Neural signatures of model-free learning when avoiding harm to self and other

Patricia L. Lockwood[a,b*^], Miriam Klein-Flügge[a,b*^], Ayat Abdurahman[a,b], & Molly J. Crockett[a,c*]

[a]Department of Experimental Psychology, University of Oxford, Oxford OX1 3PH, United Kingdom

[b]Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford

[c]Department of Psychology, Yale University, New Haven, CT, 06511, USA

*Correspondence should be addressed to:

Patricia L. Lockwood, Experimental Psychology, Tinsley Building, University of Oxford, OX1 3SR, United Kingdom. E-mail: patricia.lockwood@psy.ox.ac.uk

Miriam C Klein- Flügge, Experimental Psychology, Tinsley Building, University of Oxford, OX1 3SR, United Kingdom. E-mail: miriam.klein-flugge@psy.ox.ac.uk

Molly J. Crockett, Department of Psychology, 2 Hillhouse Avenue, New Haven, CT 06511, USA. E-mail: molly.crockett@yale.edu

^Indicates joint first authorship

1 **Abstract**

2

3 Moral behaviour requires learning how our actions help or harm others. Theoretical

4 accounts of learning propose a key division between 'model-free' algorithms that

5 efficiently cache outcome values in actions and 'model-based' algorithms that

6 prospectively map actions to outcomes, a distinction that may be critical for moral

7 learning. Here, we tested the engagement of these learning mechanisms and their

8 neural basis as participants learned to avoid painful electric shocks for themselves

9 and a stranger. We found that model-free learning was prioritized when avoiding harm

10 to others compared to oneself. Model-free prediction errors for others relative to self

11 were tracked in the thalamus/caudate at the time of the outcome. At the time of choice,

12 a signature of model-free moral learning was associated with responses in subgenual

13 anterior cingulate cortex (sgACC), and resisting this model-free influence was

14 predicted by stronger connectivity between sgACC and dorsolateral prefrontal cortex.

15 Finally, multiple behavioural and neural correlates of model-free moral learning varied

16 with individual differences in moral judgment. Our findings suggest moral learning

17 favours efficiency over flexibility and is underpinned by specific neural mechanisms.

18   **Introduction**

19

20   A central component of human morality is a prohibition against harming others[1,2].

21   People readily avoid actions that might harm another person[3–7], and this basic harm

22   aversion is so strong that many people even find it distressing to perform pretend

23   harmful actions, such as shooting someone with a fake gun[8]. Harm aversion is

24   disrupted in clinical disorders such as psychopathy that have a strong developmental

25   component[9], and although harm aversion is robust in healthy adults, anyone who has

26   watched young children fighting over a coveted toy knows that such an aversion is not

27   present from birth. Indeed, a large literature documents the emergence of moral

28   conduct over the course of development[10,11]. Cross-cultural differences in morality

29   suggest moral behavior is fine-tuned to local environmental demands[12], and lab

30   experiments demonstrate how individuals can quickly adapt moral behavior to

31   changing norms[13,14]. All this evidence highlights a critical role for *learning* in the

32   development of harm aversion and moral behaviour more broadly[6].

33

34   Recent work in computational neuroscience has advanced our knowledge of how

35   organisms learn the value of actions and outcomes via reward and punishment[15,16].

36   An important theoretical distinction has been made between 'model-based' and

37   'model-free' learning[17,18]. The computationally expensive model-based system builds

38   an internal model of the environment and selects actions by prospectively searching

39   the internal model for the best course of action[19,20]. The computationally efficient

40   model-free system assigns values to actions through trial-and-error and selects

41   actions retrospectively based on these 'cached' values; because it does not store a

42   model of the environment, it can sometimes make sub-optimal recommendations

43   (e.g., in settings where typically valuable actions lead to bad outcomes, or vice versa).

44   These two systems are somewhat neurally dissociable, with model-based learning

45   preferentially engaging lateral prefrontal cortex (LPFC), posterior parietal cortex and

46   caudate[20–22] and model-free learning preferentially engaging putamen[23,24], although

47   both systems update their representations via prediction errors encoded in

48   overlapping regions of ventral striatum[20]. Model-based and model-free systems often

49   make similar recommendations about which actions are more valuable, but when they

50 conflict, an arbitration process allocates control between them[22,25,26]. However,
51 despite extensive theorizing that the model-based/model-free distinction may help to
52 characterize puzzling features of moral learning and decision-making[3,27–29], it remains
53 unknown whether the moral consequences of actions affect the balance between
54 model-based and model-free control, and whether common or distinct neural
55 processes are engaged when learning to avoid harmful outcomes to self and others.
56
57 Past work on the neural basis of moral decision-making provides support for
58 competing hypotheses. On the one hand, the sophistication of human morality seems
59 to demand the kinds of complex representations afforded by model-based learning,
60 suggesting learning to avoid harming others may preferentially engage the model-
61 based system. Supporting this view, people are easily able to learn to avoid harmful
62 actions without directly experiencing their outcomes through trial and error, as may be
63 required for model-based learning[3,27,30]. Moreover, moral decision-making in healthy
64 adults consistently engages brain regions most strongly associated with the model-
65 based system, including lateral prefrontal cortex (LPFC), caudate, and
66 temporoparietal junction (TPJ)[23,25,31]. Deciding to follow moral norms like fairness and
67 honesty, and enforcing those norms on others via costly punishment, engages
68 LPFC[32–37], and disrupting LPFC function reduces moral norm compliance and
69 enforcement[38,39]. During decisions to avoid harming others, LPFC encodes the
70 blameworthiness of harmful choices and modulates action values in caudate and
71 thalamus[4], two subcortical areas shown to play a critical role in associative learning
72 and pain processing as well as moral decision-making[40–45]. TPJ has been implicated
73 in sophisticated representations of others' mental states and integrating these into
74 social decisions[46–48].
75
76 On the other hand, one principal function of model-free learning is to cache value in
77 actions that are reliably adaptive, sacrificing flexibility for efficiency. Given that harming
78 others is typically prohibited, actions that harm others may represent a special class
79 of actions that are prioritized for model-free learning, similar to how certain classes of
80 stimuli, like snakes and spiders, are "prepared" for aversive classical conditioning[49].
81 In other words, since avoiding harm to others is hugely important for social life,

82   learning processes that fast-track harm-avoidant action selection to a habitual,

83   automatic process may be socially adaptive. Supporting this view, recent work

84   suggests that morality constrains mental representations of what actions are

85   considered possible; harmful actions are removed from choice sets as a default[50], and

86   choices that harm others are slower than helpful choices, suggesting an automatic

87   tendency to avoid harm[5,51,52]. Furthermore, recent studies of model-free learning to

88   gain rewards for oneself and others have highlighted a distinct encoding of prediction

89   errors concerning others' outcomes in the subgenual anterior cingulate cortex

90   (sgACC)[53,54], a region that has been implicated in social and moral decision-making

91   more broadly[54–58]. Model-free processes that distinguish learning about how one's

92   actions affect others could provide a neural mechanism for prioritizing model-free

93   learning in moral contexts.

94

95   To test these competing hypotheses, we used computational modelling and fMRI to

96   probe the relative balance between model-based versus model-free processes, and

97   their neural bases, when people learn to avoid moderately painful electric shocks for

98   themselves and a stranger. After undergoing a pain thresholding procedure (see

99   Methods), participants (N=36) completed a hybrid version of two paradigms previously

100   proposed to reliably dissociate model-free versus model-based learning (Figure

101   1)[20,22,31]. We optimized the task in a way that allowed us to address the specific

102   hypotheses examined in the present study (see Supplementary Note on Experimental

103   Design for details).
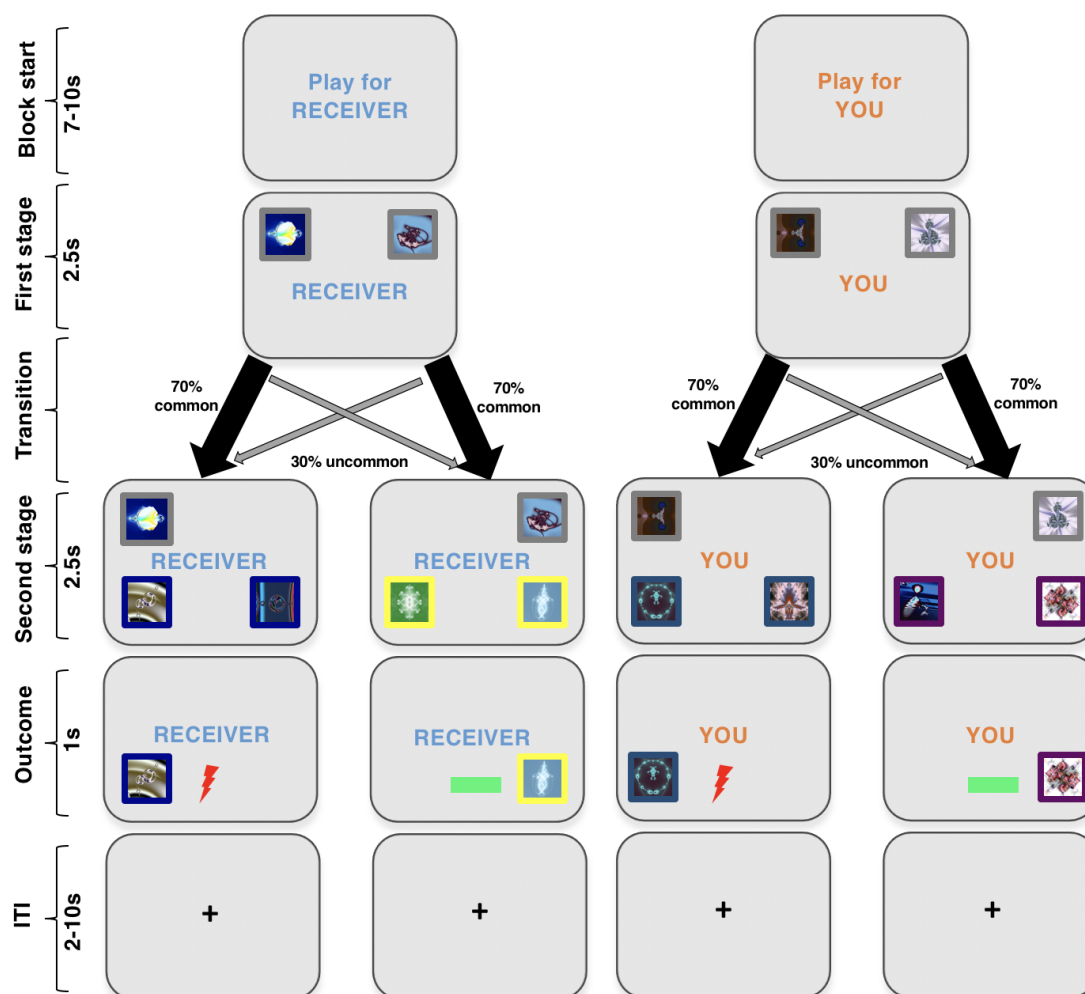
Figure 1 – Model-free and model-based aversive learning task. Participants completed a two-stage decision-making task to assess the tendency to engage in model-free and model-based learning. The task was a hybrid of two tasks previously shown to assess model-free and model-based learning processes[20,26] to probe learning to avoid aversive (shock) outcomes for either oneself or another person (the 'receiver', referred to as 'other' hereafter). At the beginning of each block an instruction cue signalled the recipient of the outcome (self or other). At the first stage two images were displayed that probabilistically led to one of two states depicted by different colours surrounding the boxes. In this example, to 'blue zone' or 'yellow zone' for the other participant, and 'turquoise zone' or 'purple zone' for self. Participants then made a second choice between two pictures in the coloured zone which was followed by an outcome of shock or no shock. The probability with which the boxes at the second stage delivered a shock or no-shock outcome drifted throughout the experiment (bounded between 0% and 100% with a drift rate of 0.2) and participants were instructed to keep learning throughout. 10% of the total electric shocks accumulated in the 'self' condition were delivered to the participant themselves at the end of the experiment whilst 10% of the electric shocks accumulated in the 'other' condition were delivered to the partner participant.

Results

Model-free learning is prioritized when avoiding harm to others vs. self

126 Participants completed a two-step decision-making task to index model-free and
127 model-based learning strategies (**Figure 1**). Prior to scanning, participants were
128 trained on the transition structure of the task using stimuli different from the main
129 experiment which allowed them to learn the probabilistic transition structure. The two-
130 step task distinguishes model-free and model-based learning by measuring people's
131 choices to stay or switch based on the outcome of the previous trial and the transition
132 structure. Theoretically, a purely model-free learner would ignore the transition
133 structure and repeat first-stage choices if they prevented pain on the previous trial but
134 switch choices if the previous choice caused pain. Thus, model-free learning is
135 reflected in a main effect of outcome (pain versus no pain) on subsequent first-stage
136 choice behaviour. In contrast, a model-based learner would take the transition
137 structure into account. While behaviour on common transitions would be similar for a
138 model-free agent, after rare transitions, a model-based agent would repeat a first-
139 stage choice if the outcome was pain, but switch if the outcome was no pain. Thus,
140 model-based learning is reflected in an interaction between outcome (pain versus no
141 pain) and transition (common versus rare). It is now well established that people
142 display a combination of model-free and model-based learning strategies when
143 learning about rewarding outcomes[20,31,59] and initial evidence indicates that the same
144 is true for learning about aversive outcomes for oneself[60,61]. We therefore first
145 examined whether participants displayed a combination of model-free and model-
146 based processes during aversive learning for oneself and others, as observed in these
147 previous studies.

148

149 We performed a logistic regression analysis predicting first-stage stay versus switch
150 choices as a function of the outcome (pain or no pain), transition on the previous trial
151 (common or rare) and recipient (Self vs. Other). We included all main effects and
152 interactions in the model. We found a significant main effect of outcome ($t(35)=4.618$,
153 $p<.001$, confidence interval (CI) for beta estimate: 0.17, 0.42, Cohen's $d = 0.77$)
154 indicating a contribution of model-free learning, but also a significant transition by
155 outcome interaction ($t(35)=-3.173$, $p=.003$, CI for beta estimate: -0.23, -0.05, $d = -$
156 0.53), indicating the presence of model-based learning (**Figure 2a,b**). This

157 demonstrates that similar to reward learning, aversive learning is underpinned by a
158 mixture of model-based and model-free processes.

159

160 Intriguingly, we also found a significant interaction between outcome (the model-free
161 contribution to learning) and recipient, showing that people were more model-free for
162 others relative to self (t(35)=-2.31, p=.027, CI of beta estimates -0.12, -0.008, d = -
163 0.39). By contrast, there was no interaction between recipient and transition x
164 outcome, the model-based component of learning (t(35)=-.459, p>.65, CI of beta
165 estimates: -0.071, 0.045, d = -0.08).

166

167 We validated these behavioural results by repeating our analysis using the lme4
168 package in R which ensured we had good estimates of random effects and accounted
169 for variability in behaviour using Bound Optimization by Quadratic Approximation (see
170 Methods). Our model predicted the tendency to switch vs. stay including all main
171 effects   and   interactions   and   a   random   effect   of   subject,
172 glmer(Switch_Stay~NoPain_Pain*Transition*Agent+(1|Subject). All results remained
173 the same (main effect of outcome (model-free) Z=5.949, p<.001; Outcome x transition
174 interaction (model-based) Z=5.484, p<.001; outcome x recipient Z=-2.154, p=0.031;
175 [outcome x transition (model-based)] x recipient Z=0.696, p=0.4864). Again, these
176 findings support the idea that people were more model-free when avoiding harming
177 others.

178

179 To further examine which outcomes most influenced the outcome x recipient
180 interaction, we performed separate post-hoc tests on the percentage of stay/switch
181 choices for self and other following (a) only pain outcomes and (b) only no pain
182 outcomes. This showed that the recipient difference was driven by increased switching
183 after pain outcomes for other versus self (p=.028), with no difference between
184 recipients in the proportion of stay/switch choices after no pain outcomes (p=.552).
185 This is important because the specificity of the effect rules out that people were simply
186 more indifferent or inattentive to the outcomes of others compared to self.

187
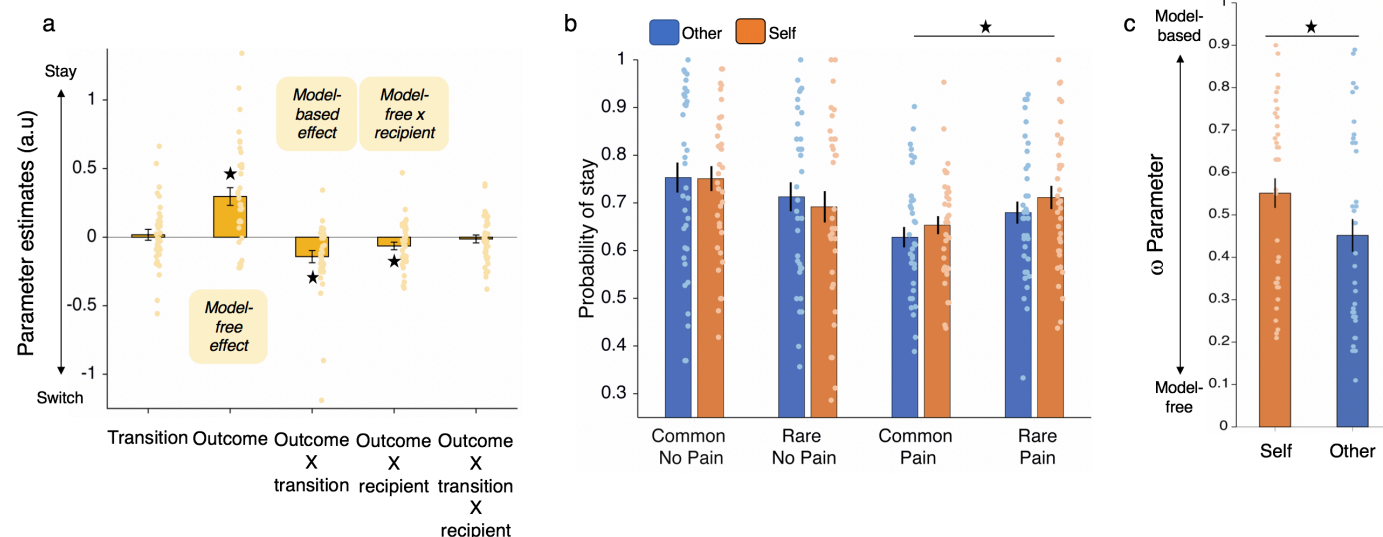
Figure 2 – Model-free and model-based choices when avoiding harming oneself and others. (a) Logistic regression coefficients predicting first-level stay/switch choices (mean ± SEM). Participants exhibited a main effect of staying after no pain which indicated model free behaviour ($t(35)=4.618$, $p<.001$, confidence interval (CI) for beta estimate: 0.17, 0.42, Cohen's $d = 0.77$) and a outcome by transition interaction which indicated model-based behaviour ($t(35)=-3.173$, $p=.003$, CI for beta estimate: -0.23, -0.05, $d = -0.53$). Intriguingly there was also a outcome x recipient interaction (($t(35)=-2.31$, $p=.027$, CI of beta estimates -0.12, -0.008, $d = -0.39$) showing that participants were more model-free, and thus more likely to switch after pain and stay after no pain, independent of transition type, when making choices for another person. (b). The probability of repeating a choice at the first-level ('stay') is plotted as a function of the transition and outcome on the previous trial. This shows that the outcome x recipient interaction in (a) is mostly driven by fewer stay trials after pain, regardless of transition (two rightmost blue vs orange bars). Thus, the more pronounced model free behaviour for others is mostly driven by a lower probability of staying after pain outcomes rather than a higher probability of staying after no pain outcomes). (c) ω estimates from the best fitting model showed that the ω parameter was significantly lower for other (0.45) than self (0.55) consistent with the regression analyses that showed people were more model free when avoiding harm to others compared to self ($p<.02$). Asterisks indicate significant difference at $p<.05$.

Computational modelling of aversive learning for self and other

Next, we fitted several trial-by-trial computational models to our data to examine further which model best captured the described behaviours during aversive learning for self and other. Deriving such trial-by-trial estimates that capture individual choice preferences was a prerequisite for modelling the fMRI data and allowed us to support our logistic regression analyses. We started with the full seven-parameter model proposed by Daw and colleagues[20] and compared this model to similar models with fewer parameters (four of five) following modifications similar to those suggested in previous studies (e.g.[62,63]); for details, see Methods and **Supplementary Table 2**). We also included variants of the same models that involved separate learning rates

218   for pain and no pain outcomes, given evidence suggesting differential learning as a

219   function of outcome valence (e.g.[64,65]). All of these models were initially fitted

220   separately on self and other blocks.

221

222   We found that a five-parameter model best explained behaviour compared to all

223   alternative models tested. This model included separate learning rates for no pain and

224   pain outcomes (αPain, αNoPain), a single temperature parameter capturing choice

225   randomness (β), a perseverance parameter capturing a tendency to stick with the

226   previously made choice (ρ) and a model-free/model-based weighting parameter (ω).

227   Importantly, this five-parameter model best explained behaviour in both the self and

228   other blocks (**Supplementary Table 2**).

229

230   We next compared the different estimated parameters for the self and other blocks.

231   This analysis showed a significant difference between the conditions in both the

232   perseverance parameter ρ (t(35)=2.41, p=.02,) and the model-free/based weighting

233   parameter ω (t(35)=3.10, p=.0039; **Supplementary Table 3**). To compare these

234   parameters more robustly, we then used maximum a-posteriori estimation performed

235   on the pooled data of self and other blocks to compare three models, one with

236   separate perseverance ρ and ω parameters for self and other (and thus a total of seven

237   parameters), one with separate ω parameters for self and other (and therefore a total

238   of six parameters), and the original 5 parameter model (which assumes the same ρ

239   and ω across self and other blocks). We used these models to examine whether

240   differences in the ρ and ω parameters between self and other when fitted separately

241   reflected true differences in the weighting of these parameters in a model comparison.

242   This analysis showed that the model with separate ω's for self and other, but not

243   separate ρ's, best explained the data. Importantly, these ω parameters were also

244   significantly different from one another (selfω = 0.55, otherω = 0.45, t(35)=2.41,

245   p=.02, d=0.40, **Figure 2c and Supplementary Table 4 and 5**). Thus, consistent with

246   the regression based behavioural analyses that did not rely on a computational model

247   (**Figure 2a**), participants were more model-free than model-based when learning to

248   avoid harming others, compared to self (**Figure 2c**).

249

Subcortical areas distinguish model-free prediction errors for self and other

Previous neuroimaging studies of model-based and model-free reward learning have reported model-free prediction error signals in ventral striatum[20]. We therefore first sought to replicate this effect in our aversive learning paradigm. To facilitate comparison with previous studies of reward learning, no-pain outcomes were coded as 1 and pain outcomes coded as -1. Therefore, a positive prediction error represents unexpected pain relief/avoidance, and a negative prediction error represents unexpected pain.

We built a general linear model (GLM1) that contained onsets for the first stage choice, second stage choice and outcome separately for self and other trials. These three time periods were each associated with parametric modulators from our winning model. These included the value difference between the two options at the first stage choice; the state prediction error based on the transition at the second stage choice; and the model-free prediction error at the time of the outcome. We focused our analysis on model-free prediction errors at the time of the outcome for two reasons. Firstly, our behavioural effects showed that self/other differences in learning emerged for model-free but not model-based learning. Secondly, model-free and model-based prediction errors are highly correlated and careful examination of their separate influences has shown that they are both encoded in ventral striatum[20].

We began our analyses by examining whether previously reported neural correlates of value difference and state prediction errors were also observed in our novel paradigm. Several areas tracked inverse value difference and thus showed larger responses for choices that had a smaller value difference between the two first-stage options, including the most dorsal parts of anterior cingulate cortex near pre-SMA, bilateral inferior parietal cortex and middle frontal gyrus (**Supplementary table 1**), regions previously associated with the tracking of inverse subjective value difference[66–68]. These signals did not differ between self and other (see **Supplementary Table 1**). Also consistent with previous findings[21], we found evidence of a main effect for state prediction errors at the second stage in dorsal ACC (x=-6,

282   y=10, z=52, Z=4.85, K=906, p<.001 FWE-corrected) that again showed overlap

283   between self and other (see **Supplementary table 1**).

284

285   Next, we tested whether model-free prediction errors were present in ventral striatum,

286   as reported in a previous study examining model-based and model-free reward

287   learning for self[20] and several studies of reward based reinforcement learning[69]. We

288   found a large bilateral cluster signalling prediction errors of harm avoidance (positive

289   for no pain, negative for pain) in ventral striatum (right x = 10, y = 12, z = -4, k=236,

290   Z= 5.84, left x = -16, y = 6 , z = -10, k=458, z= 5.77, p<.05 FWE-whole brain corrected

291   after initial thresholding at p<.001). Again, this signal did not significantly differ for self

292   and other conditions (Right $t(34)$ = -.14, p=.89, Left $t(34)$ = -1.47, p = .152)).

293

294
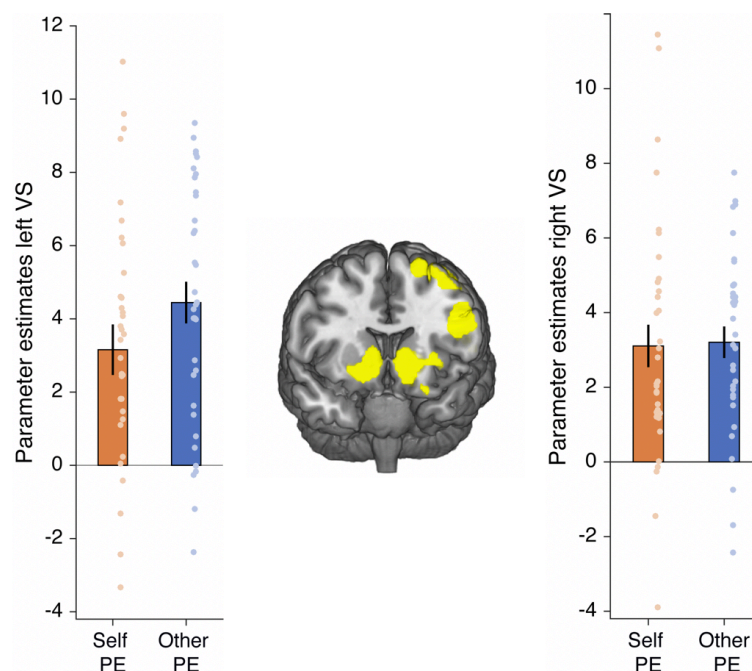
295   **Figure 3 – Ventral striatum encodes prediction errors of pain avoidance for self and other.** Ventral
296   striatum (right x = 10, y = 12, z = -4, k=236, z= 5.84, left x = -16 , y = 6 , z = -10, k=458, z= 5.77, p<.05
297   FWE-whole brain corrected after initial thresholding at p<.001) tracked model-free prediction errors for
298   both self and other bilaterally, with no significant differences between conditions.

299

300   Given that our behavioral results indicated model-free learning was prioritized when

301   avoiding harm to others (relative to self), we next sought to identify areas that

302   distinguished model-free prediction errors for others (relative to self). This analysis

303   revealed a cluster in the thalamus extending into the caudate (x =16, y = -18, z =0,

304   k=84, p=.033, Z=3.50 FWE-small volume corrected (SVC) after initial thresholding at

305   p<.001, see Methods **Figure 4a,b**). The cluster positively tracked prediction errors of

306   pain avoidance when learning for other (t(33)= 2.30, p=.028) and negatively tracked

307   prediction errors of pain avoidance when learning for self (t(33)= -2.89, p=.007).

308   Although this cluster extended into the caudate, the caudate ROI itself was not

309   significant (x =18, y = 6, z =0, Z=3.49, k=2, p=.064, FWE-SVC after initial thresholding

310   at p<.001).

311
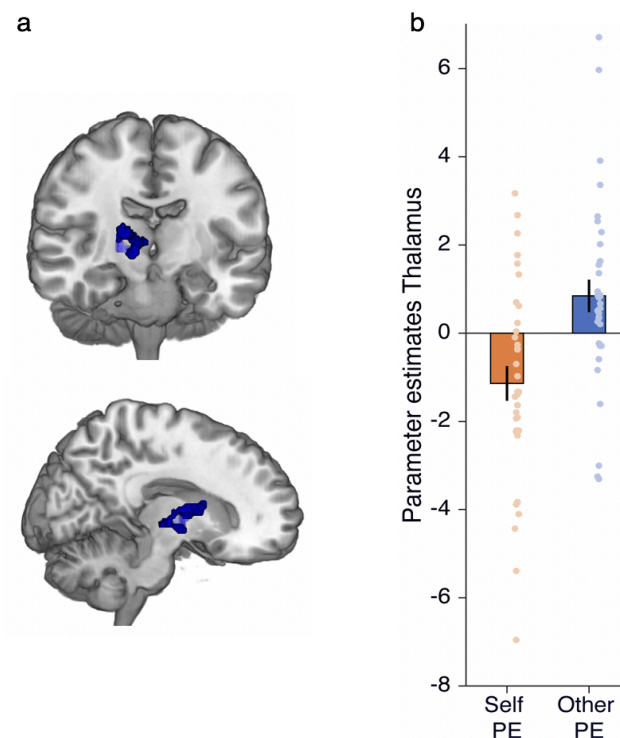
Figure 4 – Thalamus/caudate signal distinguishes model-free prediction errors for avoiding harm to other vs. self. (a) Thalamus cluster from the contrast other prediction error > self prediction error (x =16, y = -18, z =0, Z=3.50, k=84, p=.033, FWE-SVC after initial thresholding at p<.001) overlaid on an anatomical scan to show the extent of activation. (b) for illustration, parameter estimates extracted from the thalamus cluster are shown separately for self and other PE.

Signatures of model-free influence are encoded in sgACC and TPJ

321   One signature of model-free learning is a tendency to repeat previously rewarded

322   actions and avoid previously punished actions, regardless of experienced

323   transitions[20]. Such a model-free influence is thought to emerge at the time of choice

324   by activating the reinforcement histories of potential actions and driving selection of

325   the most valuable action in terms of its recent history[70]. In the context of our task,

326   through model-free influence, an action that was unpunished on the previous trial

327   should be prioritized for selection ('stay'), while an action that was punished on the

328   previous trial should be avoided ('switch'). Importantly, because model-free learning

329   is insensitive to task structure, this process should occur regardless of whether the

330   transition from the first to the second stage experienced on the previous trial was

331   common or rare.

332

333   Therefore, to probe the neural signatures consistent with a model-free influence at the

334   time of choice, and any potential differences between self and other conditions, we

335   examined neural responses during 'switch' and 'stay' choices at the first stage as a

336   function of the outcome on the previous trial (no pain or pain). We created an additional

337   GLM (GLM2) that modelled the onset of self trials after pain, self trials after no pain,

338   other trials after pain, and other trials after no pain, with stay (-1) and switch (1) coded

339   as parametric modulators of each of these onsets. Thus, our analysis examined

340   differential neural encoding of stay vs. switch decisions on the current trial, as a

341   function of the outcome on the previous trial (pain or no pain) and its recipient (self or

342   other).

343

344   Our analysis revealed a signal in sgACC consistent with a model-free influence on

345   'other' trials (x=-2, y=36, z=6, K=498; Z = 3.88, p=.028, FWE whole-brain corrected).

346   This region was more active during stay relative to switch choices on the current trial,

347   following a 'no pain' outcome on the previous trial, selectively in the 'other' condition

348   (see **Figure 5a-b** and Supplementary Analysis for more details). Corroborating the

349   view that this signal reflects a model-free influence, responses in sgACC were

350   positively associated with the model-free x recipient interaction in behaviour (r(33) =

351   .36, p=.04, **Figure 6a,b**) such that participants with the largest sgACC difference

352   between stay and switch following no pain for other also showed the strongest

353   prioritization of model-free learning for others relative to self.

354

355   We observed the inverse pattern (i.e., opposite to what would be predicted by model-

356   free influence) in right TPJ (x=54, y=-38, z=34, Z=3.56, K=39, p=.03, FWE-SVC). This

357   region was more active during switch relative to stay choices on the current trial,

358   following a 'no pain' outcome on the previous trial, in the 'other' condition specifically

359   (**Figure 5c-d** and Supplementary Analysis).

360

361

362

363
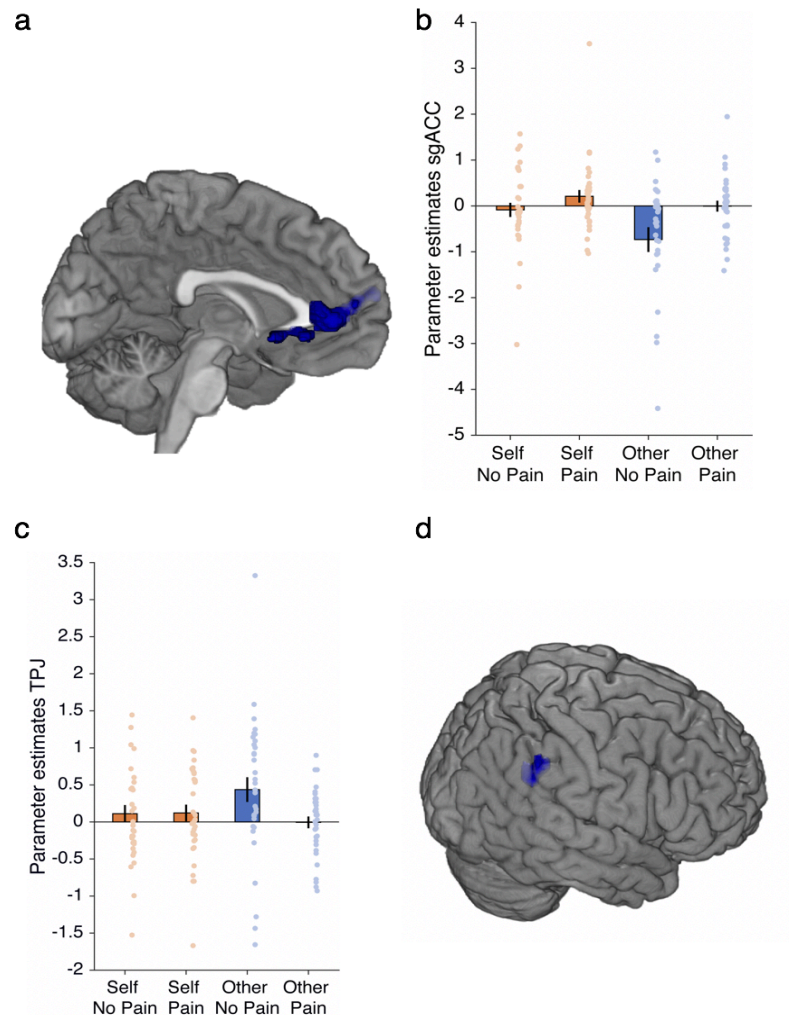
364   **Figure 5 - Subgenual anterior cingulate cortex (sgACC) and temporo-parietal junction (TPJ)**
365   **differentially encode signatures of model-free influence.** (a,b) sgACC response (x=-2, y=36, z=6,
366   K=498, p=.028, FWE whole brain corrected, after initial thresholding at p<.001) to stay vs. switch after
367   no pain for other overlaid on the medial surface of an anatomical scan. The observed BOLD pattern is
368   consistent with model-free behaviour. (c,d) Temporoparietal junction (TPJ) (x=54, y=-38, z=34, Z=3.56,
369   p=.03, FWE-SVC after initial thresholding at p<.001) tracks switch more than stay after no pain for other.
370   Cluster overlaid on an anatomical scan**.**

371

372   **Increased functional connectivity between sgACC and dorsolateral prefrontal**

373   **cortex when resisting model-free influence**

374

375   Behavioral analyses indicated that participants on average showed a mixture of

376   model-based and model-free strategies, but prioritized model-free learning when

377   avoiding harm to others. Thus, we next sought to identify regions that might modulate

378   the model-free effects observed specifically in the 'other' condition in sgACC (**Figure**

379   **6a**). We therefore conducted psycho-physiological interaction (PPI) analyses (GLM3)

380   to assess functional connectivity between sgACC and the whole brain as a

381   consequence of staying versus switching after no pain for other (see methods for

382   additional details of analyses). This analysis identified a significant negative

383   association between activity in sgACC and dorsolateral prefrontal cortex (x=-46, y=38,

384   z=26, k=382, Z=4.12, *p*=.039, FWE-whole brain corrected after initial thresholding at

385   *p*<.001).

386

387   To understand the nature of this effect we plotted the average slope of sgACC and

388   dlPFC connectivity during stay and switch conditions (**Figure 6b,c**). This showed that

389   there was a stronger positive coupling between sgACC and dlPFC during switch

390   choices compared to stay choices after receiving no pain for another person (**Figure**

391   **6b-c**). We did not observe significant coupling between sgACC and dlPFC during

392   switch or stay choices following no pain for self (Supplemental Analyses).

393
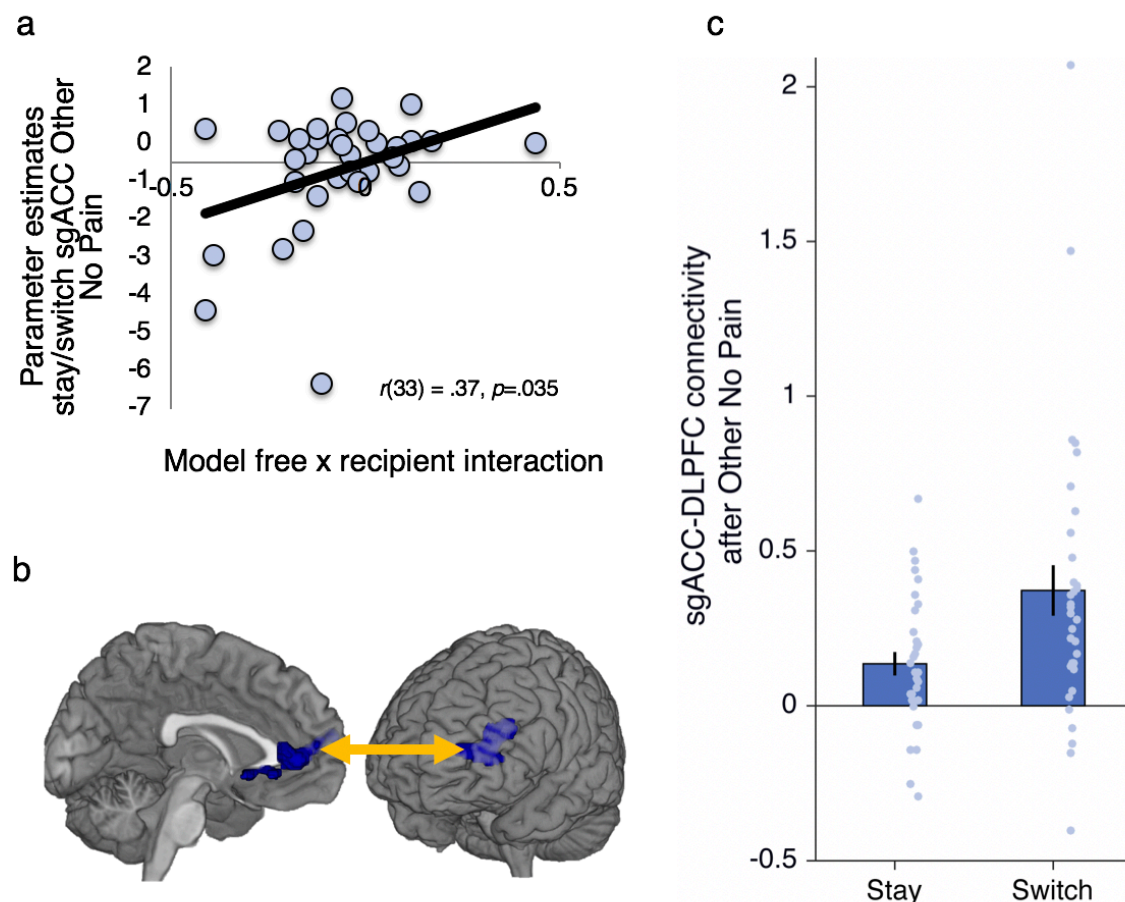
Figure 6 – Subgenual anterior cingulate cortex (sgACC) tracks stay vs. switch after no pain for other and connects more strongly to dlPFC when switching after no pain for other. (a) Bivariate association between parameter estimates for stay vs. switch after no pain for other in sgACC and greater model-free behaviour for other (more negative on x means relatively more model-free for other compared to self). (b) sgACC response (x=-2, y=36, z=6, K=498, p=.028, FWE whole-brain corrected after initial thresholding at p<.001) to stay vs. switch after no pain for other overlaid on the medial surface of an anatomical scan. The observed BOLD pattern is consistent with model-free behaviour. sgACC cluster connects to dorsolateral prefrontal cortex (dlPFC) (x=-46, y=38, z=26, k=382, Z=4.12, p=.039, FWE whole-brain corrected after initial thresholding at p<.001) during decisions to switch relative to stay after no pain for other. (c) Average slope estimate across participants shows stronger connectivity during switch decisions than stay decisions after receiving no pain for other between sgACC and dlPFC. Switching after no pain goes against a model-free influence.

Individual differences in moral judgment relate to model-free moral learning

Finally, we investigated whether individual differences in moral judgment were related to individual differences in model-free moral learning and its neural basis. We examined two aspects of moral judgment. First, motivated by theories suggesting that 'anti-utilitarian' judgments in moral dilemmas might be driven by model-free processes[3,27], we measured individual differences in utilitarian reasoning using the

415   Oxford Utilitarianism Scale[71]. Because our study concerned harmful outcomes, we

416   predicted that we would observe correlations with the 'instrumental harm' component

417   of the scale, which measures a permissive attitude towards harming one person in

418   order to help many others, but not the 'impartial beneficence' component of the scale,

419   which measures an impartial attitude towards helping others. First, we tested whether

420   utilitarianism was correlated with model-free behavior for others (relative to self). We

421   found a significant relationship between instrumental harm and model-free moral

422   behaviour, such that those who were the most anti-utilitarian were the most model-

423   free ($r(36) = 0.37$, $p = 0.026$). However, there was no correlation with the impartial

424   beneficence subscale ($r(36) = -.059$, $p = 0.731$). Next, given past work linking

425   utilitarian reasoning with dLPFC function[72,73], we predicted that utilitarianism would

426   positively predict dLPFC-sgACC connectivity when resisting model-free influence.

427   This prediction was supported, with dLPFC-sgACC connectivity positively correlated

428   with instrumental harm ($r(33) = 0.43$, $p = 0.012$), but not impartial beneficence ($r(33) =$

429    $-.11$, $p = 0.537$).

430

431   Second, we probed the sensitivity of moral wrongness judgments to how much

432   suffering an action inflicts on a victim ('outcome sensitivity'), versus how aversive it

433   feels to perform the action ('action sensitivity'[74]). Past work has connected the former

434   with model-based learning and the latter with model-free learning[3,27]. However, we

435   note that model-free learning is directly sensitive to recent outcomes[18], which might

436   lead to an association between model-free behavior and outcome sensitivity.

437   Participants evaluated the moral wrongness of 23 harmful actions that varied

438   independently in how much suffering they would cause, versus how aversive they

439   would feel to perform. Action sensitivity and outcome sensitivity were inversely

440   correlated ($r(36) = -.40$, $p = .016$). Partial correlations controlling for action sensitivity

441   revealed that outcome sensitivity was positively correlated with several aspects of

442   model-free moral learning (**Figure 7**), including the tendency to switch following harm

443   to others ($r(36) = -.37$, $p=.027$), the strength of model-free prediction error signals for

444   other vs. self in thalamus/caudate ($r(34) = .385$, $p = .025$), and the strength of model-

445   free influence in sgACC ($r(33) = -.374$, $p = .032$). The reverse partial correlations

446   testing for the effect of action sensitivity whilst controlling for outcome sensitivity were

447   not significant (all p's>.08, all |r's| <-.30). Together these findings suggest that a natural

448   tendency to engage in model-free moral learning when avoiding harm to others is

449   related to how moral judgments of others' harmful actions track with harm severity.

450

451
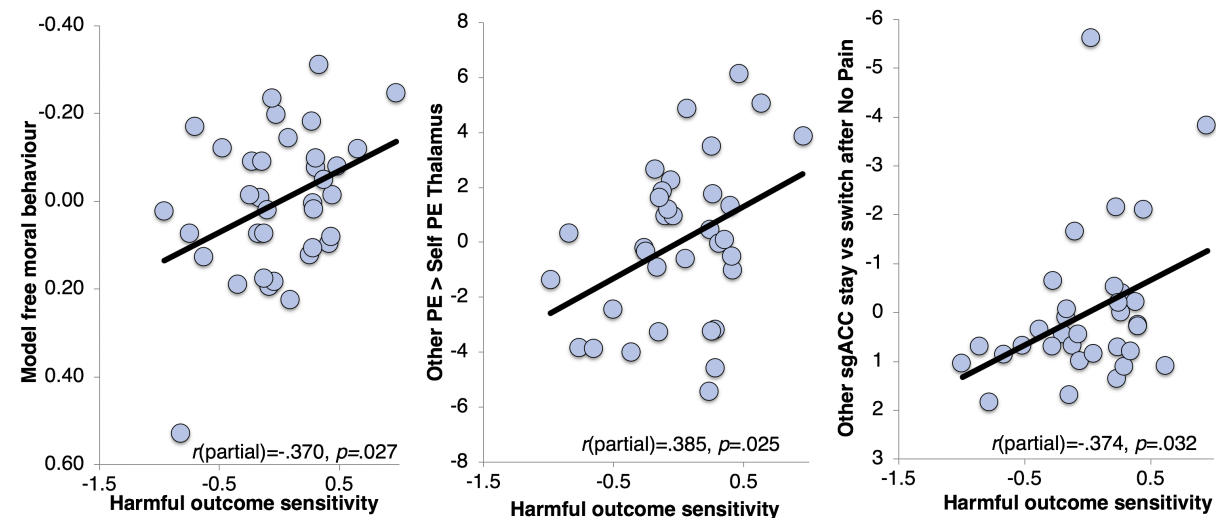
452   **Figure 7 – Sensitivity to harm in moral judgments correlates with model-free moral behaviour**
453   **and its neural correlates.** (a) Partial correlation controlling for harmful action sensitivity between
454   harmful outcome sensitivity and model-free moral behaviour. Note the values are reversed on the y axis
455   to depict that greater model-free behaviour is associated with greater harmful outcome sensitivity (b)
456   Partial correlation controlling for harmful action sensitivity, between harmful outcome sensitivity and
457   prediction errors of pain avoidance in the thalamus/caudate for Other compared to Self (c) Partial
458   correlation controlling for harmful action sensitivity between harmful outcome sensitivity and parameter
459   estimates extracted from the parametric regressor for stay vs switch after no pain for other in subgenual
460   anterior cingulate cortex (sgACC). Note that values are reversed on the y axis such that a greater
461   tendency to stay vs switch tracked in sgACC correlates with harmful outcome sensitivity.

462

463

Discussion

465

466   Learning to avoid actions that harm other people is a fundamental prerequisite for

467   moral behaviour. Here we show that people prioritize model-free learning when

468   actions have the potential to harm others, and that learning to avoid harming others

469   (vs. self) has a distinct neural signature. The thalamus/caudate differentially encoded

470   prediction errors of pain avoidance for self versus other, whilst sgACC and right TPJ

471   tracked positively and negatively with model-free influence on pain avoidance at the

472   time of choice. Overriding model-free influence when choices affected others invoked

473   stronger connectivity between sgACC and dlPFC. Finally, multiple aspects of moral

474   judgment were associated with model-free moral learning and its neural correlates.

475

476   In the context of our study, model-free moral learning manifested as a reduced

477   likelihood of repeating actions that harmed others on the previous trial, regardless of

478   whether such actions typically led to states with a high likelihood of harmful outcomes.

479   Our behavioural finding that people were more model-free when learning to avoid

480   harming others relative to themselves suggests that potentially harmful actions might

481   be prioritized for automatic avoidance as a default. Given the importance of avoiding

482   harm to others for social life, such a learning mechanism would be socially adaptive.

483   Our findings are consistent with prior work showing that harmful social choices are

484   slower than helpful choices[5,51,52] as well as evidence that morality constrains mental

485   representations of what actions are considered possible, with harmful actions removed

486   from choice sets as a default stance[50]. Repeatedly assigning negative action values

487   to harmful actions over the course of one's life might automatically remove such

488   actions from consideration, even though in some cases locally harmful actions can

489   lead to wider benefits (e.g., a surgeon cutting open a patient to remove a cancerous

490   tumor). If such a learning strategy is socially adaptive, this raises interesting questions

491   about whether model-free learning can be considered an "optimal" strategy in the

492   ecological sense.

493

494   An alternative explanation for our behavioral findings is that model-based learning is

495   effortful[26], and people choose to put in less effort to benefit others[75]. However, this

496   explanation seems unlikely given that model-free moral learning in our study was

497   specifically driven by a lower probability of repeating choices that harmed others. In

498   other words, participants were more likely to actively switch their choices following

499   harm to others, rather than passively sticking with the status quo. Our results were

500   also not due to differences in the subjective perception of the harmfulness or

501   aversiveness of the outcomes for others compared to self, as participants overall rated

502   shocks received for others in the task as being just as aversive as shocks received for

503   themselves. Furthermore, we did not observe differences in choice consistency

504   (captured by the temperature parameter in our model) during learning for self vs.

505   others. Since choice consistency is related to task engagement, if our behavioral

506   effects reflected reduced effort for others, we would expect to see lower choice

507 consistency when learning for others than self. Together these alternative
508 considerations suggest that people might naturally prioritize model-free
509 strategies/behaviours when avoiding harm to others, and these effects may not simply
510 be explained by less effort or engagement when actions affect others relative to
511 oneself.

512

513 Turning to the neural findings, we observed a signal in the thalamus, extending to the
514 caudate, that differentially encoded model-free prediction errors when learning to
515 avoid harming others versus self. These subcortical regions were previously observed
516 to encode value during moral decisions to avoid profiting from others' pain[4] and play
517 a critical role in associative learning and moral-decision-making more broadly[40,42,44,45].
518 The thalamus is often linked to the processing of the affective dimension of pain in
519 addition to its sensory properties[76]. For example, microstimulation of the thalamus can
520 invoke affective memories of previously experienced pain[45]. The thalamus/caudate
521 signal differed from the adjacent ventral striatum response that positively tracked
522 model-free prediction errors regardless of the recipient of the outcome, consistent with
523 a previous study using a similar task with rewarding outcomes for self only[20]. These
524 findings suggest that multiple sub-cortical areas support model-free moral learning,
525 perhaps with ventral striatum providing a generic model-free prediction error signal
526 that is insensitive to outcome valence and outcome recipient, and thalamus/caudate
527 providing additional information about social context.

528

529 Another signature of model-free influence was observed in sgACC at the time of
530 choice, contingent on the outcome of the previous trial and specific to the 'other'
531 condition. Specifically, signal in sgACC was higher when participants repeated actions
532 that previously avoided harming others, but not during similar choices for oneself. This
533 pattern is consistent with model-free behavior, and individual differences in model-free
534 behavior tracked with individual differences in sgACC response at the time of choice.
535 Notably, previous work has implicated sgACC in model-free learning to gain rewards
536 for others but not self[53] and in receiving unexpected positive feedback from others[77],
537 suggesting this region might compute learning signals that are specific to social

538 settings. More broadly, activity in sgACC has been positively associated with prosocial
539 and moral behaviours[15,54,55,57].

540

541 Collectively these findings suggest that sgACC might bias decision-making away from
542 choices that could harm others. However, this default strategy is not always
543 appropriate, for instance in settings where a typically harmful action might lead to a
544 better outcome. We observed signals at the time of choice in two areas that negatively
545 tracked with model-free influence. The first, in right TPJ, showed a response pattern
546 precisely opposite to that observed in sgACC. Signal in rTPJ was higher when
547 participants abandoned a choice that previously avoided pain for others, but not during
548 similar choices for oneself. In addition, overriding model-free influence at the time of
549 choice was associated with increased functional connectivity between sgACC and
550 dLPFC. These two regions showed stronger coupling on trials where participants
551 abandoned a choice that previously avoided pain for others, compared with trials
552 where participants repeated actions that previously spared others from pain. Although
553 we cannot confidently attribute these patterns to model-based control, past work has
554 implicated dLPFC and TPJ in model-based learning and decision-making[31,70]. One
555 intriguing possibility is that sgACC promotes social harm avoidance as a default, while
556 TPJ and dLPFC provide contextual information that enables this default to be
557 overridden when appropriate. Research implicating these regions in the adjustment of
558 moral decisions to blame and punishment provides initial support for this
559 possibility[5,32,33,39,47].

560

561 Finally, we observed correspondences between model-free learning, its neural
562 substrates, and moral judgments. Theoretical work has proposed links between
563 model-based/model-free learning and moral judgment[3,27,28,78], but empirical support
564 for such links has been scarce. We probed two aspects of moral judgment. First, we
565 examined individual differences in the dimension of 'utilitarian' moral reasoning that
566 justifies harming one person to help many others. Consistent with our predictions, as
567 well as work highlighting a link between dLPFC activity and 'utilitarian' judgments[73],
568 we found a positive relationship between instrumental harm and dLPFC-sgACC

569   connectivity when resisting model-free influence. We also observed that those who

570   were the most anti-utilitarian were the most model-free.

571

572   Second, using a task that asks participants to judge how morally wrong it would be to

573   perform a series of violent actions that varied independently in terms of how much

574   suffering they would inflict ('outcome sensitivity'), versus how aversive they felt to

575   perform ('action sensitivity')[74]. We found that individual variability in both the

576   behavioural and neural signatures of model-free learning was specifically correlated

577   with outcome sensitivity, but not action sensitivity. Those people whose moral

578   wrongness judgments were more sensitive to the severity of harmful outcomes were

579   less likely to repeat decisions that harmed others, and showed stronger model-free

580   prediction error signals in the thalamus and caudate, and stronger responses in

581   sgACC when repeating decisions that previously avoided harming others. Model-free

582   learning has previously been suggested to explain why actions that typically harm

583   others feel aversive to perform, even when they are not actually harmful[3,27]. Thus, one

584   might expect that a greater tendency to engage in model-free moral learning should

585   predict action sensitivity in moral judgments, rather than outcome sensitivity. However,

586   model-free learning is directly sensitive to recent outcomes[18], and in the context of our

587   task, manifested as a tendency for choices to be immediately sensitive to harmful

588   outcomes for others. Thus, individual differences in sensitivity to others' harm could

589   be commonly associated with model-free moral learning and outcome sensitivity in

590   moral judgments. Overall, these findings provide preliminary evidence linking model-

591   free learning to individual differences in moral judgments, effects that could be

592   investigated more extensively in larger samples.

593

594   More broadly, our findings highlight differences in the neurocognitive mechanisms

595   engaged in learning to avoid harming oneself versus others that underscore the unique

596   demands of social decision-making. One important feature of decisions that affect

597   others (as opposed to oneself) is that it is far more difficult to build a model that

598   incorporates others' preferences and beliefs than a model that captures only one's

599   own preferences. Furthermore, it is impossible to evaluate the accuracy of such

600   models, given that the subjective experiences of others are fundamentally

601 unknowable[79,80]. Utilitarian approaches to moral decision-making that involve

602 maximizing well-being for all sentient beings[81] may thus be computationally intractable

603 for the model-based system. Rule-based approaches, like those enshrined in

604 deontological theories of morality[82] circumvent the need for complex model building

605 and may be socially adaptive even for simple social decisions like the ones studied

606 here. Whether the prioritization of model-free learning extends to other kinds of social

607 decisions, such as acting to obtain rewards for others, avoiding monetary losses or

608 indeed even social decision-making in nonhuman species, is an important topic for

609 future study.

610

611 Overall, we observed that when learning to avoid harm to others (versus self),

612 participants showed a stronger relative balance toward model-free over model-based

613 learning. Multiple model-free learning signatures were apparent in behaviour as well

614 as cortical and subcortical areas that distinctly process harm avoidance for others

615 compared to self. These findings could have important implications for theories of

616 learning and moral-decision-making as well as disorders associated with impaired

617 avoidance learning and social cognition.

618

619 **Methods**

620

621 **Experimental procedures**

622 **Participants**

623 Forty-one right-handed healthy adults were recruited through university participant

624 databases. Exclusion criteria included previous or current neurological or psychiatric

625 disorder, non-normal or non-corrected to normal vision, previous participation in

626 studies involving social interactions and/or electric shocks and contraindications that

627 prohibited MRI scanning. One participant was excluded as they reported that they did

628 not believe their decisions would affect another person in the post-scanning debrief.

629 Three participants were excluded because the logistic regression analysis of their

630 choice behaviour did not converge. In two of these participants, this was because they

631 had less than 5% switch trials; in the third participant, there was a very high correlation

632   (>0.8) between the outcome and transition regressors on switch trials. One participant
633   was excluded from the fMRI analysis due to distortions in the scan caused by metal
634   artefacts (braces) and another for excessive head motion. This left a final sample of
635   36 participants for behavioural analyses (16 female, age 18-36) and 34 participants
636   (16 female, age 18-36) for the parametric fMRI analyses. For the stay/switch analysis
637   one further participant was excluded for having no variance in at least one regressor
638   making their fMRI GLM inestimable. With 33 subjects we had 80% power to detect a
639   'medium' effect size of $d = 0.50$ at alpha = 0.05 (two-tailed), an effect size smaller than
640   typically reported in this field, indicating sufficient power. All participants gave written
641   informed consent and the study was approved by the University of Oxford Medical
642   Sciences Division Ethics Committee.
643
644   **Procedure**
645
646   Participants completed a pain thresholding procedure which was based on previous
647   studies of self and other pain processing[4,5]. The pain thresholding procedure allowed
648   us to control for heterogeneity of skin resistance between participants to ensure the
649   delivered shocks would be rated at a matched subjective level of pain intensity and
650   also to provide participants with full experience of the shocks before the learning task
651   to ensure their choices were truly guided by knowledge of the pain and no pain
652   outcomes. Participants were then assigned to roles of either 'decider' or 'receiver'
653   using a role assignment procedure that has been used in several previous studies
654   (see[5,75]). Briefly, participants were instructed to wear coloured rubber gloves to hide
655   their identity. They then stood either side of a door and waved to one another so that
656   they knew another person was there but could not discern any information about the
657   other participant's age or gender. Next a coin was flipped to decide who would draw a
658   ball out of a box first and then each participant drew a ball. The experimental
659   participant was then told that the colour of their ball meant they had been assigned to
660   the role of decider. Before completing the task in the scanner participants performed
661   a practice task of one block that did not specify whether outcomes were for self or
662   other and they were told no actual shocks would be received during the practice. This
663   practice task, which used stimuli not used in the experimental task, allowed

664   participants to become familiar with the transition structure of the task, that they were

665   told would remain the same in the main experiment. The main experiment consisted

666   of 4 blocks of 68 trials (136 trials for self and 136 trials for other) and lasted ~45

667   minutes.

668   **Experimental task**

669   We adapted features from two variants of a task designed to distinguish model-free

670   and model-based learning[20,26] (**Figure 1, Supplemental Experimental note**).

671   Participants were presented with two fractal images that probabilistically led them to

672   one of two 'states' where they were required to make a second choice that was

673   followed by a symbol indicating the receipt of pain or the receipt of no pain (neutral).

674   At the beginning of each block and on each trial, they were told whether they were

675   playing for themselves ('You') meaning the painful outcomes would be delivered to

676   themselves, or for the other participant ('Receiver'), indicating that the painful

677   outcomes would be delivered to the other participant. In order to match the self and

678   other conditions in terms of pain stimulation, no electric shocks were delivered during

679   the scan. However, participants were told that 10% of the electric shocks that they

680   acquired during the task would be given to themselves at the end of the session and

681   10% of the electric shocks they accumulated for the other participant would be

682   delivered to the other participant at the end of the scanning session. In order to

683   account for potential differences in pain perception, participants were instructed that

684   for 'You' trials we would use the voltage setting that corresponded to *their* level 8

685   rating, and for the "Receiver" trials, we would use the voltage setting that corresponded

686   to *the Receiver's* level 8 rating (full instructions can be downloaded at (OSF

687   https://osf.io/3stp9/files/). At the end of the scan participants also rated how they felt

688   when obtaining a shock for themselves and the receiver during that task on a 0-10

689   point scale from 'very negative' (0) to 'very positive' (10). Participants perceived the

690   shocks to be aversive for both self (M=3.38, SD = 1.44) and other (M=3.30, SD = 1.61)

691   ($t(36)$ = .279, p=.782) suggesting that the perception of harm was equivalent.

692

693   Participants were instructed that the probability of reaching each of the two 'states'

694   would remain fixed throughout the task but the probability that each stimulus would

695   deliver pain or no pain would change throughout the task so that they needed to keep

696   on learning. Adding the 'receiver' trials meant that fewer trials, overall, were included

697   in each condition compared to the original paradigm[20]. We therefore made two

698   modifications to the original paradigm in order to better sample variability in behaviour.

699   We bounded the probability that the four second-stage options would deliver pain

700   between 0 and 1 (instead of 0.25 and 0.75 in[20]) and we increased the drift rate of the

701   probabilities to 0.2 (instead of 0.025 in[20]). Moreover, we specifically chose to

702   implement the main aspects off the Daw et al.[20], version of the paradigm rather than

703   using all the modifications described in[26], as the Daw et al.[20], paradigm allowed us to

704   assess peoples' relative balance between model-free or model-based behaviour[20,26]

705   when there was no particular incentive to being more model-free or model-based on

706   the task because the degree of model-basedness did not affect the number of shocks

707   received.

708   **Statistical analysis of behavioral data**

709   Analyses of behavioural data were performed in SPSS 25 (Armonk, New York: IBM

710   Corp, for bivariate correlations), R (version 1.1.423, for linear mixed-effects modelling,

711   lme4, version 1.1-21) and Matlab 2015b (for logistic regression analyses on the

712   probability of stay). For Fig 2a, we calculated the % of stay choices after common or

713   rare transitions following pain or no pain outcomes (2x2). For regression analyses

714   using lme4 in R, we coded Stay as 1 and Switch as 0 and created regressors for

715   transition type, outcome, outcome x transition type, agent x outcome and agent x

716   outcome x transition type. We used Bound Optimization by Quadratic Approximation

717   (bobyqa) with 1e5 function evaluations. We examined bivariate associations between

718   the interactions with agent from the regression analyses and neural responses with

719   individual differences in utilitarianism using the Oxford Utilitarianism Scale (OUS)[71]

720   and action and outcome sensitivity from the Harmful Action Outcome Scale[74].

721

722   **Moral judgment measures**

723   Participants completed the OUS instrumental harm (OUS-IH), OUS Impartial

724   Beneficence (OUS-IB) scales and the HAO scales via an online link in the preceding

725   weeks before the scanning session. The OUS-IH consists of 4 items reflecting a

726 relative willingness to cause harm to others in order to bring about the greater good
727 (e.g., "It is morally right to harm an innocent person if harming them is a necessary
728 means to helping several other innocent people"). The OUS-IB subscale consists of 5
729 items reflecting endorsement of the impartial maximization of the greater good even
730 at a cost to oneself (e.g. if the only way to save another person's life during an
731 emergency is to sacrifice one's own leg, then one is morally required to make this
732 sacrifice). Participants rated these items on a 7-point scale (*1 = strongly disagree, 7 =*
733 *strongly agree).* A mean score was then computed for all participants.

734

735 For the HAO participants were presented with a scenario about two people, Carl and
736 John. They are told John is terminally ill and sincerely wants to die and has asked
737 another person, Carl, to perform a mercy killing. Participants then rated how morally
738 wrong each of the 23 methods of killing were on a scale from 1 – the least morally
739 wrong, to 10 – the most morally wrong. In a previous study[74] participants were asked
740 to rate the action value and outcome value of these different methods of killing. To
741 assess action value, the researchers asked participants to rate how *upsetting* it would
742 be to "act out" performing each behavior as though it were part of a movie script. To
743 assess outcome value, they asked them to rate how much *suffering* each act would
744 impose. We used the mean action and outcome scores derived from this initial paper
745 to predict the 'wrongness' scores in our current sample. This analysis created two
746 different beta weights for each participant corresponding to the action and outcome
747 sensitivity, respectively.

748

749 **Computational modelling of behavioural data**

750

751 For modelling of choice behaviour using trial-by-trial updates, we proceeded in two
752 steps. First, we evaluated a number of models separately based on their performance
753 on the self and other blocks. This was to probe whether the same model would win for
754 self and other blocks, allowing us to rule out participants might employ entirely different
755 strategies in the two block types. The following models were fitted:

756

757   (1) 7-parameter: full model specified by Daw et al. using parameters: learning rates

758   for stage 1 and 2 (αS1, αS2), temperature parameters for stage 1 and 2 (βS1,βS2), a

759   perseverance parameter (ρ), an eligibility trace (λ) and a model-free/based weighting

760   parameter (ω); for full details of model see Supplementary information and[20]

761   (2) 6-parameter model, as (1) but with λ=1 (λ was shown to have a high mean value

762   and small variance in previous work e.g.[62])

763   (3) 5-parameter model, as (1) but with only one α and β for stage 1 and 2

764   (4) 4-parameter model, as (3) but with λ=1

765   (5) 5-parameter model, as (4) but with two learning rates for pain and no pain

766   outcomes (αPain, αNoPain)

767

768   Models were fitted using a hierarchical Bayesian model fitting approach described in

769   detail in[83,84]. It finds the maximum *a posteriori* estimate of each parameter for each

770   subject using a prior distribution for each parameter which helps to regularise and

771   constrain parameters. The algorithm uses Expectation-Maximization (EM)[85] and

772   parameters were transformed to a logistic or exponential distribution to enforce

773   constraints and ensure normality such that $0<\{\alpha,\omega\}<1, \{\beta,\lambda\}>0$.

774

775   For formal model comparison, we report the Bayesian Information Criterion (BIC)

776   based on the log-likelihood, and computed the model evidence by integrating out the

777   free parameters (BICint[83,84]; **Supplementary Tables 2 and 4**). Exceedance

778   probabilities were calculated by feeding the BICint into SPM's function spm_BMS

779   (http://www.fil.ion.ucl.ac.uk/spm/software/spm8).

780

781   The five-parameter model with separate learning rates for pain and no-pain outcomes

782   best explained behaviour in both self and other conditions (**Supplementary Table 2**).

783   We report the difference between the best-fitting parameters, but this method has a

784   caveat. Because of the nature of hierarchical fitting, which uses separate priors for self

785   and other parameters, this method is somewhat biased towards finding differences.

786   Meanwhile, fitting self and other parameters using the same priors, is overly

787   conservative and biased against finding differences.

788

789   To resolve whether the parameter ω differed between self and other blocks, in line

790   with results from the basic logistic regression analyses, and without introducing any

791   such biases, in a second step, we therefore fitted three models to the merged data of

792   both self and other blocks:

    (1) 5-parameter model (as (5) above) with all parameters shared between
793

794   self/other

    (2) 6-parameter model with αPain, αNoPain, β and ρ shared but ω split into ωSelf
795

796   and ωOther

    (3) 7-parameter model with αPain, αNoPain, β shared and ρ and ω split into ρSelf,
797

798   ρOther, ωSelf and ωOther

799   As described above, model comparison was performed based on BICint values

800   (**Supplementary Table 4**). The mean parameter estimates are shown in

801   **Supplementary Fig 3** and **Supplementary Table 5**. We also simulated data from our

802   participant schedules and showed that we had reliable parameter recovery[86]

803

804   **FMRI acquisition and analysis**

805   Multiband T2*-weighted echo planar imaging (EPI) volumes with blood oxygenation-

806   level–dependent (BOLD) contrast were acquired using a Siemens Prisma 3T MRI

807   scanner. The EPI volumes were acquired in an ascending manner, at an oblique angle

808   ($\approx$30˚) to the AC-PC line to decrease the impact of susceptibility artefacts in the

809   orbitofrontal cortex. We used the following parameters. Voxel size 2 x 2 x 2, TE=30

810   ms; repetition time=1570ms; flip angle=90°; field of view=216 mm. The structural scan

811   was acquired using a magnetization prepared rapid gradient echo (MPRAGE)

812   sequence with 192 slices; slice thickness=1 mm; TR=1900 ms; TE=3.97 ms; field of

813   view=192 mm x 192mm; voxel size=1×1×1 mm resolution.

814

815   Fmri data were analysed using SPM12 (www.fil.ion.ucl.ac.uk/spm). Images were

816   realigned and unwarped using a fieldmap and co-registered to the participant's own

817   anatomical image. The anatomical image was processed using a unified segmentation

818   procedure combining segmentation, bias correction, and spatial normalization to the

819   MNI template using the New Segment procedure; the same normalization parameters

820   were then used to normalize the EPI images. Lastly, a Gaussian kernel of 8 mm
821   FWHM (SPM default) was applied to spatially smooth the images.

822

823   Before the study, example first-level design matrices were checked to ensure that
824   estimable GLMs could be performed with independence between the parametric
825   regressors: value difference at the first-stage choice, the state prediction error at stage
826   2, and a model-free prediction error at the time of the outcome. This allowed us to look
827   at value and prediction error responses independent from one another. We also tested
828   a GLM that coded switch vs. stay trials as a parametric modulator at the time of choice
829   dependent on the previous outcome. Again this GLM could be estimated with
830   independence (See Supplementary Figure 1). We convolved these different event
831   types with SPM's canonical haemodynamic response function. All events were
832   modelled as stick functions with 0 duration.

833

834   For GLM1, each of these regressors was associated with parametric modulators
835   taken from the computational model. At the time of the first stage choice this was the
836   value difference from the hybrid model combining model-free and model based
837   learning. At the time of the second stage choice this was the state prediction error for
838   the transition from stage 1 to stage 2, and at the time of the outcome this was the
839   model free prediction error (since the behavioural differences were in the model-free
840   parameters). In all cases, values were modelled separately for the onsets of self and
841   other trials. As in[20], we fixed the parameters to the average values for self and other
842   (**Supplementary Table 4**) but allowed ω to vary.

843

844   For GLM2 we modelled whether participants stayed or switched at the first-stage
845   choice relative to the outcome on the previous trial, i.e. no pain or pain. Due to the
846   smaller number of trials included in this analysis we coded stay and switch as a
847   parametric regressor with values of 1 assigned to switch and -1 assigned to stay. One
848   participant did not have a trial in at least one of these regressors and was therefore
849   excluded from the stay-switch analysis. For all GLMs in some participants, an extra
850   regressor modelled all missed trials, on which participants did not select one of the
851   first-stage choices.

852

853   GLM3 and GLM4 corresponded to our psychophysiological interaction analyses. We

854   defined a seed region in the sgACC using a 6mm sphere based on the peak co-

855   ordinates from our analyses (for completeness we also ran PPI analyses using seed

856   regions in thalamus and TPJ, see Supplementary Text). We then extracted the

857   physiological variable and the psychophysiological interaction terms for stay vs. switch

858   after no pain for other (GLM3) and stay vs. switch after no pain for self as a control

859   analysis (GLM4). These PPI terms were entered into the GLMs along with all previous

860   regressors that specified the events of our study as described above. In all GLMs, six

861   head motion parameters modelled the residual effects of head motion as covariates

862   of no interest. Data were high-pass filtered at 128 s to remove low-frequency drifts,

863   and the statistical model included an AR(1) autoregressive function to account for

864   autocorrelations intrinsic to the fMRI time-series.

865

866   Contrast images from the first level were input into flexible-factorial designs. Following

867   standard procedures, main effects are reported at $p < .05$, family-wise error (FWE)

868   cluster corrected across the whole brain after initial thresholding at $p<.001$, or $p < 0.05$

869   FWE small volume corrected (SVC) after initial thresholding at $p<.001$ for regions

870   where we had a strong *a priori* hypothesis[87,88]. These areas were defined anatomically

871   and included subcortical regions of the caudate and thalamus (taken from the WFU

872   PickAtlas Toolbox), bilateral posterior TPJ (from[89]) the sgACC (areas s24 and 25

873   from[90] and the dlPFC (areas 46v and 9 taken from[91]). These ROIs were also used to

874   confirm anatomical labelling. We additionally applied a false discovery rate correction

875   (FDR) for the number of ROI corrections. All ROI comparisons remained significant

876   ($p< 0.05$) when controlling for the number of comparisons using FDR.

877

878   **Author contributions**

879

880   P.L.L, M. K-F and M. J. C designed study. P.L.L and A. A. collected data. P.L.L and

881   M. K-F analysed data. P.L.L, M. K-F, A. A., and M. J. C. wrote paper.

882

883   **Acknowledgements**

896

897 **Declaration of interests**

898

899 The authors have no competing interests

900

901 **Materials and correspondence**

902

903 Please address correspondence to Patricia L. Lockwood, Miriam C Klein-Flugge, and

904 Molly J Crockett.

905

906 **Data availability**

907

908 All data and code used to generate the figures can be downloaded at: [OSF

909 https://osf.io/3stp9/files/]

910

911 Unthresholded statistical maps can be downloaded at: [Identifier to be added upon

912 publication]

913

914

915 **References**

916

1. Gert, B. *Common Morality: Deciding What to Do*. (Oxford University Press, 2004).

2. Gray, K., Young, L. & Waytz, A. Mind Perception Is the Essence of Morality. *Psychol. Inq.* **23**, 101–124 (2012).

3. Crockett, M. J. Models of morality. *Trends Cogn. Sci.* 4–7 (2013). doi:10.1016/j.tics.2013.06.005

4. Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P. & Dolan, R. J. Moral transgressions corrupt neural representations of value. *Nat. Neurosci.* **20**, 879–885 (2017).

5. Crockett, M. J., Kurth-nelson, Z., Siegel, J. Z., Dayan, P. & Dolan, R. J. Harm to others outweighs harm to self in moral decision making. *Proc. Natl. Acad. Sci.* **112**, 201424572 (2015).

6. Cushman, F., Kumar, V. & Railton, P. Moral learning: Psychological and philosophical perspectives. *Moral Learn.* **167**, 1–10 (2017).

7. Decety, J. & Cowell, J. M. Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Dev. Psychopathol.* **30**, 153–164 (2018).

8. Cushman, F., Gray, K., Gaffey, A. & Mendes, W. B. Simulating murder: The aversion to harmful action. *Emotion* **12**, 2 (2012).

9. Blair, R. J. R. The neurobiology of psychopathic traits in youths. *Nat. Rev. Neurosci.* 786–799 (2013).

10. Kohlberg, L. Stages in the development of moral thought and action. (1969).

939   11. Tremblay, R. E. The development of agressive behaviour during childhood: What have we learned in the past century? *Int. J. Behav. Dev.* **24**, 129–141 (2000).

941   12. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).

943   13. Gino, F., Ayal, S. & Ariely, D. Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychol. Sci.* **20**, 393–398 (2009).

945   14. Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J. & Griskevicius, V. Normative social influence is underdetected. *Pers. Soc. Psychol. Bull.* **34**, 913–923 (2008).

948   15. Wittmann, M. K., Lockwood, P. L. & Rushworth, M. F. S. Neural Mechanisms of Social Cognition in Primates. *Annu. Rev. Neurosci.* (2018). doi:10.1146/annurev-neuro-080317-061450

951   16. Schultz, W. Updating dopamine reward signals. *Curr. Opin. Neurobiol.* **23**, 229–238 (2013).

953   17. Dayan, P. & Daw, N. D. Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* **8**, 429–453 (2008).

955   18. Sutton, R. S. & Barto, A. G. *Reinforcement learning: an introduction*. (MIT press, 1998).

957   19. O'doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, reward, and decision making. *Annu. Rev. Psychol.* **68**, 73–100 (2017).

959   20. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).

962   21. Glascher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus Rewards:

963   Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-

964   Free Reinforcement Learning. *Neuron* **66**, 585–595 (2010).

965   22. Lee, S. W., Shimojo, S. & O'Doherty, J. P. Neural computations underlying

966   arbitration between model-based and model-free learning. *Neuron* **81**, 687–699

967   (2014).

968   23. Wunderlich, K., Dayan, P. & Dolan, R. J. Mapping value based planning and

969   extensively trained choice in the human brain. *Nat. Neurosci.* **15**, 786 (2012).

970   24. Tricomi, E., Balleine, B. W. & O'Doherty, J. P. A specific role for posterior

971   dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* **29**, 2225–2232

972   (2009).

973   25. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between

974   prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*

975   **8**, 1704–11 (2005).

976   26. Kool, W., Cushman, F. A. & Gershman, S. J. When does model-based control

977   pay off? *PLoS Comput. Biol.* **12**, e1005090 (2016).

978   27. Cushman, F. Action, outcome, and value: A dual-system framework for morality.

979   *Personal. Soc. Psychol. Rev.* **17**, 273–292 (2013).

980   28. Gęsiarz, F. & Crockett, M. J. Goal-directed, habitual and Pavlovian prosocial

981   behavior. *Front. Behav. Neurosci.* **9**, 135 (2015).

982   29. Greene, J. D. The rat-a-gorical imperative: Moral intuition and the limits of

983   affective learning. *Moral Learn.* **167**, 66–77 (2017).

984   30. Ayars, A. Can model-free reinforcement learning explain deontological moral

985   judgments? *Cognition* **150**, 232–242 (2016).

986   31. Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D. & Dolan, R. J.

987   Disruption of dorsolateral prefrontal cortex decreases model-based in favor of

988   model-free control in humans. *Neuron* **80**, 914–919 (2013).

989   32. Carlson, R. W. & Crockett, M. J. The lateral prefrontal cortex and moral goal

990   pursuit. *Curr. Opin. Psychol.* (2018).

991   33. Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G. & Fehr, E. The neural

992   signature of social norm compliance. *Neuron* **56**, 185–196 (2007).

993   34. Greene, J. D. & Paxton, J. M. Patterns of neural activity associated with honest

994   and dishonest moral decisions. *Proc. Natl. Acad. Sci.* **106**, 12506–12511 (2009).

995   35. Buckholtz, J. W. & Marois, R. The roots of modern justice: cognitive and neural

996   foundations of social norms and their enforcement. *Nat. Neurosci.* **15**, 655

997   (2012).

998   36. Chang, L. J., Smith, A., Dufwenberg, M. & Sanfey, A. G. Triangulating the neural,

999   psychological, and economic bases of guilt aversion. *Neuron* **70**, 560–572

1000   (2011).

1001   37. Feldmanhall, O. *et al.* Differential Neural Circuitry and Self-Interest in Real versus

1002   Hypothetical Moral Decisions. *Soc. Cogn. Affect. Neurosci.* 1–23 (2012).

1003   doi:10.1093/scan/nss069

1004   38. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. Diminishing

1005   reciprocal fairness by disrupting the right prefrontal cortex. *science* **314**, 829–832

1006   (2006).

1007   39. Ruff, C. C., Ugazio, G. & Fehr, E. Changing social norm compliance with

1008   noninvasive brain stimulation. *Science* **342**, 482–484 (2013).

1009   40. Fumagalli, M. & Priori, A. Functional and clinical neuroanatomy of morality. *Brain*

1010   **135**, 2006–2021 (2012).

1011   41. Gaffan, D. & Murray, E. A. Amygdalar interaction with the mediodorsal nucleus of

1012   the thalamus and the ventromedial prefrontal cortex in stimulus-reward

1013   associative learning in the monkey. *J. Neurosci.* **10**, 3479–3493 (1990).

1014   42. Kim, H., Shimojo, S. & O'Doherty, J. P. Is Avoiding an Aversive Outcome

1015   Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. *PLOS*

1016   *Biol.* **4**, e233 (2006).

1017   43. Oyoshi, T., Nishijo, H., Asakura, T., Takamura, Y. & Ono, T. Emotional and

1018   Behavioral Correlates of Mediodorsal Thalamic Neurons during Associative

1019   Learning in Rats. *J. Neurosci.* **16**, 5812–5829 (1996).

1020   44. Treede, R.-D., Kenshalo, D. R., Gracely, R. H. & Jones, A. K. The cortical

1021   representation of pain. *Pain* **79**, 105–111 (1999).

1022   45. Lenz, F. A. *et al.* Stimulation in the human somatosensory thalamus can

1023   reproduce both the affective and sensory dimensions of previously experienced

1024   pain. *Nat. Med.* **1**, 910–913 (1995).

1025   46. FeldmanHall, O., Mobbs, D. & Dalgleish, T. Deconstructing the brain's moral

1026   network: dissociable functionality between the temporoparietal junction and

1027   ventro-medial prefrontal cortex. *Soc. Cogn. Affect. Neurosci.* **9**, 297–306 (2014).

1028   47. Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A. & Saxe, R.

1029   Disruption of the right temporoparietal junction with transcranial magnetic

1030   stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci.*

1031   *U. S. A.* **107**, 6753–6758 (2010).

48. Van Bavel, J. J., FeldmanHall, O. & Mende-Siedlecki, P. The neuroscience of moral cognition: from dual processes to dynamic systems. *Curr. Opin. Psychol.* **6**, 167–172 (2015).

49. Öhman, A., Eriksson, A. & Olofsson, C. One-trial learning and superior resistance to extinction of autonomic responses conditioned to potentially phobic stimuli. *J. Comp. Physiol. Psychol.* **88**, 619 (1975).

50. Phillips, J. & Cushman, F. Morality constrains the default representation of what is possible. *Proc. Natl. Acad. Sci.* **114**, 4649–4654 (2017).

51. Crockett, M. J. *et al.* Dissociable Effects of Serotonin and Dopamine on the Valuation of Harm in Moral Decision Making. *Curr. Biol.* **25**, 1852–1859 (2015).

52. Zaki, J. & Mitchell, J. P. Intuitive prosociality. *Curr. Dir. Psychol. Sci.* **22**, 466–470 (2013).

53. Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E. & Roiser, J. P. Neurocomputational mechanisms of prosocial learning and links to empathy. *Proc. Natl. Acad. Sci.* **113**, 9763–9768 (2016).

54. Lockwood, P. L. & Wittmann, M. K. Ventral anterior cingulate cortex and social decision-making. *Neurosci. Biobehav. Rev.* **92**, 187–191 (2018).

55. Cutler, J. & Campbell-Meiklejohn, D. A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *NeuroImage* **184**, 227–241 (2019).

56. FeldmanHall, O., Dalgleish, T., Evans, D. & Mobbs, D. Empathic concern drives costly altruism. *Neuroimage* **105**, 347–356 (2015).

57. Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F. & Grafman, J. The neural basis of human moral cognition. *Nat. Rev. Neurosci.* **6**, 799 (2005).

1055 58. Moll, J. & Schulkin, J. Social attachment and aversion in human moral cognition.

1056   *Neurosci. Biobehav. Rev.* **33**, 456–465 (2009).

1057 59. Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine enhances model-based

1058   over model-free choice behavior. *Neuron* **75**, 418–424 (2012).

1059 60. Seymour, B. Pain: A Precision Signal for Reinforcement Learning and Control.

1060   *Neuron* **101**, 1029–1041 (2019).

1061 61. Wang, O., Lee, S. W., O'Doherty, J., Seymour, B. & Yoshida, W. Model-based

1062   and model-free pain avoidance learning. *Brain Neurosci. Adv.* **2**,

1063   2398212818772964 (2018).

1064 62. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory

1065   capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci.* **110**,

1066   20941–20946 (2013).

1067 63. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D.

1068   Characterizing a psychiatric symptom dimension related to deficits in goal-

1069   directed control. *Elife* **5**, e11305 (2016).

1070 64. Costa, V. D., Monte, O. D., Lucas, D. R., Murray, E. A. & Averbeck, B. B.

1071   Amygdala and ventral striatum make distinct contributions to reinforcement

1072   learning. *Neuron* **92**, 505–517 (2016).

1073 65. Eldar, E., Hauser, T. U., Dayan, P. & Dolan, R. J. Striatal structure and function

1074   predict individual biases in learning to avoid pain. *Proc. Natl. Acad. Sci.* **113**,

1075   4812–4817 (2016).

1076 66. Lockwood, P. L. *et al.* Neural mechanisms for learning self and other ownership.

1077   *Nat. Commun.* **9**, (2018).

67. Chong, T. T.-J. *et al.* Neurocomputational Mechanisms Underlying Valuation of Effort Costs. *Plos Biol.* (in press).

68. Klein-Flügge, M. C., Kennerley, S. W., Friston, K. & Bestmann, S. Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *J. Neurosci.* **36**, 10002–10015 (2016).

69. Tobler, P. N., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Human neural learning depends on reward prediction errors in the blocking paradigm. *J. Neurophysiol.* **95**, 301–310 (2006).

70. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).

71. Kahane, G. *et al.* Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. (2017).

72. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).

73. Patil Indrajeet *et al.* Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. https://doi.org/10.31234/osf.io/q86vx. *PsyArXiv* (2018).

74. Miller, R. M., Hannikainen, I. A. & Cushman, F. A. Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion* **14**, 573 (2014).

75. Lockwood, P. L. *et al.* Prosocial apathy for helping others when effort is required. *Nat. Hum. Behav.* **1**, 0131 (2017).

76. Price, D. D. Psychological and neural mechanisms of the affective dimension of pain. *Science* **288**, 1769–1772 (2000).

1102   77. Will, G.-J., Rutledge, R. B., Moutoussis, M. & Dolan, R. J. Neural and

1103   computational processes underlying dynamic changes in self-esteem. *eLife* **6**,

1104   e28098 (2017).

1105   78. Railton, P. Moral Learning: Conceptual foundations and normative relevance.

1106   *Cognition* **167**, 172–190 (2017).

1107   79. Nagel, T. What is it like to be a bat? *Philos. Rev.* **83**, 435–450 (1974).

1108   80. Harsanyi, J. C. Cardinal welfare, individualistic ethics, and interpersonal

1109   comparisons of utility. *J. Polit. Econ.* **63**, 309–321 (1955).

1110   81. Bentham, J. *An introduction to the principles of morals and legislation*.

1111   (Clarendon Press, 1879).

1112   82. Kant, I., Semple, J. W. & Calderwood, H. *The Metaphysic of Ethics*. (T. & T.

1113   Clark, 1871).

1114   83. Huys, Q. J. M. *et al.* Bonsai trees in your head: how the pavlovian system sculpts

1115   goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* **8**, e1002410

1116   (2012).

1117   84. Huys, Q. J. M. *et al.* Disentangling the Roles of Approach, Activation and

1118   Valence in Instrumental and Pavlovian Responding. *PLOS Comput. Biol.* **7**,

1119   e1002028 (2011).

1120   85. Mackay, D. Model comparison and Occams Razor. *Inf. Theory Inference Learn.*

1121   *Algorithms* 343–355 (2003).

1122   86. Palminteri, S., Wyart, V. & Koechlin, E. The Importance of Falsification in

1123   Computational Cognitive Modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).

1124   87. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for

1125   spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* **113**, 7900

1126   (2016).

1127   88. Poldrack, R. A. *et al.* Guidelines for reporting an fMRI study. *Neuroimage* **40**,

1128   409–414 (2008).

1129   89. Mars, R. B. *et al.* Connectivity-Based Subdivisions of the Human Right

1130   'Temporoparietal Junction Area': Evidence for Different Areas Participating in

1131   Different Cortical Networks. *Cereb. Cortex* **22**, (2012).

1132   90. Palomero-Gallagher, N. *et al.* Functional organization of human subgenual

1133   cortical areas: Relationship between architectonical segregation and

1134   connectional heterogeneity. *NeuroImage* **115**, 177–190 (2015).

1135   91. Sallet, J. *et al.* The organization of dorsal frontal cortex in humans and

1136   macaques. *J. Neurosci.* **33**, 12255–12274 (2013).

1137