Supplementary Results I) broadly confirmed these predictions but was underpowered to find a between-group effect of moderate size. We therefore ran a better powered pre-registered replication on which we report next.

Experiment 2 – Binomial outcomes - Decision-making performance

As can be seen in Figure 5B, in both groups choice performance was statistically significantly above chance (CIs do not overlap .5, see also Supplementary Results II). As in Experiment 1, the Crossed group made statistically significantly better decisions in All-pairs following learning (Figure 5B, $t(222) = 2.30$, $p = .011$, $CI = .011$-inf, $d = .31$, one-tailed independent t-test). As in Experiment 1, the worse performance in the Uncrossed group was accompanied by slower decision times (Supplementary Results VI).

Next, we further constrained our comparison to those item pairs for which a divisive normalisation model would make opposing predictions to an absolute value code (Supplementary Methods II). Even for this restricted analysis (Figure 4C), for which choosing is more difficult (differences between values are smaller, Supplementary Figure 3A) choice performance was statistically significantly above chance in both groups (non-overlapping CIs, Figure 5C). However, for this sub-selection, the Crossed group again made statistically significantly better decisions than the Uncrossed group ($t(222) = 3.56$, $p < .001$, $CI = .073$–inf, $d = .48$, one-tailed independent t-test).

Restricting the analysis further to the single diagnostic stimulus pair (Supplementary Methods II) replicates Experiment 1 (Figure 5D): Crossed group participants chose the higher-value option more frequently than Uncrossed (Figure 5C; $U(224) = 4722$, $p < .001$, $r = .25$, one-tailed Mann-U Whitney test). Thus, as in Experiment 1, decision-making was consistent with the Crossed group learning a more absolute-like value representation than the Uncrossed group.

Experiment 2 – Binomial outcomes - Value representation

Next, we turned our attention again to the RSA analyses. Figure 6A-B show the group-wise average RDMs for Experiment 2. As in Experiment 1, Figure 6C-D highlight similarities between the empirical average RDMs and the model RDMs. As in Experiment 1, participants' value representation was not consistent with a Valence code (Figure 6E, for formal tests see Supplementary Results V, and for raw judgements see Supplementary Results IV).

As in Experiment 1, the Crossed group learned a more absolute value representation than the Uncrossed group: whether one considers the Range-adaptation RDM ($t(222) = 3.25$, $p < .001$, $CI =$ .17-inf, $d = .43$), or the Divisive normalisation RDM ($t(222) = 3.09$, $p = .001$, $CI = .15$-inf, $d = .41$, both one-tailed independent t-tests).

Comparing the empirical RDMs to the finer-grained model RDMs, the 'cross-type' pattern in the two-remaining relative RDMs (Figure 6F-H) is apparent in Uncrossed group (Figure 6C), but largely absent in the Crossed group (Figure 6D). The latter instead seems to reflect a gradient of dissimilarity approximating the underlying outcome probabilities as in the Absolute Model (.1 vs the remaining item values, Figure 6C).

Next, we turned to the partial correlation analyses plotted in Figure 7A-B. For the Uncrossed group (A), there was a trend towards the relative models performing better than the absolute model. However, as in Experiment 1, no model consistently outperformed another (mix of sloped lines). In the Crossed group, however, participants were substantially better accounted for by absolute encoding (upward sloping lines); regardless of whether the comparison is a Range adaptation RDM or a Divisive normalisation RDM.

Figure 7C shows the within-group contrast between the Absolute model and the two relative models from the data in Figure 7A-B. Positive Δr indicate evidence in favour of the absolute model, and negative indicate evidence in favour of the relative model. As can be seen, there is no statistically significant evidence of any model being consistently favoured in the Uncrossed group (all CIs overlap 0). However, in the Crossed group, the absolute model is clearly favoured (CIs do not overlap 0).

Figure 7D shows an analysis identical to that in Figure 7C except that it has been carried out on independently run correlations. As can be seen, the results persist with independent correlations. The Crossed group learned a more absolute value representation than the Uncrossed group: whether one considers the Range-adaptation RDM ($t(222) = 3.01$, $p = .002$, $CI = .05$-inf, $d = .40$), or the Divisive normalisation RDM ($t(222) = 3.15$, $p < .001$, $CI = .02$-inf, $d = .42$, both one-tailed independent t-tests).

Thus, Experiment 2 replicated and generalised the results of Experiment 1, using an online study with binomial outcome distributions. Choice task data showed that the Crossed group learned a different value representation than the Uncrossed group – despite identical learning experiences. The choices in the Crossed group were better than those in the Uncrossed group and were better specifically for item pairs for which an absolute-like representation will result in improved choice. The RSA analyses further show that the Crossed group learned an absolute-like representation, and that they learned a more absolute-like representation than the Uncrossed group.

Discussion

We sought to reconcile the theoretical and empirical tension between two diametrically opposing accounts of value learning and encoding: a context-independent but potentially computationally costly absolute value representation[1,2,4], and an efficient local, but potentially irrational, relative value representation [7,11,13,15]. We proposed that humans do not use a single fixed mechanism – learning either absolute or relative value codes – but adapt their learning to expected task demands in an efficient and rational manner: learning sufficient and necessary value representations.

We tested this hypothesis in two value-learning experiments: one involving real-valued items and the other involving binomial outcomes. In each study, the first phase was equivalent to the full experience of participants in many experimental paradigms[(e.g., 15,37)]. The second phase gave participants the chance to use their prior experience with the task to tune their learning mechanism to optimise task performance. Phase 2 thus mimicked the opportunity to adapt learning mechanisms that arise in many real-life tasks (and which are performed more than once).

Despite identical learning experiences, the two groups learned different value codes. Specifically, across the two experiments, the Crossed group made decisions consistent with a higher-fidelity representation (Figs. 3,6), made fewer irrational choices (Figs. 2,5), and learned representations that were more absolute-like than the Uncrossed group (Figs. 4,7). Importantly, participants learned more absolute representations only when it was expected to be useful. Thus, people learn neither absolute nor relative value codes, but adapt their learning to what they expect to code to be used for.

Nevertheless, the reliable group differences were not always reflected at the individual level. In the Uncrossed condition, many participants appeared to have learnt absolute-like codes. This may be driven by the fact that both absolute and relative codes yield good results for the Uncrossed group. Thus, whichever code participants favour as their “default” would be expected to persist. Relatedly, in the Crossed condition some participants appeared to have learnt relative codes. This may be due to factors beyond the scope of our current study, such as cognitive capacity limitations[40], intrinsic computational noise[41,42], or mechanisms relating to working memory or attention[43,44]. Future work might manipulate task demands and difficulty[38,45, (c.f. 46] to address these factors.

A second outstanding question relates to precise encoding of the flexibly adapted representation. Whilst our design allowed us to determine that people switch to an absolute value representation when they expect it to be necessary, it does not allow finer-grade discrimination between different kinds of relative encodings. Apart from the Valence model, which we were able to largely reject (Supplementary Results V), different experimental designs are required to address this question. It is also possible that any relative value code is itself adaptable and/or determined by task constraints.

A related question is what mechanisms give rise to the flexible and adaptive value representations we observe. Our studies were designed for well-controlled measurement of value representation following learning, with an emphasis on being able to dissociate absolute from relative value encoding. The trade-off is that the design is not effective in characterizing learning mechanisms themselves - as opposed to the codes they give rise to. Nevertheless, our design allowed us to successfully recover and discriminate between relative and absolute models in simulations, thus supporting our key RDM contrasts (Supplementary Results VIII), and allowing us to ask: Is value encoding adaptive, and if so – is it rationally adaptive?

In principle, a single mechanism could underlie the observed flexibility. Such a mechanism could, for example, be implemented with a free parameter governing the extent to which learning is relative, such that $\text{item}^{\text{left}} = \frac{\text{item}_{left}}{1+\text{w}*(\text{item}_{left}+\text{item}_{right})}$, where W is a free parameter between 0 (for wholly absolute encoding) and 1 (for wholly relative encoding, see Supplementary Results IX). Alternatively, the space over which item values are normalised could be expanded to render the code absolute-like. However, it is also possible that mechanisms rely at least in part on different substrates as in, for example, model-based and model-free learning[47–49].

Finally, extrapolating beyond the behavioural data at hand, one might reasonably expect that relative values behave like “cached” values in Reinforcement Learning, in that they incorporate context into their code (without later being able to retrieve context values), whereas absolute-like encoding may rely on memory systems that separate item and context representations, allowing the system to flexibly combine them at decision time. Thus one might expect the absolute-like representation to preferentially recruit hippocampal-medial prefrontal circuits, whereas relative encoding may rely more heavily on striatal-prefrontal circuits, as approximately in the model-free / model-based distinction in RL[49]. However, further research is needed to identify the neural mechanisms arbitrating between the two encodings.

In summary, our results highlight the highly dynamic and rational nature of value representation: humans do not simply have a single, fixed form of representation, but rather adjust their value code in a rational[50–52] manner according to expected task demands. In relation to the ongoing debate about whether the brain encodes values at all[53,54], where relative encoding is sometimes taken as evidence that it does not, our results suggest the it may well do - if the circumstances merit. In other words, perhaps both absolute and relative codes previously found can be explained by participants inferring which code would be sufficient for the task at hand.

Methods

Experiment 1 - Participants

The study complied with all relevant ethical regulations and was approved by the local ethics committee at City, University of London. Sixty participants (37 female) were recruited via the local participation panel. Participants provided written informed consent and were debriefed. Participants had normal, or corrected-to-normal, vision, were fluent in English, healthy (no known physical or psychological conditions), and between 18-45 years old. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those in previous work[10,15].

Participants were reimbursed for their time and were paid a performance-related bonus: a base pay of £5 and an additional bonus between £0 and £6. The average bonus was for a total of £2.78 (range £0-6). The performance bonus was determined by choice performance across all Decision Phases as well as during the final two tasks. The greater the number of high-value choices, the greater the bonus, and the closer the judgement to the true item value the higher the bonus.

We excluded participants who did not fulfil minimal task requirements. Criteria apply to the Learning phases only (Figure 1A-B), and are therefore orthogonal to the target behaviour in the final tasks (Figure 1A). Exclusion criteria were based on 1) sampling behaviour and 2) below-chance performance for the preliminary decisions in the first sampling phase. Participants who only sampled once (or fewer times), per item per item-pair sampling opportunity, were excluded (Learning 1-2, Figure 1). This cut-off represents <= 18 samples per Context and is far lower than the median of 123 (*IQR*=118) and 143 (*IQR*=91) for Phase 1 and 2 respectively. There were 9 preliminary decisions in the first Phase (3 pairs presented three times each, Figure 1 A-B). Someone who responded randomly when making these decisions, would be expected to achieve a choice accuracy between .22 and .78 (with a mean choice accuracy of .5). This range reflects the lower and upper 95% confidence interval on a hypothetical agent who responds randomly (i.e., selects each option with $p = .5$). Participants who performed worse than the upper confidence interval (i.e., did not achieve at a greater choice accuracy than expected by chance) were excluded.

In summary, we excluded participants who showed no or little evidence of learning – a pre-condition for encoding value (whether in an absolute or relative form). In total, fourteen participants met one or both exclusion criteria for a final sample size of n=46: 24 of which had been assigned to the Uncrossed condition, and 22 of which had been assigned to the Crossed condition.

Experiment 1 - Materials

Participants took on the role of a consultant to a manufacturer of reproduction items in two different contexts (antiques/cars, Figure 1). The item-values and item-pairs were Phase-specific (Figure 1A). However, the mapping of item type (antiques/cars) to Phase, the mapping of specific items (e.g., typewriter) to item-values (e.g., $N(180,18)$), and the side on which items were presented during sampling, were all randomized across participants.

Item-values (Figure 1A) were selected primarily so that absolute-value and relative-value representations dissociate (Supplementary Methods I-II, Supplementary Figure 1-3), and secondarily to achieve a balance between task-difficulties in the Learning and Decision phases (Supplementary Methods III, Supplementary Figure 4-7). A single sample from one item resulted in a draw from the corresponding normal value distribution (truncated at ±2 SD). The Learning phases (Figure 1A-B, D) were self-paced, and participants had a wide range of different strategies as evidenced by the wide range of the number of samples drawn (range Phase 1: min=48, max=478: range Phase 2; min=32, max=509).

The Decision phases (Figure 1A, C-D) involved 18 decisions per Phase. The Uncrossed group decided between the pairs they had experienced during sampling (repeated 6 times = 18 decisions). The Crossed group made decisions between novel pairs (6 novel pairs x 2 = 12 decisions, see Figure 1 for examples), in addition to learnt pairs (3 pairs x 2 = 6 decisions).

The two final tasks (Figure 1A, E) were identical across groups. The All-pairs task involved 15 pairs, representing a full factorial combination of all possible pairs from Phase 2 (excluding identical pairs), repeated three times for a total of 45 pairs. The Value judgment task involved the 6 items in Phase 2, presented one at a time along with a slider-interface (min=100, max=450). For all tasks, the presentation order and presentation side (where applicable) were randomized across participants.

Experiment 1 Procedure

Participants read the information sheet, provided written informed consent, and completed the tasks. After completing the behavioural tasks, participants completed three questionnaires. These formed part of one author's MSc dissertation project and are not reported on here.

Experiment 1 Apparatus

Stimuli were displayed on a touchscreen (Ilyama T2245MSC) and code was written in MATLAB (Mathworks) using PsychToolbox[55] on Linux (Xubuntu 18.04) with a soft real-time kernel.

Experiment 2 Participants

The study was approved by the local ethics committee at City, University of London, and complied with all relevant ethical regulations. Participants were recruited via Prolific Academic, fully informed, provided written informed consent and were debriefed. Participants were between the age of 18 and 40, were UK residents, were healthy (no ongoing mental health conditions, dementia/mild cognitive impairment, no daily impact of mental illness), had not participated in similar studies of ours, had a minimum approval rate on Prolific of 99 and minimum of 10 submissions. We sought to include a minimum of 280 participants, conditional on having at least 100 participants in each condition passing post-completion exclusion criteria. The sample size was determined based on power calculations, which in turn were based on the pilot study (Supplementary Results I). Power calculation, exclusion criteria, and sampling strategy were pre-registered (https://osf.io/xjsmh).

Online panels provide little experimental control and the potential for poor participant engagement (see also discussion in Supplementary Results VII). To minimise this issue, we employed an initial check that participants had read and understood task instructions. To be eligible, potential participants had to answer 8 multiple-choice questions correctly. In addition, participants were allowed to make only one error in the first Decision block for the stimuli they had just learnt about. Specifically, if after experiencing 10 learning trials per item-pair, participants were unable to choose the higher value items 2 out of the first 3 presentations the study ended prematurely, and participants' pay was pro-rated. We chose to allow 1 error as even engaged participants might be expected to make mistakes especially for the more difficult stimulus pair (.8 vs .9). In total 888 participants expressed interest and 352 completed the full study. Most non-completers (92%) failed the initial knowledge test.

Participants were reimbursed for their time and were paid a performance-related bonus. Participants were paid a base pay of £2.92 for participation (the experiment took ~35 mins) and an additional bonus between £0 and £2.92. The average bonus was £1.46 (range £0.50–£2). The performance bonus was determined by choice performance across all Learning, Decision Phases and the final two tasks. Correct choices in the Decision phases and the All-pairs task were weighted x10 compared to Learning. This was done to encourage participant engagement for the tasks which did not involve feedback. In general, the reward structure was as in Experiment 1 in that the greater the number of high-value choices, the higher the bonus, and the closer the judged value to the true item value, the higher the bonus.

In addition to the pre-registered *a priori* exclusions, we also employed pre-registered exclusion criteria based on participants' not fulfilling minimal task performance criteria after completing the full study. Because each participant experienced the same number of trials, sampling behaviour cannot be used for excluding disengaged participants (unlike in Experiment 1). Instead, we excluded participants who did not learn to choose among the pairs experienced during Learning. All participants were trained on the following binomial probability pairs: [.1 vs .6], [4 vs .7] and [.8 vs .9.] - irrespective of condition. We excluded participants who made more than two errors in two repeats of these three pairs (i.e., more than 2/6 errors) at the end of the experiment (in the All-pairs task). In other words, we include only participants who showed evidence of encoding these learning phases for later recall. Note that these exclusion criteria are orthogonal to the question of absolute and relative value codes. Both absolute and relative models of learning will allow participants to learn to choose between the items in the Learning phase. In other words, choices between items of pairs that participants directly learned about - unlike novel combinations of the component items - are not diagnostic with regards to value representation.

Applying these exclusion criteria, which are orthogonal to which model participants may use to encode value, leaves N=224 participants of which n=119 participants were from the Uncrossed group and n=105 were from the Crossed group. That is, it resulted in the exclusion of ~36% of participants. We report analyses also including these excluded participants in Supplementary Results VII (these analyses replicate those reported here).

Experiment 2 Materials and Procedure

Experiment 2 was a pre-registered version of a previous study (https://osf.io/xjsmh). As in Experiment 1, participants took on the role of a consultant to a manufacturer of reproduction items in two different contexts (antiques/cars, Figure 5). Key design features were identical to Experiment 1. However, outcomes were binomial (successful sale/unsuccessful sale), the task was not self-paced, and the learning experience was not 'blocked' by item-pairs (item-pairs were randomly intermixed during learning) and involved a relatively rapid stimulus display sequence.

In the Learning Phases, participants saw each item pair presented side-by-side (~1 sec), followed by a response phase in which participants had ~1.5 second to make a choice, followed by sequential feedback, in which the chosen item was presented first followed by the unchosen item. Outcome feedback was in the form of a green double-rectangle image outline (successful sale) or a single-rectangle red image outline (unsuccessful sale).

Experienced outcomes matched the expected outcome of the binomial distributions (Figure 5A). This was achieved by pre-allocating and shuffling an outcome vector (of 1's and 0's) for each item. This design minimizes the impact of sampling error[56] on differences between participants and/or conditions. There were two learning blocks per Learning Phase. In each block each pair was presented 10 times, for a total of 30 trials per block and 60 trials per Phase. The presentation order was randomized.

Each of the two blocks of Decision trials (two for each Learning Phase), involved 12 decisions without feedback. The Uncrossed group made decisions between pairs experienced during learning (3 pairs x 4). In addition to experienced pairs (3 pairs presented once), the Crossed group made decisions also between novel pairs composed of items from different learning pairs (9 novel pairs, randomized across participants). Thus, each group experienced 24 Decision trials per Learning Phase.

The two final tasks (Figure 5A) were identical across groups. The All-pairs task involved 15 pairs, representing a full factorial combination of all possible pairs from Phase 2 (excluding identical pairs), repeated twice (controlling for presentation side) for a total of 30 pairs. The Value judgment task involved the 6 items in Phase 2, presented one at a time along with a slider-interface (min=0%, max=100%) representing the probability of an item selling. The presentation order and presentation side (where applicable) were randomized across participants for all tasks.

Design and Statistical Analyses – Experiment 1 & 2

Both experiments used a between-subject design with participants assigned randomly and blindly to one of two conditions: Uncrossed and Crossed. Our analyses focus on differences between the two groups for the two final tasks and within-task contrasts against reference magnitudes.

Prior to any analyses, we first explored data for normality by inspecting Q-Q plots and boxplots, and for independent tests equality of variances (F test). The primary inferential statistic was the t-test, and these tests relatively robust to violations of assumptions and were used unless deviations were extreme.

For data with clear deviations from parametric assumptions (e.g., Figure 2B), less powerful non-parametric equivalent tests were used. To rule out potential limits to t-test robustness affecting inferences we also ran all our t-tests reported here using non-parametric tests. All contrasts for which the t-test was used replicate with non-parametric tests (i.e., a significant test for the t-test also yields a significant result for the non-parametric equivalent). The results therefore do not depend on the assumptions of the independent and paired-samples t-tests.

We also report 95% CIs (parametric or bootstrapped) for all descriptive statistics here. CIs can be used for inference by comparing them to reference magnitudes. For example, if the mean choice accuracy is above .5, and the 95% CI of that mean does not overlap .5, choice performance was statistically significantly greater than chance (though overlapping CIs do not necessarily imply a non-significant contrast).

All reported tests for Experiment 1 are two-tailed. Predictions for Experiment 2 were pre-registered (https://osf.io/xjsmh) and derived from results from Experiment 1, and the initial pilot version of Experiment 2 (Supplementary Results I), and all between-group contrasts were one-tailed. Reported effect sizes are Cohen's *d* for t-tests ($d >= .2$ small; $d >= .5$ medium; $d >= .8$ large) and rank-biserial correlation *r* for non-parametric tests ($r >= .1$ small; $r >= .3$ medium; $r >= .5$ large). Cohen's *d* was computed as $d = \frac{abs(t)}{\sqrt{N}}$ and $d = abs(t)\sqrt{\frac{1}{n1} + \frac{1}{n2}}$ for paired-samples and independent t-statistic respectively. Rank-biserial correlation was computed as $r = 1 - \frac{2U}{n1\, n2}$ for the Mann-Whitney U statistic.

Standard RSA protocols[33] were followed. Empirical value RDMs were computed as the pair-wise Euclidean distance between each participant's value judgements. Average RDMs were computed by averaging (arithmetic mean) over participants' RDMs separately for each group. Model RDMs were computed as the pair-wise Euclidean distance between item values defined by the relevant model equations (Main Text). For display purposes RDMs were rank-transformed (equal stays equal) and scaled to 0-1, where 0 implies identical item-values and 1 means maximally dissimilar item-values.

We computed the similarity between model RDMs and participant RDMs by partial correlation (Spearman). Partial correlation accounts only for unique variance. This means that a correlation between one model RDM and a participant's RDM cannot be explained by the second model RDM. Because our interest lay in dissociating absolute from relative encoding (not distinguishing between various relative models), and because relative models were highly correlated, we performed these analyses separately for each contrasting relative model (Figure 4 & 7). We also performed analyses with independent correlations (i.e., any shared variance between models is not taken into account). Like the partial correlation analyses, these used the Spearman correlation coefficient. For all correlational analyses, large positive *r*'s imply a high degree of similarity between participants value RDMs and model RDMs (and $r = 0$ implies no relationship).

For the key statistical analysis, to establish whether the evidence in favour of the absolute model over the relative model was greater in the Crossed group than the Uncrossed group, we computed the difference between the absolute and the relative models independently for each group and contrasted those differences with t-tests. A positive difference in r indicates evidence in favour of absolute encoding and a negative difference in r indicates evidence in favour of relative encoding. These differences can also be used to infer whether there was a tendency to favour relative or absolute encoding within each group by contrasting the 95% CIs of those average differences to 0.

All statistical analyses were performed in MATLAB 2020b and 2022a (MathWorks).

Data availability

Data is available online on the Open Science Framework (https://osf.io/h32u6/).

Code availability

Analysis code is available on Open Science Framework (https://osf.io/h32u6/).

Acknowledgements:

Andreas Jarvstad was supported by a British Academy Postdoctoral Fellowship in developing this research (D-MAD, PF150005). The funders had no role in study design, data collection and analysis, decision to publish or the preparation of the manuscript. We thank Peter Barr for programming Experiment 2. We thank the Palminteri lab for helpful suggestions on these data.

Author Contributions:

KJ, AJ designed the research; TA, RH conducted the research; AJ analysed the data; KJ, AJ contributed materials/analysis tools and wrote the paper.

Competing Interests:

The authors declare no competing interests.

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-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. 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 ($Item_3$, $Item_5$). Bar height reflect means and error bars are bootstrapped 95% CIs. Grey discs represent individual participants. X-axis coordinates of participants' data have been jittered for presentation purposes. P-values reflect two-tailed Mann-Whitney U tests: $U(46) = 183$, $p = .055$, $r = .31$ and $U(46) = 145$, $p = .003$, $r = .45$ respectively. For all panels Uncrossed n = 24 and Crossed n = 22.

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). Panel A-D reflect averages over Uncrossed n=24 and Crossed n=22 respectively.

Figure 4. Experiment 1 RDM correlations. Partial Spearman participant x model correlations for the Uncrossed (A, green squares) and Crossed group (B) 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 ($t(44) = 2.88$, $p = .006$, $CI = .13 - .74$, $d = .85$). All t-statistics reflect two-tailed independent tests. All panels reflect Uncrossed n = 24 and Crossed n = 22.

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; Crossed=blue triangle,). Error bars are 95% CIs. Gray dots represent individual participants. The p-value reflects a one-tailed independent t-test $t(222) = 2.30$, $p = .011$, $CI = .011 - inf$, $d = .31$. (C) Sub-set of All-pairs trials for which Divisive normalisation and Absolute encoding make different predictions (see Supplementary Methods II). Coloured symbols represent group means and 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, plotted separately for Uncrossed and Crossed. Bar height reflects means and error bars reflect 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. Across all panels: Uncrossed n=119 and Crossed n=105.

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). Panel A-D reflect averages over Uncrossed n=119 and Crossed n=105 respectively.

Figure 7. Experiment 2 Model RDM correlations. Partial Spearman participant x model correlations for the Uncrossed group (A, green squares) and Crossed group (B, blue triangles). Each plot 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 key Crossed-Uncrossed group-wise contrasts assessing whether the evidence in favour of the absolute model over the relative model was stronger in the Crossed group: $t(222) = 3.25$, $p < .001$, lCI = .17-*inf*, $d = .43$); $t(222) = 3.09$, $p = .001$, CI = .15-*inf*, $d = .41$). (D) As (C) but for independent correlations. The p-values reflect key across-group contrasts: ($t(222) = 3.01$, $p = .002$, CI = .05-*inf*, $d = .40$); ($t(222) = 3.15$, $p < .001$, CI = .02-inf, $d = .42$). All t-tests reflect two-tailed independent t-tests. All panel reflects n=119 for the Uncrossed group and n=105 for the Crossed group.

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).
11. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat Rev Neurosci* **13**, 51–62 (2012).
12. Normann, R. A. & Werblin, F. S. Control of retinal sensitivity: I. Light and dark adaptation of vertebrate rods and cones. *The Journal of general physiology* **63**, 37–61 (1974).
13. Stewart, N., Chater, N. & Brown, G. D. A. Decision by sampling. *Cognitive Psychology* **53**, 1–26 (2006).
14. Yamada, H., Louie, K., Tymula, A. & Glimcher, P. W. Free choice shapes normalized value signals in medial orbitofrontal cortex. *Nat Commun* **9**, 162 (2018).
15. Klein, T. A., Ullsperger, M. & Jocham, G. Learning relative values in the striatum induces violations of normative decision making. *Nat Commun* **8**, 16033 (2017).
16. Palminteri, S. & Lebreton, M. *Context-dependent outcome encoding in human reinforcement learning*. https://osf.io/4qh2d (2021) doi:10.31234/osf.io/4qh2d.
17. Rigoli, F. Reference effects on decision-making elicited by previous rewards. *Cognition* **192**, 104034 (2019).
18. Rigoli, F., Mathys, C., Friston, K. J. & Dolan, R. J. A unifying Bayesian account of contextual effects in value-based choice. *PLoS Comput Biol* **13**, e1005769 (2017).
19. Ciranka, S. *et al. Asymmetric learning facilitates human inference of transitive relations*. http://biorxiv.org/lookup/doi/10.1101/2021.04.03.437766 (2021) doi:10.1101/2021.04.03.437766.

20. Gluth, S., Kern, N., Kortmann, M. & Vitali, C. L. Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour* **4**, 634–645 (2020).
21. Rustichini, A., Conen, K. E., Cai, X. & Padoa-Schioppa, C. Optimal coding and neuronal adaptation in economic decisions. *Nat Commun* **8**, 1208 (2017).
22. Bhui, R. & Gershman, S. J. Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review* **125**, 985 (2018).
23. Polanía, R., Woodford, M. & Ruff, C. Efficient coding of subjective value. *Nature Neuroscience* **22**, 134–142 (2019).
24. Pompilio, L. State-Dependent Learned Valuation Drives Choice in an Invertebrate. *Science* **311**, 1613–1615 (2006).
25. Kool, W., Gershman, S. J. & Cushman, F. A. Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychol Sci* **28**, 1321–1333 (2017).
26. Griffiths, T. L. *et al.* Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences* **29**, 24–30 (2019).
27. James, W. *The principles of psychology*. vol. 1 (Henry Holt & Co, 1890).
28. Anderson, J. R. *The adaptive character of thought*. (Psychology Press, 2013).
29. Payne, J. W., Bettman, J. R. & Johnson, E. J. *The adaptive decision maker*. (Cambridge university press, 1993).
30. Anderson, J. (1990). The adaptive character of thought. *Hillsdale, NJ: Erl-baum*.
31. Kahneman, D. & Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**, 263–292 (1979).
32. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences* **103**, 3863–3868 (2006).
33. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol* **10**, e1003553 (2014).
34. Luyckx, F., Nili, H., Spitzer, B. & Summerfield, C. Neural structure mapping in human probabilistic reward learning. 1–16 (2018) doi:10.1101/366757.

35. Sheahan, H., Luyckx, F., Nelli, S., Teupe, C. & Summerfield, C. *Neural state space alignment for magnitude generalisation in humans and recurrent networks*. http://biorxiv.org/lookup/doi/10.1101/2020.07.22.215541 (2020) doi:10.1101/2020.07.22.215541.
36. Hunt, L. T. *et al.* Triple dissociation of attention and decision computations across prefrontal cortex. *Nature neuroscience* **21**, 1471–1481 (2018).
37. Hertwig, R., Barron, G., Weber, E. U. & Erev, I. Decisions From Experience and the Effect of Rare Events in Risky Choice. *Psychological Science* **15**, 6 (2004).
38. Bavard, S., Rustichini, A. & Palminteri, S. The construction and deconstruction of sub-optimal preferences through range-adapting reinforcement learning (2020). http://biorxiv.org/content/early/2020/07/29/2020.07.28.224642.abstract
39. Hotaling, J. M., Jarvstad, A., Donkin, C. & Newell, B. R. How to Change the Weight of Rare Events in Decisions From Experience. *Psychol Sci* **30**, 1767–1779 (2019).
40. Shenhav, A. *et al.* Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience* **40**, 99–124 (2017).
41. Prat-Carrabin, A. & Woodford, M. *Efficient coding of numbers explains decision bias and noise*. http://biorxiv.org/lookup/doi/10.1101/2020.02.18.942938 (2020) doi:10.1101/2020.02.18.942938.
42. Juechems, K., Balaguer, J., Spitzer, B. & Summerfield, C. Optimal utility and probability functions for agents with finite computational precision. *Proceedings of the National Academy of Sciences* **118**, (2021).
43. Spektor, M. S., Gluth, S., Fontanesi, L. & Rieskamp, J. How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological Review* **126**, 52–88 (2019).
44. Collins, A. G. E. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working memory in reinforcement learning. *European Journal of Neuroscience* **35**, 1024–1035 (2012).

45. Hayes, W. M. & Wedell, D. H. Regret in experience-based decisions: The effects of expected value differences and mixed gains and losses. *Decision* **8**, 277–294 (2021).

46. Edwards, D. J., Pothos, E. M. & Perlman, A. Relational versus absolute representation in categorization. *Am J Psychol* **125**, 481–497 (2012).

47. Collins, A. G. E. & Cockburn, J. Beyond dichotomies in reinforcement learning. *Nat Rev Neurosci* **21**, 576–586 (2020).

48. Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J. & Daw, N. D. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput Biol* **13**, e1005768 (2017).

49. Koechlin, E. Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology* **37**, 1–6 (2016).

50. Gershman, S. J., Horvitz, E. J. & Tenenbaum, J. B. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science* **349**, 273–278 (2015).

51. Hunter, L. E. & Gershman, S. J. *Reference-dependent preferences arise from structure learning*. http://biorxiv.org/lookup/doi/10.1101/252692 (2018) doi:10.1101/252692.

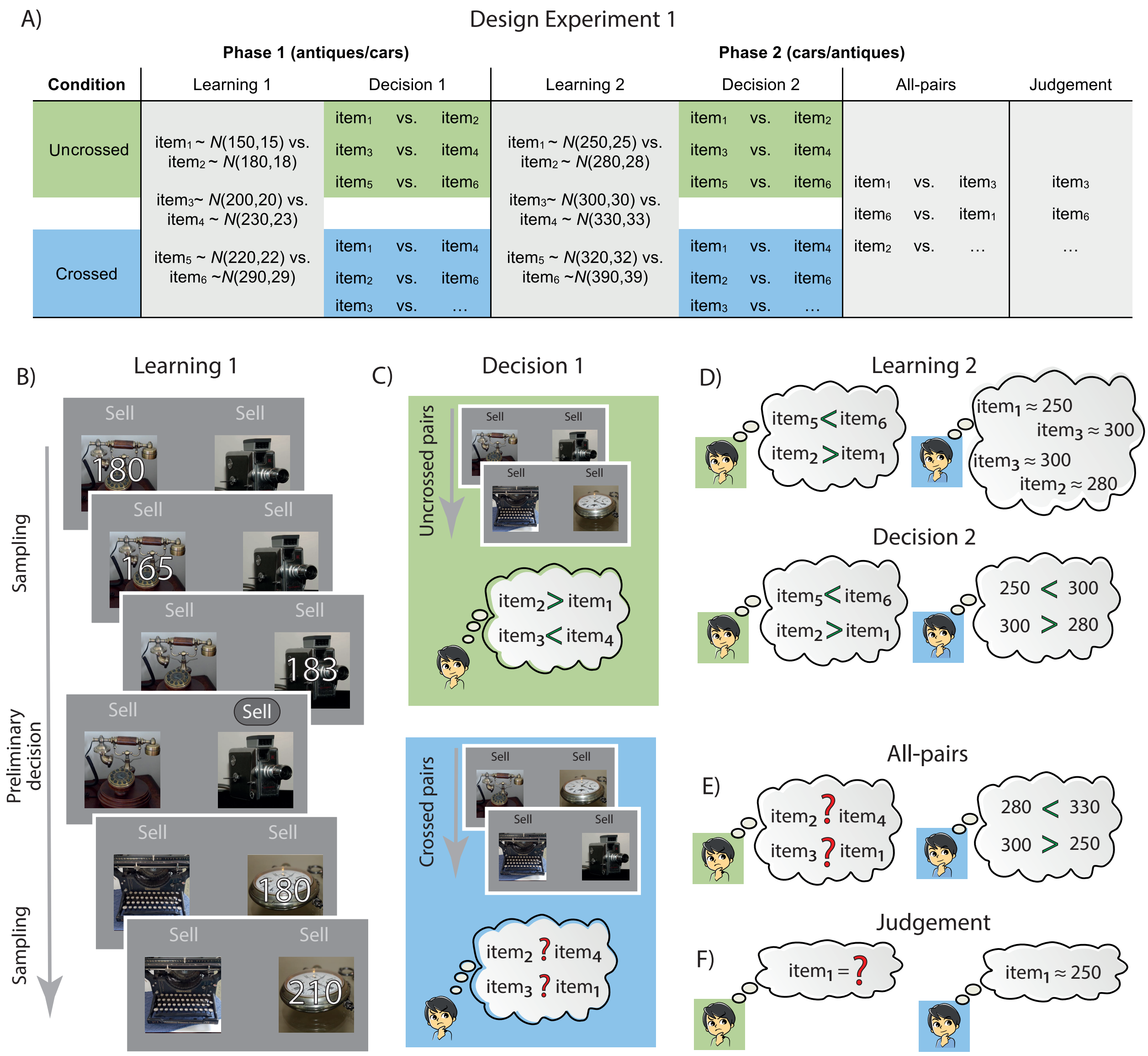
52. Lieder, F., Shenhav, A., Musslick, S. & Griffiths, T. L. Rational metareasoning and the plasticity of cognitive control. *PLoS Comput Biol* **14**, e1006043 (2018).

53. Vlaev, I., Chater, N., Stewart, N. & Brown, G. D. A. Does the brain calculate value? *Trends in Cognitive Sciences* **15**, 546–554 (2011).

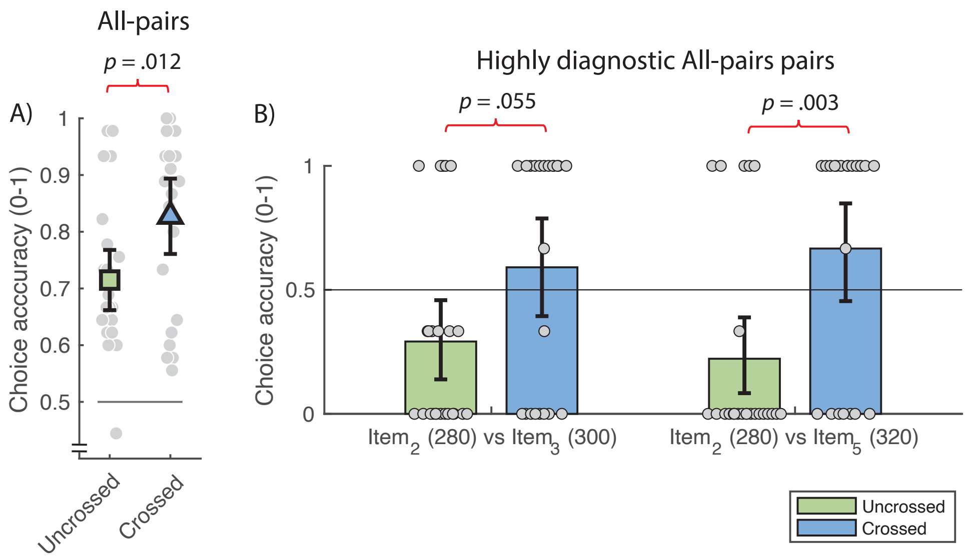
54. Hayden, B. & Niv, Y. *The case against economic values in the orbitofrontal cortex (or anywhere else in the brain)*. https://osf.io/7hgup (2020) doi:10.31234/osf.io/7hgup.

55. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? *Perception* **36**, 1–16 (2007).

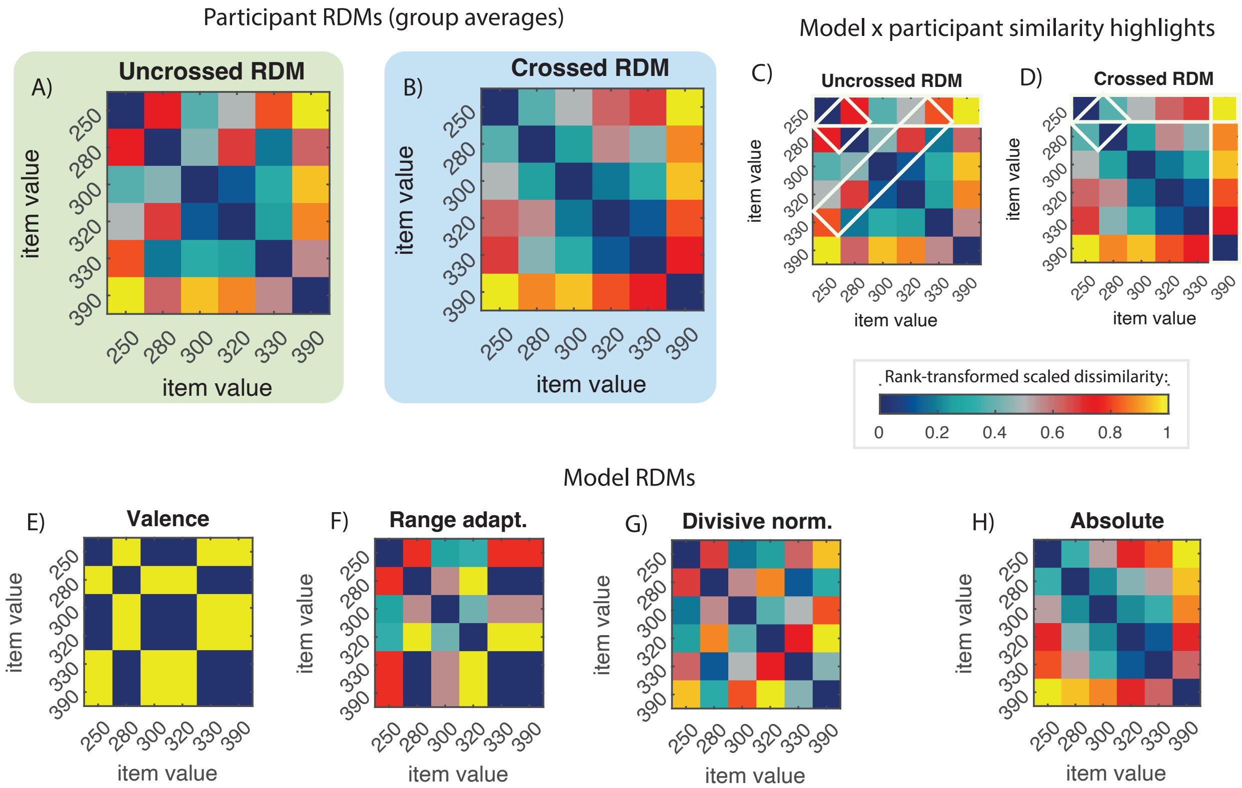
56. Fox, C. R. & Hadar, L. "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making* **1**, 3 (2006).

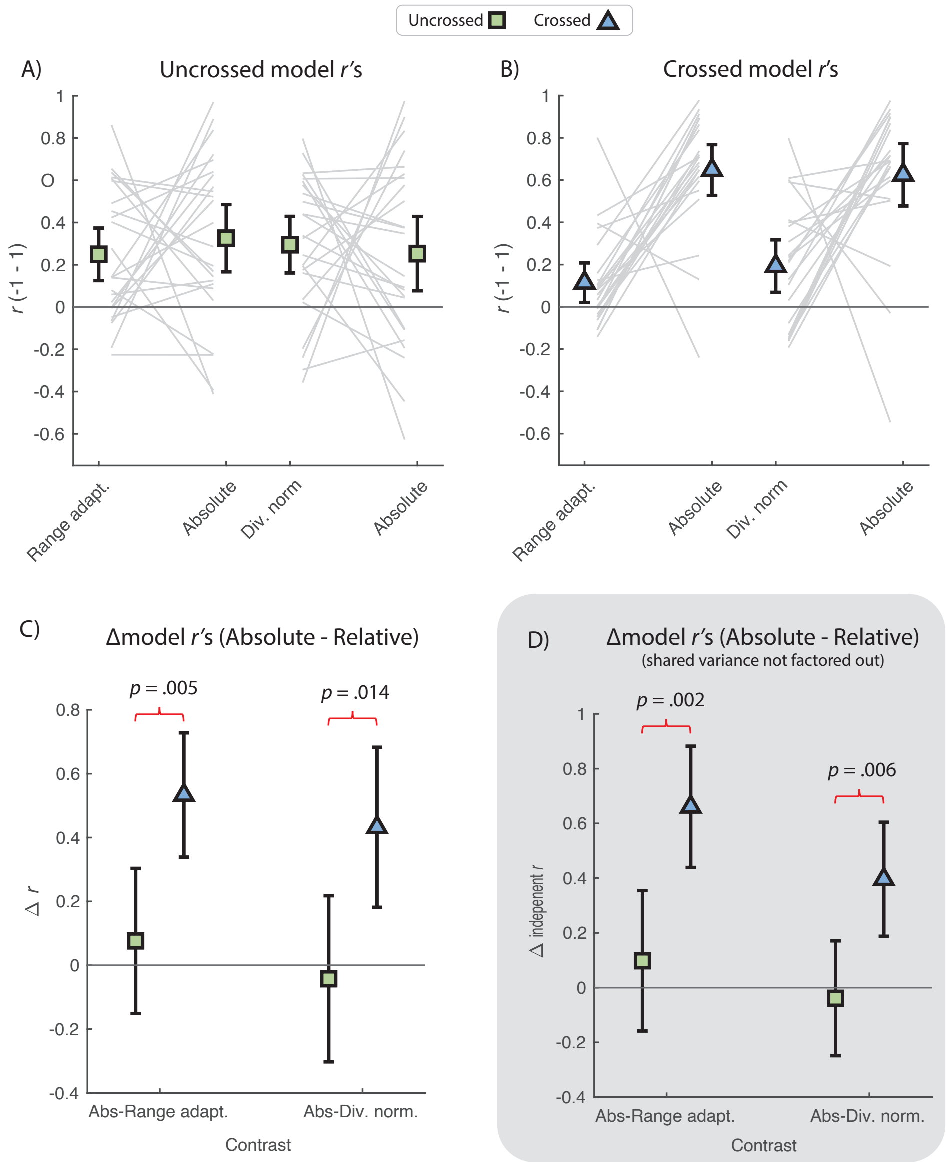


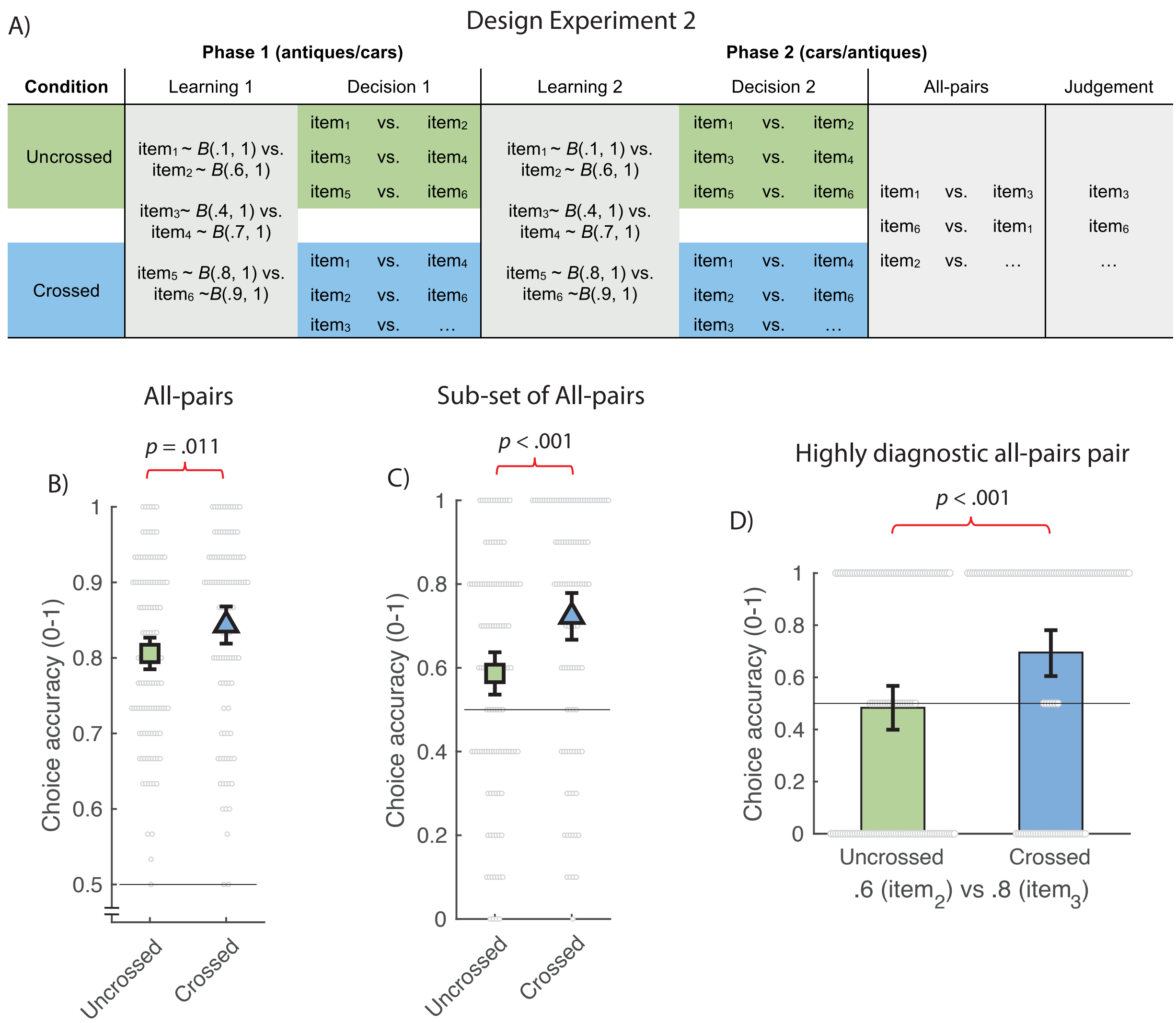
A)
Design Experiment 1
Phase 1 (antiques/cars)
Phase 2 (cars/antiques)
Condition
Learning 1
Decision 1
Learning 2
Decision 2
All-pairs
Judgement
Uncrossed
Crossed
$item_1 \sim N(150,15)$ vs. $item_2 \sim N(180,18)$
$item_3 \sim N(200,20)$ vs. $item_4 \sim N(230,23)$
$item_5 \sim N(220,22)$ vs. $item_6 \sim N(290,29)$
$item_1$ vs. $item_2$
$item_3$ vs. $item_4$
$item_5$ vs. $item_6$
$item_1$ vs. $item_4$
$item_2$ vs. $item_6$
$item_3$ vs. …
$item_1 \sim N(250,25)$ vs. $item_2 \sim N(280,28)$
$item_3 \sim N(300,30)$ vs. $item_4 \sim N(330,33)$
$item_5 \sim N(320,32)$ vs. $item_6 \sim N(390,39)$
$item_1$ vs. $item_2$
$item_3$ vs. $item_4$
$item_5$ vs. $item_6$
$item_1$ vs. $item_4$
$item_2$ vs. $item_6$
$item_3$ vs. …
$item_1$ vs. $item_3$
$item_6$ vs. $item_1$
$item_2$ vs. …
$item_3$
$item_6$
…
B)
Learning 1
Sampling
Preliminary decision
Sampling
Sell
180
165
183
180
210
C)
Decision 1
Uncrossed pairs
Crossed pairs
Sell
$item_2 > item_1$
$item_3 < item_4$
$item_2$? $item_4$
$item_3$? $item_1$
D)
Learning 2
$item_5 < item_6$
$item_2 > item_1$
$item_1 \approx 250$
$item_3 \approx 300$
$item_3 \approx 300$
$item_2 \approx 280$
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
$item_1 =$?
$item_1 \approx 250$

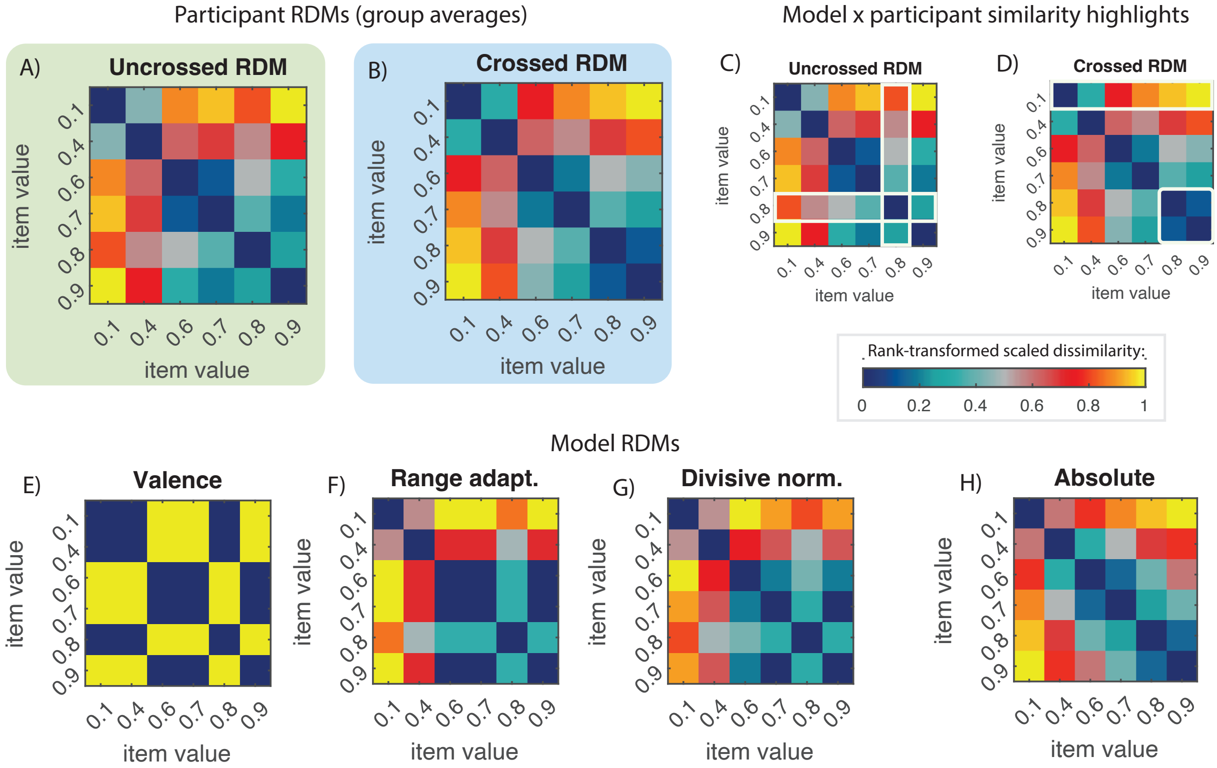


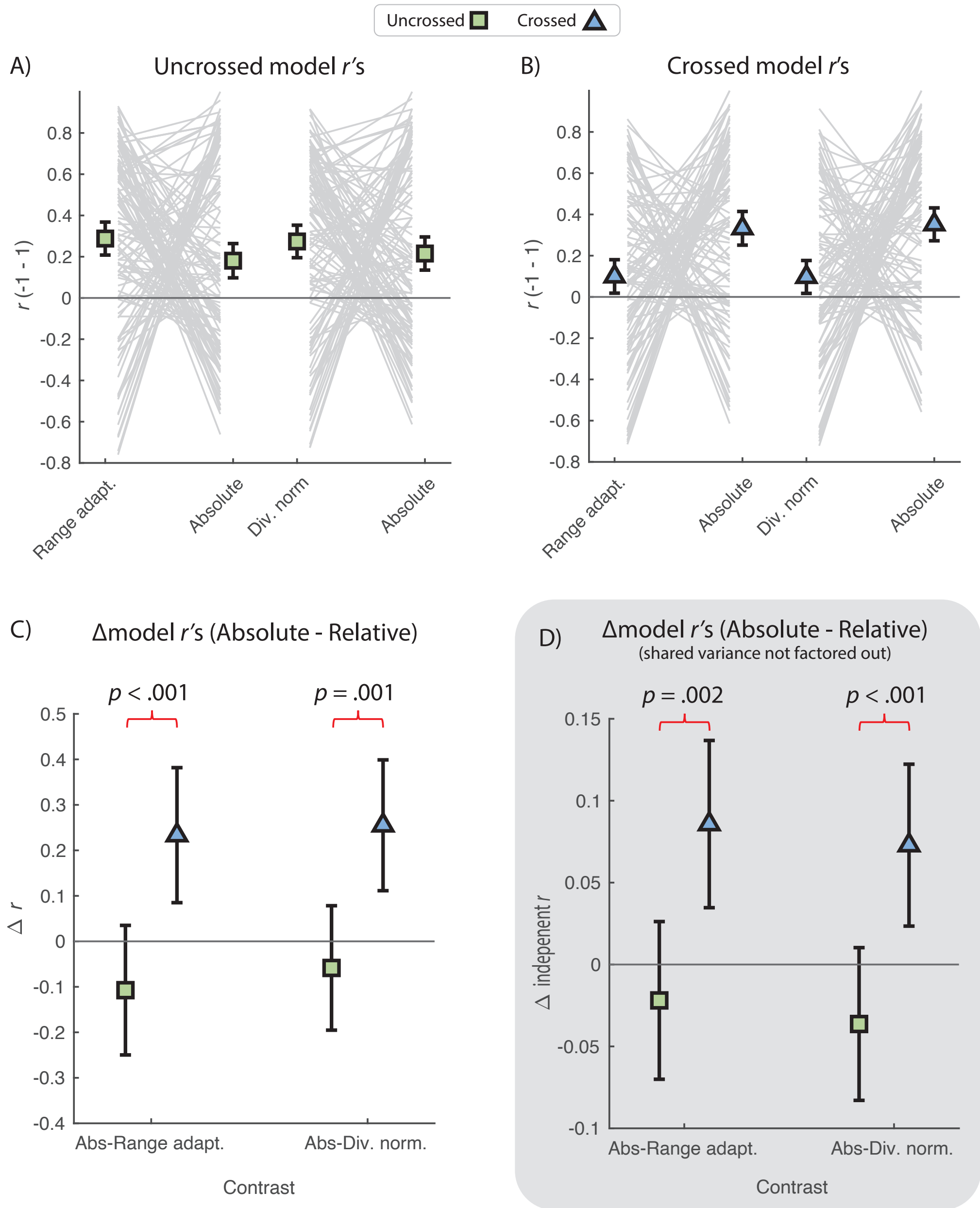
All-pairs
p = .012
A)
Choice acccuracy (0-1)
1
0.9
0.8
0.7
0.6
0.5
Uncrossed
Crossed
Highly diagnostic All-pairs pairs
B)
p = .055
p = .003
Choice accuracy (0-1)
1
0.5
0
Item2 (280) vs Item3 (300)
Item2 (280) vs Item5 (320)
Uncrossed
Crossed


Participant RDMs (group averages)
A)
Uncrossed RDM
B)
Crossed RDM
item value
250
280
300
320
330
390
Model x participant similarity highlights
C)
Uncrossed RDM
D)
Crossed RDM
Rank-transformed scaled dissimilarity:
0
0.2
0.4
0.6
0.8
1
Model RDMs
E)
Valence
F)
Range adapt.
G)
Divisive norm.
H)
Absolute


Uncrossed
Crossed
A)
Uncrossed model r's
B)
Crossed model r's
r (-1 - 1)
1
0.8
O
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Range adapt.
Absolute
Div. norm
Absolute
C)
Δmodel r's (Absolute - Relative)
p = .005
p = .014
Δ r
Abs-Range adapt.
Abs-Div. norm.
Contrast
D)
Δmodel r's (Absolute - Relative)
(shared variance not factored out)
p = .002
p = .006
Δ indepenent r
Abs-Range adapt.
Abs-Div. norm.
Contrast


A)
Design Experiment 2
Phase 1 (antiques/cars)
Phase 2 (cars/antiques)
Condition
Learning 1
Decision 1
Learning 2
Decision 2
All-pairs
Judgement
Uncrossed
Crossed
item1 ~ B(.1, 1) vs. item2 ~ B(.6, 1)
item3~ B(.4, 1) vs. item4 ~ B(.7, 1)
item5 ~ B(.8, 1) vs. item6 ~B(.9, 1)
item1 vs. item2
item3 vs. item4
item5 vs. item6
item1 vs. item4
item2 vs. item6
item3 vs. …
item1 vs. item3
item6 vs. item1
item2 vs. …
item3
item6
…
All-pairs
B)
p = .011
Choice accuracy (0-1)
Uncrossed
Crossed
Sub-set of All-pairs
C)
p < .001
Choice accuracy (0-1)
Uncrossed
Crossed
Highly diagnostic all-pairs pair
D)
p < .001
Choice accuracy (0-1)
Uncrossed
Crossed
.6 (item2) vs .8 (item3)


Participant RDMs (group averages)
Model x participant similarity highlights
A)
Uncrossed RDM
B)
Crossed RDM
C)
Uncrossed RDM
D)
Crossed RDM
item value
0.1
0.4
0.6
0.7
0.8
0.9
Rank-transformed scaled dissimilarity:
0
0.2
0.4
0.6
0.8
1
Model RDMs
E)
Valence
F)
Range adapt.
G)
Divisive norm.
H)
Absolute



Uncrossed
Crossed
A)
Uncrossed model r's
B)
Crossed model r's
r (-1 - 1)
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Range adapt.
Absolute
Div. norm
Absolute
C)
Δmodel r's (Absolute - Relative)
p < .001
p = .001
Δ r
Abs-Range adapt.
Abs-Div. norm.
Contrast
D)
Δmodel r's (Absolute - Relative)
(shared variance not factored out)
p = .002
p < .001
Δ independent r
0.15
0.1
0.05
0
-0.05
-0.1
Abs-Range adapt.
Abs-Div. norm.
Contrast

893