**Moral judgments are value-based decisions driven by culturally stable valuations and culturally variable decision biases.**

Word Count: 5466

**Abstract**

Many theorize that cultural similarities in moral judgments arise from a specialized cognitive system devoted to morality. We claim, in contrast, that people make moral judgments using a general-purpose, value-based decision-making process. We present a computational cognitive model to predict response time and response choice to moral dilemmas using valuations as input. Cultural similarities in moral judgment are explained by a *culturally stable set of valuations* that drives choices that aid survival. Corresponding cultural differences are explained by changes in a decisional bias parameter that accounts for differences in the perceived costs of making various kinds of decisional errors. The model accurately predicts the timed choices of both UK and Chinese respondents from values collected from USA respondents.

Keywords: Cross Culture; Moral Judgment; Value; Psychological Values; Social Cognition

**Moral judgments are value-based decisions driven by culturally stable valuations and culturally variable decision biases**

When people are faced with a moral judgment, they often — though not always — behave similarly despite their different cultural backgrounds (Awad et al., 2020). One explanation of cultural similarity in moral judgments is that humans evolved a specialized cognitive system that deploys to identify and produce judgments to questions that involve issues of “right” and “wrong” (e.g., a *universal moral grammar*, Hauser, 2006; Mikhail, 2007). A specialized cognitive system, however, is only necessary if a more general-purpose cognitive system cannot explain the data. Here we propose that moral judgments are made by deploying a general-purpose, value-based decision process that favors the option with the highest value. According to our view, cultural similarities arise because people share a *culturally stable set of valuations*. Cultural differences result from decisional biases induced by cultural variability in the perceived cost of committing decisional errors.

Central to evolutionary theory is the assumption that certain objects and events carry higher value in terms of adaptive usefulness than others (e.g., Ao, 2005), and accurate valuation is critical to successful survival of the species (e.g., Sznycer, 2022). Both behavioural and neuroscientific evidence supports an evolved valuation system. For example, Sznycer (2022) shows that assessments of value are consistent across individuals and correspond, to some degree, with their survival advantages/disadvantages. Additionally, there is extensive neuroscientific evidence that the neurons in the orbitofrontal cortex encode for subjective value (Fellows, 2011; Hare et al., 2008; Padoa-Schioppa, 2011, 2013; Padoa-Schioppa & Assad, 2006), whereas those in the ventromedial prefrontal cortex are activated when computing values of incommensurable objects (Chib et al., 2009; Levy & Glimcher, 2011, 2012; Lin, Adolphs, & Rangel, 2011).

We posit that this generalized value-based decision mechanism explains moral judgments and their cultural influences. To test this supposition, we call upon *Psychological Value Theory* (PVT; Cohen et al., 2022). PVT is a computational cognitive model of value and choice (for a tutorial on PVT, see https://[BLINDED] – REVIEWERS: Please see Psychological Value Theory: A Tutorial – in supplementary files for review). A central assumption of PVT is that people automatically perceive the “importance, worth or usefulness of an item to the observer” (op. cit. p. 2016), termed *Psychological Value*. Psychological Value is assumed to be a perceptual event, and thus noisy *within* the individual (Cohen & Ahn, 2016). As such, PVT quantifies Psychological Value as a distribution, rather than a discrete quantity. Like other perceptions, Psychological Value can only be directly experienced from the individual’s perspective. The concept of Psychological Value is different from the concept of a *util* as defined in subjective expected utility theory (Fishburn, 1981). A util indicates the subjectively judged benefit of the consequence of the choice, whereas Psychological Value represents the value of the *stimulus itself*.

We contend that a stable set of Psychological Values has evolved to drive choices. Therefore, the distribution representing the Psychological Value of a stimulus is assumed to be relatively stable across individuals. For example, whereas I value my child more than yours and you value your child more than mine, we both value “our own children” more than we value others. The value of “our own children” however is noisy and the distribution that represents it is essentially the same across different individuals.

Cohen et al., (2022) estimates Psychological Value distributions using a magnitude estimation task. Magnitude estimation is widely studied and has been consistently demonstrated to be a valid measure of perceptual events (e.g., Gescheider, 1988; Marks, 1974). Cohen and colleagues asked USA undergraduates, on every trial, to provide their Psychological Value of a probe relative to that of a chimpanzee valued at 1000 points (e.g., if the participant valued the probe half as much as the chimpanzee, they should assign it a value of 500, etc.). Participants were free to respond with any rational number. Each probe could contain the description of a human being (e.g., a nun), a group of humans (e.g., 5 terrorist), a concrete object (e.g., a toaster) or a group of such objects (5 TVs). PVT maintains that an individual’s magnitude estimate of Psychological Value for a probe represents a single sample from the distribution of the Psychological Value of the stimulus. Assuming Psychological Value is stable across individuals, the sample of estimates across individuals for a probe will be an accurate estimate of that probe’s Psychological Value distribution.

Because PVT quantifies Psychological Value and the decision process, PVT makes precise point predictions about the RT and choice in various experimental tasks (Cohen et al., 2022). According to PVT, the observer will compare the Psychological Values associated with the options and attempt to choose the option with the highest Psychological Value. Cohen et al. (2022) developed a Robust Random Walk (RRW) that quantifies this assumption by incrementally accruing evidence from the Psychological Value distributions of the options until it meets the threshold (termed a *boundary* in the RRW) for one of the two response options. In general, the evidence will accumulate toward the threshold that represents the higher valued option (the HVO). However, because this is a stochastic process, the system makes errors and will sometimes cross the threshold that represents the lower valued option (the LVO). For distributions that overlap considerably, the time to reach a boundary is relatively large as is the probability of an error. The opposite is true for distributions that overlap minimally. Any individual differences in performance might be taken as evidence for idiosyncratic differences in Psychological Values. PVT, however, predicts variable responding based solely on the stochastic decision mechanism. Therefore, whereas any individual response cannot be predicted, the average response is highly predictable.

If people evolved a common set of Psychological Values and these Psychological Values drive moral judgments, then the RRW should be able to predict the moral judgments made by one group of people using the Psychological Values of another group of people as input. Amongst the moral judgments most widely studied across cultures are sacrificial moral dilemmas, typified by trolley problems (Ahlenius, & Tännsjö, 2012; Awad et al., 2018; Awad et al., 2020; Hauser et al., 2007). In the classic trolley problem (Thompson, 1985), a runaway train will kill five individuals located on its track if nothing is done (the *do nothing* option). However, one can pull a lever that diverts the train onto a siding with the consequence that one person is killed (the *do something* option). Cohen et al. (2022) demonstrated that the Psychological Values collected from one set of USA undergraduates accurately predicted the RT and response choices to sacrificial moral dilemmas (and economic choices) of a different sample of USA undergraduates (see also Cohen & Ahn, 2016). These experiments provided strong evidence that moral judgments were value-based decisions. Here we take this prediction one step further by examining whether estimates of Psychological Values taken from USA undergraduates predict the moral judgments of UK and Chinese undergraduates.

To predict responses from UK and Chinese participants accurately, PVT must account for the cultural variation in the likelihood of sacrificing the bystander in a sacrificial moral dilemma (Awad et al., 2018). It is well established that Chinese respondents are significantly less likely to endorse the “do something” option than Western counterparts (see e.g., Ahlenius, & Tännsjö, 2012). Nevertheless, across three variants of the classic problem all cultural groups maintain the same rank order for the prevalence of making the do something response (Awad et al., 2020). For example, all cultural groups reduce the likelihood of the do something response if, to save the five, the respondent must forcibly kill the one. Currently there is no compelling psychological account of exactly what may be responsible for such patterns of data. PVT provides one such account.

We predict that culture primarily influences a *decisional* *bias parameter* of the random walk process, rather than the decision process itself or Psychological Values. Specifically, we hypothesize that culture influences the position where evidence begins to accumulate in the RRW, termed the *start point*. When the start point is positioned equidistant between the two thresholds, it is *unbiased*. The start point is *biased* when it is set closer to one threshold than the other. When a person has a biased start point, they are favoring one type of error over another (Busemeyer, & Johnson, 2004; Cohen & Ahn, 2016). For example, when a person has a start point closer to the “do nothing” boundary in the trolley problem, they require more evidence to “do something” than to “do nothing”. This is because they perceive the cost of erroneously failing to act as lower than erroneously acting. Evidence for this shift would be present in the pattern of RTs and response choices (Busemeyer & Johnson, 2004).

Here, we examined whether the Psychological Values collected from USA undergraduates by Cohen and colleagues (Cohen & Ahn, 2016; Cohen et al., 2022) can be used to predict moral judgments made by UK and Chinese undergraduates, with differences across cultures accounted for by a shift in the start point parameter.

All procedures were approved by the Ethics Committee of the Department of Psychology, University of [XBLINDEDX] IRB (Experiments 1 & 2: #653; Experiment 3: #21274-1). All data and R code for the analysis of Experiments 1-3 can be downloaded at:
https://github.com/ccpluncw/ccpl\_data\_UKmorals2021.git. For each Experiment, all data were collected before analysis began.

**Experiment 1**

In Experiment 1, we presented participants with sacrificial moral dilemmas that asked respondents to choose between two individuals who differ by social status (e.g., ‘a soldier’ vs ‘a nun’; replicating Cohen et al., 2022). We used the Psychological Values of USA participants to predict UK participants’ RTs and response choices.

**Methods**

**Participants.**

Thirty-five naïve participants were recruited from the participant panel of the Department of Psychology, The University of York. The panel comprises members of the University: predominantly undergraduate students (aged 18–21 yrs.; see Supplementary Materials for more detailed description). Participants either received course credit or £4 for taking part.

Given the huge effect sizes present in Cohen et al. (2022) (e.g., *r2*>0.85) and 120 trials per participant, power≈1.0 to detect the effect in an individual participant. Because group data is important to show consistency in results, rather than increasing power, sample size is somewhat arbitrary. Sample sizes reflected the number of participants that voluntarily signed up for testing once the adverts had been posted.

**Stimuli.**

The text of the scenario was as follows: “Through circumstances out of your control, Probe1 is about to be killed, but Probe2 will not be affected. You have the opportunity to save Probe1. However, if you save Probe1, Probe2 will be killed. Would you save Probe1 causing Probe2 to be killed?” The terms *Probe1* and *Probe2* serve as placeholders for the actual probes, which were randomly selected from a list of 23 probes (see Table 1). The probes used consisted of individuals who differed in social roles (e.g., a nun, a judge), as well as their associated Psychological Values, acquired from Cohen et al, (2022, see Supplementary Materials). Cohen et al. (2022) showed that this simplified scenario produced identical results as those with more elaborate scenarios.

**Table 1: The probes collected by Cohen et al., (2022) and used in Experiment 1**.

| pedophile | thief | congressman | Billionaire | college student | elderly person |
| --- | --- | --- | --- | --- | --- |
| rapist | gang member | celebrity | homeless adult | police officer | orphan |
| terrorist | convict | judge | Olympian | mentor | soldier |
| assassin | addict | nun | Astronaut | adult |  |

Notes:

1. To ensure cultural appropriateness, the names of the following probes were changed for the UK participants; ‘hooligan’ was substituted for ‘gang member’, ‘Member of Parliament’ was substituted for ‘congressman’, and ‘university student’ was substituted for ‘college student’. The corresponding normed Psychological Values were those used previously for the substituted items.
2. All stimuli were basic level categories (e.g., astronaut) rather than specific exemplars (e.g., John Glen) because specific exemplars are do not carry the same information cross-culturally (see Cohen et al., 2022). For example, the value of a cow is likely much higher for a person of the Hindu faith than for a person of the Christian faith. The reverse is likely true when assessing the value of Jesus. Critically, these two specific stimuli are exemplars of the basic level category *deity*.

(c) These probes were selected by Cohen et al. (2022) because they were social roles whose distributional overlaps were relatively evenly dispersed across the entire range (i.e., 0-1; see Supplemental Materials).

**Procedure.**

All participants were tested individually in a windowless testing cubicle. The experimental scripts were written in JAVA and run on a Windows PC laptop (HP model: ProBook 450 G4).

To estimate RT without including reading time, we presented participants with a masked form of the scenario prior to unveiling the actual probes. In the masked scenario, the text was presented in full, with the exceptions that each character in Probe1 and Probe 2 were replaced with ‘+’ and “=”, respectively. A vertical progress bar acted as a countdown timer to the point that the probes would be revealed. The duration of the masked scenario was presented so that the participants had sufficient time (as gauged from pilot work) to read and understand the scenario in absence of any information about the probes. Cohen and Ahn (2016) demonstrated that this presentation method does not alter participants’ pattern of response choices relative to an unmasked variation.

On each trial, the presentation of the central fixation point, for 500 ms, was followed by the timed masked scenario, and then the unmasked scenario remained on screen until the participant pressed either the ‘d’ or ‘k’ key on the keyboard. The assignment of the “Yes” and “No” responses (i.e., *do something* and *do nothing*) to ‘d’ and ‘k’ was counterbalanced across participants. RT was recorded from the onset of the unmasked scenario to the participant’s response. There was no time-out interval.

Four practice trials were presented prior to the experimental trials. The practice trials consisted of the same trial format, but were innocuous questions about food preferences, rather than moral dilemmas. Following the practice trials there was a further single block of 120 experimental trials. On each trial, the pair of probes were randomly sampled from the 253 possible combinations of the 23 types of individuals (presented as singletons ‘a terrorist’, ‘a nun’, etc) taken 2 at a time.

**Results**

If the Psychological Values of US students predict the moral judgments of UK participants, then RT and response choices of UK participants should be a function of the overlap of the US Psychological Value distributions of the probes in the scenarios.Cohen et al. (2022) calculated a non-parametric measure of the overlap of each pair of probes. This overlap measure ranged from 0 (no overlap, highly dissimilar values) to 1 (complete overlap, identical values). Similar to Cohen and Ahn (2016), distributional overlap was rounded to the nearest 0.1 (defined as O0.1) for all analyses.

The data filtering procedure was developed to emphasize the automatic removal of potentially inattentive responses that may contaminate data collected via the internet. First, we calculated the overall RT Coefficient of Variation for each participant (CVRT=SDRT/MRT). CVRT is a measure of attention fluctuation that corrects for average RT (because SD systematically varies with RT). We automatically removed the 5% of participants with the greatest CVRT (2 participants). Individual outlier trials were filtered by removing the fastest and slowest 2.5% of trials for each level of Overlap. Individual participants who responded at or below chance, indicating that they were not devoting attention to the task were removed (1 participant). The remaining participants were analyzed (32 participants). Finally, we removed the effects of practice from RT for each participant by fitting an exponential decay function to log(RT) and taking the residual as our measure of RT (see Supplementary Materials).

We modelled participants’ responses using the RRW, a value-based sequential sampling procedure (VSSP) that was developed to instantiate the assumptions of PVT (for detailed description, see Cohen et al. 2022; the RRW and associated analysis are presented in more detail in the Supplementary Materials). Similar to other VSSPs (Busemeyer & Johnson, 2004; Krajbich, 2019), the RRW is based on the classic two-dimensional random walk process of decision making (Balakrishnan & Ratcliff, 1996). Unlike traditional VSSPs, the RRW uses direct measures of value (Overlap, O0.1) to drive the evidence accumulation, rather than estimating value. This provides strong constraints on the ability of the model to fit the data.

To fit the model, we first coded each trial to identify whether the group condemned to death by default is the HVO (as determined by the Psychological Values collected by Cohen et al., 2022), termed DefaultHVO, or the LVO, termed DefaultLVO. This categorization is critical to identifying a start point bias toward “do nothing” or “do something.” For example, if participants make their choices based on the Psychological Value of the options, but there is a start point bias toward “do nothing,” the probability of choosing the higher valued option, p(HVO), will be a function of Overlap for both DefaultHVO and DefaultLVO, but p(HVO) will be greater for the DefaultLVO trials than the DefaultHVO trials.

We summarized the data by calculating p(HVO) and mean RT by Overlap and Default HVO/LVO. We compared two model in their abilities to predict the data: an Unbiased Model and a Biased Model that includes a start point bias. Both models had one fixed parameter set at a value other than 0: the threshold parameter on the function estimating the influence of time on the importance of the data (*dA*=0.2). Both models had four free parameters: non-decision time, *Ter*; 0.5\*boundary separation, *b*; the standard deviation of the Gaussian noise added to the drift, *nSD*; and an exponent that estimates the influence of time on the importance of information, *dB*. Finally, the Biased model included a free parameter that quantifies the start point bias as the proportion of *b* that the start point shifts toward one of the response boundaries, *s*. The Biased Model instantiates the *a priori* preference for “no action” that is often discussed in the moral judgment literature as a deontological ethical strategy (Bago & De Neys, 2019). Although the Biased Model captures this inclination that is often found in the literature, it does not assume this inclination manifests from any specific ethical strategy (e.g., a deontological ethical strategy). To determine the best fit model, we compared the BIC of each model. When the BIC’s were essentially equivalent (within 1 or 2 points), we accepted the Unbiased Model (simpler) as the better model based on parsimony.

 ****

**Figure 1. The PVT computational model fit to the data from Experiments 1-3.**

The black symbols/lines are trials in which the condemned individual is the HVO and bystander is the LVO, whereas the grey symbols/lines represent the opposite. Within each panel, the top graph illustrates how the probability of choosing the HVO, *p*(HVO), changes with the distributional overlap of Psychological Value of the options. The bottom graphs illustrate the same with RT: on the left are the trials that respondents saved the HVO and on the right they saved the LVO. A bias toward the do-nothing choice is present when participants are slower and less accurate on the DefaultHVO trials relative to the DefaultLVO trials. The computational model produced extremely accurate fits (*r*2 > .88 for Experiments 1-3) and revealed a 10% shift toward the do-nothing boundary in Experiment 1 and a 57% shift toward the do-nothing boundary in Experiment 3.

The behavioral data are shown in Fig. 1. Table 2 presents the *r*2 and BIC values of each model. Table 3 presents the parameter values, *r*2, and BIC for the best fit model. The data were best fit by the Biased Model (*r2*=0.88), with a 10% start point shift towards the “do nothing” boundary. To assess the skill of the model to fit the data, we ran a 3-fold cross validation analysis (Bischl, Mersmann, Trautmann & Weihs, 2012; see Supplementary Materials for details). As can be seen in Table 3, the average r2 for the *k*-fold cross validation analysis is close to that fit to the entire dataset (though predictably lower because they were fit to 1/3 the data), demonstrating the skill of the model to fit the data.

**Table 2. The PVT Unbiased and Biased model fits for Experiments 1-3**

|  |  |  |  |  | Model Fit |  |
| --- | --- | --- | --- | --- | --- | --- |
| Exp. | Country | Probes | Model |  | *r2* | BIC | Best Fit |
| 1 | UK | Individual |  |  |  |  |  |
|  |  |  | Unbiased Model |  | 0.84 | -119 |  |
|  |  |  | Biased Model |  | 0.88 | -128 | **✓** |
| 2 | UK | Group |  |  |  |  |  |
|  |  |  | Unbiased Model |  | 0.91 | -176 | **✓** |
|  |  |  | Biased Model |  | 0.91 | -177 |  |
| 3 | China | Group |  |  |  |  |  |
|  |  |  | Unbiased Model |  | 0.32 | -33 |  |
|  |  |  | Biased Model |  | 0.93 | -147 | **✓** |

**Table 3.** **The free parameter values for the best fit PVT models for Experiments 1-3**

| Exp. | Country | Probes | Free Parameters |  | Model Fit | k-fold validity |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Ter | b | dB | nSD | s |  | *r2* | BIC | *r2* |
| 1 | UK | Individual | -0.92 | 145 | 0.07 | 3 | 0.1 |  | 0.88 | -128 | 0.82 |
| 2 | UK | Group | -0.84 | 69 | 0.06 | 3.35 | NA |  | 0.91 | -176 | 0.89 |
| 3 | China | Group | -0.46 | 158 | 0.13 | 2.75 | 0.57 |  | 0.93 | -147 | 0.88 |

**Discussion**

In Experiment 1, PVT produced an excellent fit to the data. It has been repeatedly shown that the prevalence of a particular responses (i.e., do nothing vs. do something) is essentially the same across Western individuals (Awad et al., 2018; Awad et al., 2020; Hauser et al., 2007). Here we provide for the first time a detailed theoretical account as to why this pattern is the same – participants are making value-based decisions that are based on a common set of Psychological Values. By accommodating the tendency toward doing nothing with the same process as the drive to choose the highest valued option, PVT provides a unified explanation of participants’ judgment process.

**Experiment 2**

In Experiment 2, we replicated the methodology of Experiment 1 with the exception that the number of people in each option was systematically varied (e.g., 3 nuns vs 47 soldiers). This study pits the predictive power of the PVT against the contrasting idea that judgments in sacrificial moral dilemmas reflect rumination involving a utilitarian desire to save more rather than fewer lives (Paxton & Greene, 2010).

**Methods**

**Participants.**

Sixty-seven naïve participants were recruited from the same participant panel as Experiment 1 (see Supplementary Materials for demographic information). Participants either received course credit or £4 for taking part. Sample size was determined by the same procedure used in Experiment 1.

**Stimuli.**

The stimuli in Experiment 2 were identical to Experiment 1 with the exception that the number of people in the probes varied by group size: single (1, e.g., “a nun”), small (3-9; e.g., “3 nuns”), and large (47-53, e.g., “47 nuns”).

**Procedure**

The procedure in Experiment 2 was identical to Experiment 1 with the exception that in Experiment 2 there were 69 probes (23 individuals \* 3 quantity conditions). On each trial, the pair of probes were randomly sampled from the 2346 possible combinations of the 69 probes taken 2 at a time. The scenarios were edited appropriately for group size (i.e., “is” changed to “are”).

**Results**

The analysis in Experiment 2 followed that of Experiment 1. We removed four participants with the CVRT criterion, filtered individual outlier trials, and removed five participants who responded at or below chance. The remaining 58 participants were analyzed. Finally, we removed the effects of practice from RT.

We compared the same two RRW models used in Experiment 1 to predict the RT and choice of UK participants from the Psychological Values of US participants. Summaries of the key data are provided in Figure 1 and Tables 1 and 2. The best fit model from Experiment 2 was ambiguous. Both the Unbiased Model (*r2*=0.91, BIC= -176) and the Biased Model (*r2*=0.91, BIC= -177), with a 4% start point shift towards the “do nothing” boundary fit equivalently. For this reason, we accepted the model with fewer parameters as the best fit model – the Unbiased Model. The results of the 3-fold cross validation analysis are presented in Table 3 (and see Supplementary Materials).

 We also compared the ability of the difference in group size to predict RT and response choice to that of Psychological Value. A quantity-based model predicts that RT will be positively related to the difference in the quantity of people in Probe A and Probe B, termed *qDiff*, and the probability the participant chose to save the probe containing the greater quantity of people, termed *p*(HQO), will be negatively related qDiff. We fit two linear regressions, one defined as *RTres*=*a*+(*m*\*qDiff) and one defined as *p*(HQO)=*a*+(*m*\**q*Diff), where *a* and *m* are free parameters. Group size difference did not predict RT, *F*(1,21)=0.36, *p*=0.55, *r2*=0.016; but did significantly predict choice, *F*(1,21)=10.97, *p*=0.003, *r2*=0.34.

We ran an analysis to determine if Psychological Values could predict *p*(HQO) more accurately than group size difference. The probe pairs were divided into two sets: those pairs in which the higher valued option was also the higher quantity option (termed *consistent*) and those pairs in which the higher valued option was the lower quantity option (termed *inconsistent*). If psychological values are playing an instrumental role, then for the consistent trials, *p*(HQO) should decrease as Overlap increases. The inconsistent trials should show an equivalent, but exactly opposite effect, *p*(HQO) should increase as Overlap increases. We quantified the above predictions in the following linear model:

$p\left(HQO\right)=a+ m\_{1}\left(-1\*trialType\right)+m\_{2}\left(trialType\*Overlap\right),$ (1)

where trialType=1 for inconsistent trials, trialType= -1 for consistent trials, and Overlap is the average Overlap collapsed across subjects for each group size difference in the scenario. This model was significant, *F*(2,40)=139.8, *p*<0.001, *r2*=0.87. All parameters were significant, *a*=0.69, *m1*=0.86, and *m2*=1.12, *t*’s>6, *p*<0.001.

**Discussion**

Experiment 2 demonstrated that the estimates of Psychological Value taken from USA respondents accurately predicts the responses of UK participants. Critically, group size difference did not explain a significant amount of variance in RT and explained far less variance (slightly over 0.3) than Psychological Value (approximately 0.9) in choice. This suggests that the influence of group size difference is a mere artifact of Psychological Value changes. The current results replicate the findings of Cohen et al. (2022), who conducted an extensive examination of the influence of group size difference on choice in sacrificial moral dilemmas.

**Experiment 3**

In Experiment 3, we replicated the methods of Experiment 2 with undergraduates in China. It has been repeatedly shown that the prevalence of the kinds of responses vary across Western and Asian cultures (Awad et al., 2018; Awad et al., 2020; Hauser et al., 2007) with Chinese respondents being ‘significantly less prone to support utility-maximizing alternatives (Ahlenius, & Tännsjö, 2012), but, aside from establishing the robustness of this finding, no detailed account has been forthcoming. PVT provides one such an account: Moral judgments are value-based decisions and Chinese respondents have a start point bias towards the do-nothing boundary.

**Methods**

**Participants.**

Forty-two naïve participants were undergraduate and graduate students from Beijing Normal University and Tsinghua University (Age:19-28 yr., Mean=22.7, SD=2.6, 28 female, 13 male, 1 prefers not to say). Participants were recompensed with 35 RMB for taking part. Sample size was determined by the same procedure used in Experiment 1.

**Stimuli.**

The stimuli in Experiment 3 were identical to Experiment 2 except all textual materials were translated and presented in Chinese typeset, and we removed all variations of ‘Member of Parliament’ because of political sensitivities in China.

**Procedure**

The procedure in Experiment 3 were identical to Experiment 2 with the following exceptions. The experiment was re-written in JavaScript (including JSPsych library, de Leeuw, 2015) and a link to the html file was disseminated to the Chinese participants. These participants ran the experiment on their own computers in their own time. Furthermore, a descending dot replaced the progress bar due to programming constraints and no practice trials were included.

**Results**

The analysis in Experiment 3 followed that of Experiment 1. We removed three participants with the CVRT criterion, filtered individual outlier trials, and removed no participants who responded at or below chance. The remaining 39 participants were analyzed. Finally, we removed the effects of practice from RT.

We compared the same two RRW models used in Experiments 1 and 2 to predict the RT and choice of Chinese participants from the Psychological Values of US participants. Summaries of the key data are provided in Figure 1 and Tables 1 and 2. The Biased Model provided the best fit (*r2*=0.93), with a large 57% start point shift towards the “do nothing” boundary. This bias can be seen in the *p*(HVO) data, whereby participants are much more inclined to save the HVO in the DefaultLVO trials than the DefaultHVO trials. The results of the 3-fold cross validation analysis are presented in Table 3 (and see Supplementary Materials).

In addition, the group size analysis on the data from the Chinese undergraduates reveals a high degree of concordance with that of the UK undergraduates. As in Experiment 2, group size difference did not predict RT: *F*(1,20)=0.1, *p*=0.75, *r2*=0.01, but group size difference did significantly predict choice, *F*(1,20)=8.98, *p*=0.007, *r2*=0.31. Additionally, Psychological Value accounted for a significant amount of the variance in choice; *F*(2,39)=134.4, *p*<0.001, *r2*=0.87, and all parameters were significant, *a*=0.61, *m1*=0.46, and *m2*=0.48, *t*’s>2.9, *p*<0.006. As before the group size difference explained far less variance than Psychological Value.

**Discussion**

Experiment 3 demonstrated that the estimates of Psychological Value taken from USA respondents accurately predicts the responses to sacrificial moral dilemmas of Chinese participants. A 57% start point bias toward the “do nothing” boundary explains the Chinese participants’ tendency to respond “do nothing.” Thus, the underlying decision process and evidence structure is the same for Chinese and UK participants.

**General Discussion**

Researchers have proposed that cultural similarity in moral judgments arises from a cognitive system that specializes in moral judgments (e.g., Hauser, 2006; Mikhail, 2007). Such a system, however, is only necessary if a more general-purpose cognitive system cannot explain the data. Here we demonstrate that PVT, a computational cognitive model of a general-purpose, value-based system, can explain moral judgments across cultures. In three experiments, PVT accurately predicted UK and Chinese participants’ moral judgments using Psychological Values collected independent of choice and context from USA participants (Cohen et al., 2022). In Experiments 2 and 3, PVT predicted response choice and RT more accurately than a quantity-based model. These results strongly indicate that moral judgements are value-based decisions driven by culturally stable valuations (see also validation and sensitivity tests in the Supplementary Materials).

There is extensive neuroscientific evidence for a specialized valuation system in primates (for a review, see Wallis, 2012). Furthermore, this valuation system is used to drive a multitude of choices, including food preferences, partner preferences, economic decisions, etc. (for a review, see Dixon & Christoff, 2014). By accurately modeling moral judgments with PVT, we provide evidence that moral judgments are also value-based decisions. PVT is a highly constrained computational cognitive model of this system. PVT’s predictions will fail for many reasons: if moral judgments are not value-based, if the measurements of Psychological Value are inaccurate, if the decision mechanism is poorly modelled, if Psychological Values are not stable across individuals and cultures, etc. Nevertheless, PVT accurately predicted participants’ response choices and RTs when making moral judgments.

It has been hypothesized that a set of culturally stable valuations evolved to facilitate decisions that aid in survival (e.g., Sznycer, 2022). By demonstrating that valuations from USA participants predicted moral judgments of both western (UK) and eastern (Chinese) participants, we provide evidence for the existence and use of a culturally stable set of Psychological Values. Once a system evolves to make relatively fine distinctions in valuations, then that system will likely be deployed in a variety of complex ways. For example, an “astronaut” is a modern concept, so humans could not have evolved to value an “astronaut” in any specific manner. Nevertheless, astronauts are likely perceived to have traits that humans did evolve to value, such as rarity, bravery, skill, resources, etc. It is probable that the evolved valuation system automatically accessed these traits in the stored representation of an astronaut to make the valuation (future research should explore this issue). As such, cultural similarity in valuations of modern concepts may arises from shared learned conceptualizations and a common set of evolved Psychological Values.

Although it has been reported previously that Chinese participants present with a resistance to respond ‘do something’ relative to that of Western participants (see e.g., Ahlenius, & Tännsjö, 2012), there has as yet been no detailed account of why. Here we show that both UK and Chinese participants are attempting to save the HVO in sacrificial moral dilemmas and cultural variation result from changes in the start point bias. This type of decision bias is typically a function of the perceived cost of committing decisional errors. Within the context of PVT, there are two ways to make an error: “doing nothing” when the HVO is the condemned individual (termed a *miss*) and “doing something” when the HVO is the safe bystander (termed a *false alarm*). The start point is a quantification of how costly the respondent views a miss relative to a false alarm (Macmillan & Creelman, 2005, Chapt. 2). This suggests that Chinese respondents view a false alarm as more costly than a miss relative to UK respondents. Future research may address the cultural factors that influence this cost/benefit calculation. Furthermore, because start point biases are sensitive to the perceived cost of committing decisional errors (Macmillan & Creelman, 2005, Chapt. 2), future research should test how manipulating the cost of a miss relative to a false alarm influences the start point bias.

Our populations samples are limited to undergraduates (albeit in both western and eastern cultures), and we presented them with a limited set of probes (23 individuals with different social roles). It is important to ask, will the same results obtain with different stimuli (different social roles, animals, objects, etc.) and cultures (South Korean, Brazilian, Malian, etc.)? Furthermore, if exceptions are found, will that invalidate evolutionary claims? Tooby and Cosmidies addressed these questions in their 1989 commentary. They claim that evolved traits will *always* exhibit some variability across cultures because (a) genotypes may not manifest consistently (i.e., phenotype vs genotype) and (b) stochasticity in expression. Therefore, identifying variation across cultures and stimuli will not invalidate the claims about evolved systems. Rather, if evolution is not at play, then one should be able to identify cultures that have opposing biases. For example, Tooby and Cosmidies state, “The assertion that ‘culture’ explains human variation will be taken seriously when there are reports of women war parties raiding villages to capture men as husbands or of parents cloistering their sons but not their daughters to protect their sons' ‘virtue.’” (p. 37). Therefore, if valuation has evolved, we expect small shifts in valuation across cultures due to stochasticity (e.g., preference for an elderly man over a female child). To invalidate the evolutionary claim, researchers must identify cultures in which large valuation differences are identified (e.g., a culture that values rapists over police officers and gang members over Olympians). We encourage future research that assesses the range of cultural variability in Psychological Value associated with different stimuli/stimulus classes.

Unfortunately, we are unable to assess whether the extant data is consistent with our conclusions because researchers addressing cross cultural moral judgment typically do not use a model-based approach, nor do they measure the variables necessary for us to apply our model to their data (see e.g., Awad et al., 2018). Based on our data, we suspect that previously identified cross-cultural differences in response choices are also the result of decision biases. Rather than speculate further, however, we recommend that researchers adopt the model-based approach in the future (and we provide free access to our PVT analysis programs for such purposes).

Finally, PVT may provide a (perhaps controversial) answer to the question, “How do society’s deontological concepts arise?” Assuming people evolved a common set of Psychological Values and moral judgments are value-based, one may conclude that Psychological Values precede, and thus drive, people’s deontological concepts of what is “right” and “wrong”. For example, we have evolved to value children highly. We propose that this high valuation drives us to protect our children *and* drives our deontological belief that it is “right” and “good” to protect children.

**References**

Ahlenius, H., & Tännsjö, T. (2012). Chinese and Westerners respond differently to the trolley dilemmas. *Journal of Cognition and Culture, 12*, 195-201. doi: <http://dx.doi.org/10.1163/15685373-12342073>

Ao, P. (2005). Laws in Darwinian evolutionary theory. *Physics of Life Reviews, 2*, 117-156. doi: <http://dx.doi.org/10.1016/j.plrev.2005.03.002>

Awad, E., Dsouza, S., Kim, R., Schultz, J., Henrich, J., Shariff, A., . . . Rahwan, I. (2018). The moral machine experiment. *Nature, 563*, 59-64. doi: <https://doi.org/10.1038/s41586-018-0637-6>

Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J. F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 2332-2337. doi: <https://doi.org/10.1073/pnas.1911517117>

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective Dual Process Model of moral cognition. *Journal of Experimental Psychology: General, 148*, 1782-1801. doi: <http://dx.doi.org/10.1037/xge0000533>

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making under confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 615-633. doi: <http://dx.doi.org/10.1037/0096-1523.22.3.615>

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C. (2012). Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation, 20*, 249-275. doi: <https://doi.org/10.1162/EVCO_a_00069>

Busemeyer, J. R., & Johnson, J. G. (2004). *Computational models of decision making*. Oxofrd, UK: Blackwell Publishing.

Chibb, V. S., Rangel, A., Shimojo, S., & O’Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience, 29*, 12315-12320. doi: <https://doi.org/10.1523/JNEUROSCI.2575-09.2009>

Cohen, D. J., & Ahn, M. (2016). A subjective utilitarian theory of moral judgment. *Journal of Experimental Psychology: General, 145*, 1359-1381. doi: <http://dx.doi.org/10.1037/xge0000210>

Cohen, D. J., Cromley, A. R., Freda, K. E., & White, M. (2022). Psychological value theory: The psychological value of human lives and economic goods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*, 2015–2050. doi: <http://dx.doi.org/10.1037/xlm0001047>

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1-12. doi: <http://dx.doi.org/10.3758/s13428-014-0458-y>

Dixon, M. L., & Christoff, F. (2014). The lateral prefrontal cortex and complex value-based learning and decision making. *Neuroscience and Biobehavioral Reviews, 45*, 9-18. doi: <http://dx.doi.org/10.1016/j.neubiorev.2014.04.011>

Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: Evidence from humans with frontal lobe damage. *Annals of the New York Academy of Sciences, 1239*, 51-58. doi: <https://doi.org/10.1111/j.1749-6632.2011.06229.x>

Fishburn, P. C. (1981). Subjective expected utility: A review of normative theories. *Theory and decision, 13*, 139-199. doi: <http://dx.doi.org/10.1007/BF00134215>

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39*, 169-200.

Hare, T. A., O’Doherty, J. P., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience, 28*, 5623-5630. doi: <https://doi.org/10.1523/JNEUROSCI.1309-08.2008>

Hauser, M. D. (2006). *Moral minds. How nature designed our universal sense of right and wrong*. London, UK: Abacus.

Hauser, M. D., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*, 1-21. doi: <https://doi.org/10.1111/j.1468-0017.2006.00297.x>

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology, 29*, 6-11. doi: <https://doi.org/10.1016/j.copsyc.2018.10.008>

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. *Journal of Neuroscience, 31*, 14693-14707. doi: <https://doi.org/10.1523/JNEUROSCI.2218-11.2011>

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology, 22*, 1027-1038. doi: <https://doi.org/10.1016/j.conb.2012.06.001>

Lin, A., Adolphs, R., & Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PLoS ONE, 6*, e21074. doi: <https://doi.org/10.1371/journal.pone.0021074>

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York, NY: Lawrence Erlbaum Associates.

Marks, L. E. (1974). On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics, 16*, 358-376. doi: <https://doi.org/10.3758/BF03203956>

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences, 111*, 143-152. doi: <http://dx.doi.org/10.1016/j.tics.2006.12.007>

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: A good-based model. *Annual Review of Neuroscience, 34*, 333-359. doi: <https://doi.org/10.1146/annurev-neuro-061010-113648>

Padoa-Schioppa, C. (2013). Neuronal origins of choice variability in economic decisions. *Neuron, 80*, 1322-1336. doi: <https://doi.org/10.1016/j.neuron.2013.09.013>

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*, 223-226. doi: <https://doi.org/10.1038/nature04676>

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science, 2*, 511-527. doi: <http://dx.doi.org/10.1111/j.1756-8765.2010.01096.x>

Sznycer, D. (2022). Value computation in humans. *Evolution and Human Behavior, 43*, 367-380. doi: <http://dx.doi.org/10.1016/j.evolhumbehav.2022.06.002>

Thomson, J. J. (1985). The trolley problem. *Yale Law Journal, 94*, 1395-1415. doi: <http://dx.doi.org/10.2307/796133>

Tooby, J., & Cosmides, L. (1989). The innate versus the manifest: How universal does universal have to be? *Behavioral and Brain Sciences, 12*, 36-37.

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience, 15*, 13-19. doi: <http://dx.doi.org/10.1038/nn.2956>