Journal of Neuropsychology (2017)
© 2017 The British Psychological Society

The British Psychological Society

www.wileyonlinelibrary.com

Counter-intuitive moral judgement following traumatic brain injury

Dane A. Rowley[1], Miles Rogish[2], Timothy Alexander[1] and Kevin J. Riggs[3]*

[1]Clinical Psychology Programme, University of Hull, UK
[2]Brain Injury Rehabilitation Trust, The Disabilities Trust, West Sussex, UK
[3]Department of Psychology, University of Hull, UK

Several neurological patient populations, including traumatic brain injury (TBI), appear to produce an abnormally 'utilitarian' pattern of judgements to moral dilemmas; they tend to make judgements that maximize the welfare of the majority, rather than deontological judgements based on the following of moral rules (e.g., do not harm others). However, this patient research has always used extreme dilemmas with highly valued moral rules (e.g., do not kill). Data from healthy participants, however, suggest that when a wider range of dilemmas are employed, involving less valued moral rules (e.g., do not lie), moral judgements demonstrate sensitivity to the psychological intuitiveness of the judgements, rather than their deontological or utilitarian content (Kahane *et al.*, *Social Cognitive and Affective Neuroscience*, 7, 2011, 393). We sought the moral judgements of 30 TBI participants and 30 controls on moral dilemmas where content (utilitarian/deontological) and intuition (intuitive/counter-intuitive) were measured concurrently. Overall TBI participants made utilitarian judgements in equal proportions to controls; disproportionately favouring utilitarian judgements *only when* they were counter-intuitive, and deontological judgements *only when* they were counter-intuitive. These results speak against the view that TBI causes a specific utilitarian bias, suggesting instead that moral intuition is broadly disrupted following TBI.

Research on the cognitive and neural bases of moral judgments has blossomed in the last 15 years and a clear finding appears to have emerged: utilitarian judgements (i.e., those that maximize aggregate welfare) are associated with increased activation in a core group of frontal brain areas implicated in deliberate controlled processing; deontological judgements (i.e., those judgements that conform to moral laws) are associated with those brain areas associated with automatic processing (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Moral judgement has also been investigated in neurological populations with frontal lobe lesions (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs *et al.*, 2007) and traumatic brain injury (TBI; Martins, Faísca, Esteves, Muresan, & Reis, 2012); populations who characteristically show emotional blunting, impaired empathy and social cognition, egocentrism (Mitchell, Avny, & Blair, 2006; Müller, Schuierer, Marienhagen, Putzhammer, & Klein, 2003), and demonstrate

*Correspondence should be addressed to Kevin J. Riggs, Department of Psychology, University of Hull, Hull HU6 7RX, UK (email: k.riggs.@hull.ac.uk)

DOI:10.1111/jnp.12117

socially inappropriate behaviour (Beer, John, Scabini, & Knight, 2006; Cicerone & Tanenbaum, 1997; Pitman, Haddlesey, Ramos, Oddy, & Fortescue, 2014). This profile is observed routinely in TBI, where neuropathology is caused by an impact to, or rapid acceleration/deceleration of, the brain (Lezak, Howeison, Bigler, & Tranel, 2012). Neural damage is characteristically diffuse in TBI, but the frontal cortex is especially vulnerable to lesion (Lezak *et al.*, 2012). In addition, subcortical and white matter tract damage caused by traumatic axonal injury compromises the integrity of neural networks, causing disruption of functions reliant on the integrity of these networks (Hayes, Bigler, & Verfaellie, 2016; Lipton *et al.*, 2009). In this study, we investigate the moral judgements made by TBI patients to further our understanding of the cognitive and neural bases of moral judgement.

In a two-system cognitive account of moral judgement, Greene *et al.* (2004) characterize ‘system 1’ as the rapid and automatic processes delivering moral judgement, while higher order processes of deliberative reasoning are engaged by ‘system 2’. The automatic system biases towards ‘deontological’ moral judgements – judgements that conform to moral laws such as *do not lie; do not harm others* (Kant, 1785/1959), whereas controlled processing allows us to override these judgements in favour of more *reasoned* ‘utilitarian’ judgements – ones that maximize aggregate welfare (Greene *et al.*, 2008). Greene *et al.* (2004) provide data to support this assertion. In dilemmas where utilitarian judgements required the maiming or killing of another person (e.g., the infamous ‘trolley dilemma’, where participants are asked whether they should pull a lever to divert the course of a train, thus condemning one bystander to death but saving five others) participants took longer to endorse utilitarian actions than deontological ones. Additionally, neural activity in dorsolateral prefrontal and anterior cingulate cortices (areas associated with controlled processing) correlated with utilitarian moral judgement. This was taken as evidence for the involvement of effortful cognition in utilitarian judgements, both in these extreme scenarios and more broadly (Greene *et al.*, 2001, 2004, 2008).

However, data from patient studies would appear to pose a problem for Greene’s model. Populations with TBI (Martins *et al.*, 2012), circumscribed ventromedial prefrontal cortex (VMPFC) lesions (Ciaramelli *et al.*, 2007; Koenigs *et al.*, 2007), fronto-temporal dementia (Mendez, Anderson, & Shapira, 2005) and psychopathy (Koenigs, Kruepke, Zeier, & Newman, 2012) all show a utilitarian bias compared to healthy controls in moral dilemmas, which appears at odds with the view that these judgements require careful and controlled moral processing. One explanation for this is that patients have impairments in social cognition (e.g., empathy, perspective taking, theory of mind [ToM]) and consequently have a reduced aversion to harming others (i.e., a reduced aversion to killing the lone protagonist in the service of saving many others – see the example above). Support for this idea is seen in skin conductance response (SCR) studies: a strong SCR precedes utilitarian judgements in healthy controls, but no such response is seen when patients with VMPFC lesions make identical judgements (Moretto, Làdavas, Mattioli, & Di Pellegrino, 2010). Further, in healthy participants, reduced aversion to harming others (Cushman, Gray, Gaffey, & Mendes, 2012), lower trait empathy (Choe & Min, 2011), and higher psychoticism (characterized by reduced empathy and emotional blunting; Wiech *et al.*, 2013) all correlate with increased levels of utilitarian judgement.

Taken together then, the evidence suggests a link between social cognition and utilitarian judgements in both clinical and non-clinical populations, and that this link may arise because the extent of our aversion to harming others may influence how appealing utilitarian solutions to moral dilemmas are. In this study, we therefore included measures of social cognition alongside moral dilemmas.

Another problem for the two systems theory is that only extreme moral dilemmas have been employed to test it, where utilitarian judgements required the violation of highly regarded deontological rules, such as *do not kill.* As such, the observed association between controlled processing and utilitarian judgement could be an artefact of the limited range of dilemmas employed. One possibility is that in extreme dilemmas, deontological judgements are psychologically intuitive, whereas utilitarian judgements are psychologically counter-intuitive. This possibility has *prima facie* appeal; judgements which endorse murder in the service of aggregate welfare might not be immediately appealing, whereas judgements which only require a lie or a broken promise may be immediately compelling, or intuitive.

For this reason then, Kahane *et al.* (2011) devised new dilemmas which captured the tension between maximizing aggregate welfare and adherence to moral rules, while controlling for psychological intuitiveness. They collected normative data for these dilemmas; recording the non-reflective judgements of a group of independent judges and assigning dilemmas to one of two categories. A dilemma was categorized as *Intuitively Utilitarian (UI)* when most judges intuitively violated the moral rule to maximize aggregate welfare. For example:

> 'You know a man called Fred. Fred is a prejudiced and grumpy person who often takes a disliking to people for no good reason. You also have a friend who admires Fred and gives great weight to his opinions. However, Fred despises your friend. One day, your friend asks you what Fred thinks of him. Your friend would be devastated to discover that Fred despises him, but will only find out if you tell him. Should you tell your friend that Fred despises him?' [Adapted from the original]

In this case, normative data indicated that people disregarded the deontological rule 'do not lie' in favour of the course of action which maximized welfare (preserving your friend's self-esteem) and so it was categorized as a UI dilemma. Thus, a utilitarian judgement in these dilemmas is also an 'intuitive' judgement, while a deontological judgement is a 'counter-intuitive' judgement.

Conversely, dilemmas were categorized as *Intuitively Deontological* (DI) when most judges upheld the deontological rule. This category involved deontological rules which were considered more absolute, such as the impermissibility of killing. For example:

> 'You are a Doctor. You have five very poorly patients who are all about to die of various failing organs. You have another patient who is healthy. The only way you can save the lives of the first five patients is to remove this man's organs, and transplant them into the five poorly patients. The healthy man does not want you to take away his organs. If you do this, the health man will die, and the five will live. Should you perform these transplants?' [Adapted from the original]

The normative sample overwhelmingly rejected utilitarianism here, choosing to uphold the deontological rule despite the net harm (five deaths rather than one). As such, this was categorized as a DI dilemma. In this category then, a utilitarian judgement is a 'counter-intuitive' judgement, and a deontological judgement is an 'intuitive' judgement (the exact inverse of the UI dilemmas, thus allowing preferences for utilitarian and deontological judgements to be measured independently of their intuitiveness).

In an fMRI study using these new stimuli (Kahane *et al.*, 2011), the previously reported neural and behavioural association between controlled processing and utilitarian judgements disappeared. Healthy participants rated counter-intuitive judgements as more difficult than intuitive judgements, but did not rate utilitarian judgements as more difficult than deontological ones. Furthermore, the pattern of neural activation was related to the intuitiveness of judgements.

During counter-intuitive judgements, activation was recorded in the rostral and dorsal cingulate cortex, primary, and secondary somatosensory cortex, insula, ventro-lateral pre-frontal cortex, and lateral orbitofrontal cortex, irrespective of the (deontological/utilitarian) content of the judgement. Kahane *et al.* (2011) concluded that previous findings associating utilitarian judgements with controlled processing were an artefact of the limited dilemmas employed, and that healthy people use controlled processing when making any counter-intuitive moral judgement, regardless of its content (though see Paxton, Bruni, & Greene, 2014). They note, however, that it remains unclear precisely *which* controlled processes are involved in moral judgements (e.g., inhibitory control, attentional flexibility or working memory).

During intuitive judgements, activation was recorded in the visual, pre-motor, and ventromedial pre-frontal cortices, and the temporal lobe; areas which have been associated with various aspects of social cognition: empathy (Nummenmaa, Hirvonen, Parkkola, & Hietanen, 2008), affective ToM (Shamay-Tsoory, Tibi-Elhanany, & Aharon-Peretz, 2006), and emotional perspective taking (Lamm, Batson, & Decety, 2007). Indeed, a trait tendency towards empathy increases preference for deontological judgements in extreme dilemmas (Crockett, Clark, Hauser, & Robbins, 2010; Gleichgerrcht & Young, 2013). ToM in particular is understood to rely on a distributed cortical and subcortical network comprising (at least) the medial pre-frontal cortex, left and right temporo-parietal junctions, the temporal poles, and the amygdala circuitry (Apperly, 2010; Siegal & Varley, 2002). It is noteworthy that the VMPFC, a necessary area for affective ToM (Shamay-Tsoory *et al.*, 2006), was implicated in intuitive moral judgement in Kahane and colleagues' fMRI study.

In sum then, the evidence suggests that the intuitiveness of a moral judgement, rather than its content, is the key factor in controlled versus automatic processing, and thus, there is reason to doubt reports of utilitarian bias in focal frontal injury, TBI, and other clinical populations including autism, fronto-temporal dementia, and psychopathy (Ciaramelli *et al.*, 2007; Gleichgerrcht *et al.*, 2013; Koenigs *et al.*, 2007; Martins *et al.*, 2012; Mendez *et al.*, 2005), as all of these studies employed a limited range of extreme dilemmas which did not control for intuitiveness. It also appears that social cognition, including empathy and ToM, likely plays a role in moral judgement. Moreover, a wealth of evidence demonstrates that ToM is compromised following TBI (Bibby & McDonald, 2005; Martín-Rodríguez & León-Carrión, 2010; Muller *et al.*, 2010), as are other abilities implicated in moral judgement, such as empathy and emotional expressiveness and regulation (Beer *et al.*, 2006; Cicerone & Tanenbaum, 1997; Mitchell *et al.*, 2006; Müller *et al.*, 2003; Pitman *et al.*, 2014; Stuss, 2011).

The present study

To date, no study has investigated the effect of brain pathology on both content and intuitiveness in moral judgement, and therefore, their relative importance in explaining atypical moral judgement patterns is unknown. To address this issue, this study employed a cross-sectional case–control design in which participants with TBI and healthy controls gave their moral judgements on dilemmas devised by Kahane *et al.* (2011). Participants also completed a range of social cognition measures and cognitive assessments.

If TBI causes a *specific* bias towards utilitarianism, then these participants should make more utilitarian judgements compared to controls, regardless of intuitiveness. Such a finding would suggest that the content (utilitarian/deontological) of moral judgement is relevant to the processes of automatic and controlled moral judgement, and would support, and extend previous findings to less extreme dilemmas involving lying and breaking promises.

However, if intuitiveness is the crucial factor, TBI participants should make more utilitarian judgements than controls only in DI dilemmas, *where utilitarianism is* counter-intuitive. This would indicate that the neural networks impacted by TBI are not sensitive to the content of a judgement *per se*, but instead its intuitiveness.

In dilemmas where the utilitarian option is *intuitive*, it remains unclear whether TBI participants would show a preference for counter-intuitive judgements. One possibility is that TBI causes a preference for counter-intuitive judgements only in extreme (DI) dilemmas where serious physical harm is at stake. Alternatively, TBI may result in a general tendency to make counter-intuitive judgements, irrespective of dilemma type.

Finally, if reduced aversion to harm underlies counter-intuitive judgement following TBI, then the TBI group should be able to make these judgements with relative ease. As such, we expect TBI participants to find counter-intuitive responses easier to make compared to controls, within both UI and DI dilemmas. In addition, if disruption of social cognition modulates moral judgement disturbance following TBI, then ToM processes (particularly its emotional components) should be associated with counter-intuitive moral judgement.

Method

Participants

Thirty adults (five female; mean age = 41.3 [*SD* = 13.67]) with non-penetrating TBI were recruited via NHS community neuropsychology services, Brain Injury Rehabilitation Trust inpatient and community services and the Headway charity across England. Inclusion criteria were as follows: (1) history of TBI, (2) at least 12 months post-injury, (3) fluent in English. Exclusion criteria were as follows: (1) significant visual, perceptual or language impairment; (2) TBI incurred before 18 years; (3) other neurological disorder; (4) current major depressive disorder, PTSD or psychosis; (5) developmental disorder. Self-report was the primary method used to determine eligibility, although the medical records of those recruited from clinical services were screened for eligibility in the first instance by treating clinicians.

Traumatic brain injury severity was categorized according to available information on post-traumatic amnesia duration, length of unconsciousness and lowest Glasgow Coma Scale (Jones, 1979) score, in that order of preference. Table 1 displays cut-offs for injury severity categorization and the number of TBI participants in each category.

Thirty healthy controls (11 female; mean age = 39.8 [*SD* = 14.56]) were recruited to match the demographic of TBI participants. Exclusion criteria were as follows: (1) neurological disorder; (2) current major depressive disorder, PTSD or psychosis; (3) developmental disorder. All participants gave informed consent, and the study was

Table 1. Classification of severity by post-traumatic amnesia duration (PTA), length of loss of consciousness (LOC) and Glasgow Coma Scale (GSC), and number of participants (*n*) in each group

Severity classification	PTA	LOC	GCS	*n*
Mild	<1 hr	<15 min	13–15	1
Moderate	1–24 hr	15 min–6 hr	9–12	3
Severe	24 hr–7 days	6–48 hr	3–8	8
Very severe	>7 days	>48 hr		18

approved by an NHS research ethics committee, in accordance with the World Medical Association (2013).

The TBI and control groups were comparable in terms of gender, $\chi^2(2, 60) = 3.068$, $p = .080$, and level of education ($U = 404.5$, $z = -0.706$, $p = .480$). The groups did not differ significantly in age or the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) scores, but differed significantly in verbal (VIQ), performance (PIQ), and full-scale (FSIQ) intellectual ability as measured by the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999; see Table 2).

Materials and procedure

All participants were tested on a range of moral dilemmas followed by testing on a number of social cognition, IQ, and depression and anxiety measures.

Moral dilemmas

Ten of the eighteen dilemmas from Kahane *et al.* (2011) were adapted for the study. These were selected to encompass the range of deontological rules involved in the originals, comprising five UI and five DI dilemmas. The dilemmas were rearranged into storyboards, and cartoon drawings were created to aid comprehension (see Table S1). In a piloting exercise, two groups (total $n = 18$) of independent judges gave their *non-reflective* responses to the original or the adapted dilemmas. On average, judges placed each dilemma in its originally assigned category 77% of the time (range = 67–100%; see

Table 2. Demographic, clinical, and cognitive characteristics of traumatic brain injury (TBI) and control groups

	TBI Group (*N* = 30) M (*SD*)	Control Group (*N* = 30) M (*SD*)	*t*	*p*
Age	41.3 (13.67)	39.8 (14.56)	0.402	.689
Years post-injury	9.3 (9.83)	N/A		
HADS depression	4.3 (3.67)	3.5 (4.35)	0.770	.445
HADS anxiety	6.0 (3.84)	6.2 (5.05)	−0.144	.886
HADS total	10.2 (7.01)	9.7 (8.97)	0.273	.786
VIQ	100.8 (19.17)	113.2 (12.47)	−2.954	.005**
PIQ	101.5 (17.21)	113.8 (21.11)	−2.487	.016*
FSIQ	101.3 (18.07)	117.5 (11.13)	−4.190	.000***

Notes. *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

supplementary data), and as such, all ten of the adapted dilemmas retained their original categorization based on previously employed cut-off of 67% agreement (Kahane *et al.*, 2011).

Dilemmas were presented to participants first, in a fixed randomized order on laminated paper. The experimenter read the dilemmas aloud once, before inviting participants to make a judgement on what they *should* do. Participants were then asked to rate the difficulty of each judgement on a 1 (*not difficult at all*) to 10 (*very difficult*) scale. Answers to each dilemma were recorded and subsequently categorized (intuitive/counter-intuitive; utilitarian/deontological).

Cognitive measures

In a perspective taking task (Tversky & Hard, 2009), participants were shown a photograph and asked to give the spatial location of an object, where the answer differed depending on whether participants took their own or another's visual perspective. This was taken as a measure of spontaneous perspective taking (an automatic process relevant to ToM). Participants were then administered the WASI (Wechsler, 1999). Following a break, participants undertook the Faux Pas (FP) test (Stone, Baron-Cohen, & Knight, 1998), which measures the ability to identify a social *faux pas*, and represent both the beliefs, intentions (cognitive ToM), and feelings (affective ToM) of characters involved. Then came the revised version of the Reading the Mind in the Eyes (RME: Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) which measured affective ToM by asking participants to ascribe an emotional experience to actors in 36 images of eyes, choosing one of four adjectives. Finally, participants completed the HADS (Zigmond & Snaith, 1983).

Data analysis

Within group analyses

Moral judgement data were transformed into proportions for each participant. Proportions of intuitive and counter-intuitive moral judgement were analysed separately for UI and DI dilemmas using one-sample *t*-tests, reporting 95% CI's. Preferences for utilitarian versus deontological judgements were analysed similarly, pooled across all dilemmas. Mean proportions were tested against a value of 0.5; the value expected if participants showed no preference for either option during moral judgement. Paired-samples *t*-tests were used to compare difficulty ratings between intuitive and counter-intuitive judgements in both UI and DI dilemmas, and between utilitarian and deontological judgements in all dilemmas.

Between group analyses

One-way mixed ANOVA analysed differences between groups and dilemma type (UI/DI) in proportion of counter-intuitive judgements. Dilemmas were then pooled across dilemma type, and group differences in counter-intuitive judgement were investigated using independent samples *t*-tests with 95% CI's. Group differences in utilitarian judgement were analysed similarly. As these data were proportional, the sum of intuitive and counter-intuitive judgements (and utilitarian and deontological judgements) for each participant was 1.0.

The difficulty cost of selecting the counter-intuitive response over the intuitive response, and the utilitarian response over the deontological response, was calculated by subtracting the latter from the former for each case. These were computed because both utilitarian judgements and counter-intuitive judgements should theoretically be more difficult than their opposites, according to the positions of Greene *et al.* (2008) and Kahane *et al.* (2011). A one-way mixed ANOVA was used to analyse the differences between groups and dilemma type in the difficulty cost of counter-intuitive judgements. Again, all dilemmas were then pooled and group differences investigated using independent samples *t*-tests.

Cognition analyses

Independent samples *t*-tests were employed to test for group differences on ToM and IQ variables, and Pearson's correlation coefficients were calculated for the whole sample between moral judgement and cognitive variables. BCa 95% CI's are reported.

For the TBI group, ToM variables were entered into a hierarchical multiple regression model, with proportion of counter-intuitive responses as the dependent variable. Bootstrapped *p*-values were computed. Affective ToM variables (FP empathy and RME) were entered at step one, and the cognitive ToM variable (FP cognitive index) at step two. Bootstrapping was used in these analyses due to non-normal distribution in the FP data.

Results

Within group analyses

Control group moral judgements

The proportion of intuitive judgements was significantly higher than the 0.5 baseline in both UI, $t(29) = 8.361, p < .001$, 95% CI (0.227, 0.373), and DI, $t(29) = 4.110, p < .001$, 95% CI (0.101, 0.300), dilemmas (see Figure 1A). The control group showed no significant preference for utilitarian (or deontological) judgements when all dilemmas were pooled together and compared against the 0.5 baseline, $t(29) = 1.455$, $p = .156$, 95% CI (−0.019, 0.112).

Control group difficulty ratings

Controls rated counter-intuitive judgements as significantly more difficult than intuitive judgements in both UI, $t(19) = -3.931$, $p = .001$, 95% CI (−3.24, −0.988), and DI, $t(20) = -3.839, p = .001$, 95% CI (−3.987, −1.179) dilemmas (see Figure 1B). However, difficulty ratings did not differ significantly between utilitarian and deontological judgements overall, $t(29) = 0.300, p = .766$, 95% CI (0.543, 0.730).

TBI group moral judgements

The proportion of intuitive judgements was significantly higher than the 0.5 baseline in UI, $t(29) = 3.137, p = .004$, 95% CI (0.044, 0.209), but not DI, $t(29) = 0.377, p = .709$, 95% CI (−0.089, 0.129) dilemmas (see Figure 1C). The TBI group showed no significant preference for utilitarian (or deontological) judgements when all dilemmas were pooled and compared against the 0.5 baseline, $t(29) = 1.306, p = .202$, 95% CI (−0.028, 0.128).

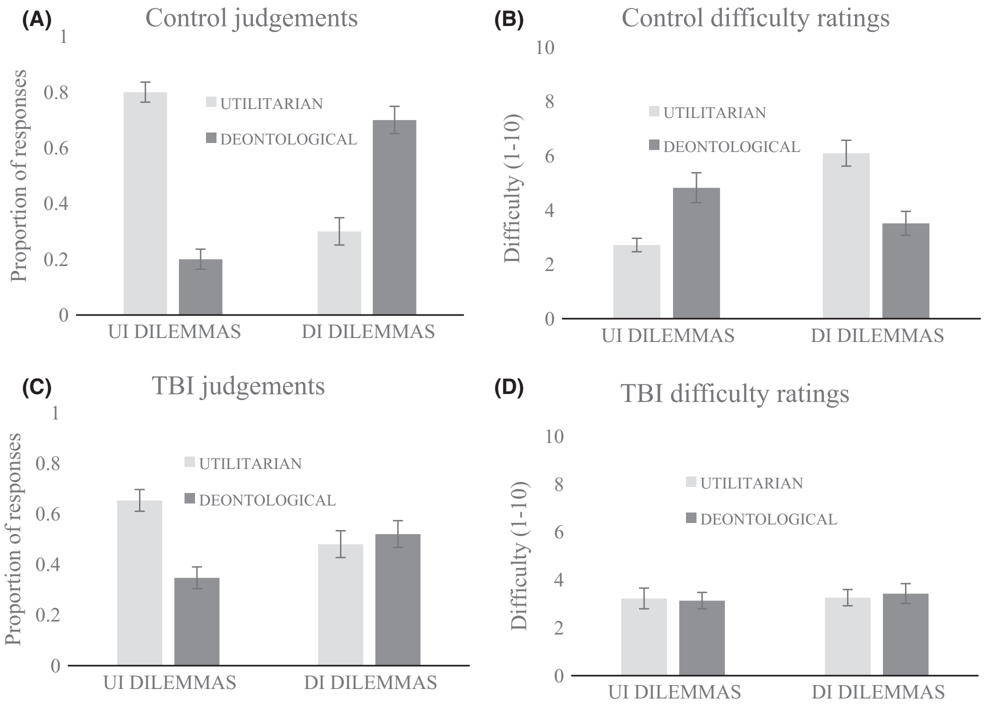

Figure 1. Judgement and difficulty rating data for traumatic brain injury (TBI) and controls individually. (A) Average proportion of utilitarian and deontological responses in the control group, in dilemmas where the utilitarian option is intuitive (UI) and where the deontological option is intuitive (DI). (B) Average difficulty ratings of utilitarian and deontological responses in the control group, in UI and DI dilemmas. (C) Average proportion of utilitarian and deontological responses in the TBI group, in UI and DI dilemmas. (D) Average difficulty ratings of utilitarian and deontological responses in the TBI group, in UI and DI dilemmas. Error bars are standard error of the mean.

TBI group difficulty ratings

In the TBI group, there was no significant difference in the difficulty ratings of intuitive versus counter-intuitive judgements in UI, $t(26) = 0.232$, $p = .818$, 95% CI (−0.703, 0.882), or DI, $t(26) = 0.419$, $p = .679$, 95% CI (−0.669, 1.010), dilemmas (see Figure 1D). Additionally, difficulty ratings did not differ significantly between utilitarian and deontological judgements overall, $t(29) = -0.180$, $p = .858$, 95% CI (−0.644, 0.539).

Between group analyses

Moral judgements

There was a main effect of group on the proportions of counter-intuitive judgements, $F(1, 58) = 19.484$, $p < .001$, with more counter-intuitive judgements in the TBI group, and a main effect of dilemma type, $F(1, 58) = 4.362$, $p = .041$, with more counter-intuitive judgements in DI dilemmas. There was no significant group x dilemma type interaction, $F(1, 58) = 0.005$, $p = .947$.

Group comparisons pooled across both dilemma types (see Figure 2A) indicated that overall the TBI group made a significantly higher proportion of counter-intuitive

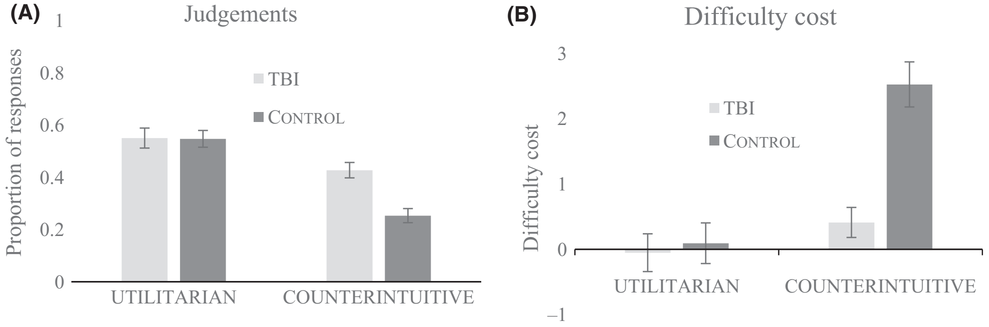

Figure 2. Judgement and difficulty rating data in control and traumatic brain injury (TBI) groups. (A) Average proportions of utilitarian and counter-intuitive judgements in TBI and controls, across all dilemmas. (B) Average difficulty cost of counter-intuitive judgements (over intuitive judgements) and utilitarian judgements (over deontological judgements) separately for TBI and control groups. Error bars are standard error of the mean.

judgements than controls, $t(58) = 4.331$, $p < .001$, 95% CI (0.093, 0.253), but the two groups did not differ significantly in their preference for utilitarian judgements, t $(58) = 0.067$, $p = .947$, 95% CI (−0.097, 0.103).

Difficulty cost data

There was a main effect of group on the difficulty cost of counter-intuitive judgements, with the control group exhibiting higher difficulty costs, $F(1, 33) = 27.065$, $p < .001$. There was no significant effect of dilemma type, $F(1, 33) = 0.364$, $p = .550$, and no significant group x dilemma type interaction, $F(1, 33) = 0.154$, $p = .697$.

Group comparisons pooled across dilemma type revealed that the control and TBI groups differed significantly in the difficulty cost exhibited when they selected the counter-intuitive response, $t(55) = -5.132$, $p < .001$, 95% CI (−2.938, −1.288), with the control group exhibiting a higher mean difficulty cost than the TBI group (see Figure 2B). TBI and control groups did not differ significantly in the difficulty cost associated with utilitarian judgements, $t(58) = -0.342$, $p = .733$, 95% CI (−0.996, 0.705).

Social cognition

The TBI group attained significantly lower scores than controls on cognitive, t $(33.61) = -3.465$, $p = .004$, BCa 95% CI (−0.112, −0.031), and affective, t $(30.31) = -3.360$, $p = .012$, BCa 95% CI (−0.193, −0.051), FP indices, and the RME, $t(58) = -2.097$, $p = .035$, BCa 95% CI (−0.136, −0.011). There were no significant group differences in the tendency towards spontaneous perspective taking, $\chi^2(2, 60) = 0.084$, $p = .959$. As such, no further analyses of this measure were conducted.

Moral judgement and social cognition

Whole sample

Neither the proportion of utilitarian judgements nor the difficulty cost associated with utilitarian decisions significantly correlated with any ToM or IQ variables. However, all IQ

Table 3. Pearson product moment correlations between moral judgement variables and cognitive and social-cognitive variables in the whole sample (n = 57)

	Proportion of utilitarian responses	Cost of utilitarian response	Proportion of counter-intuitive responses	Cost of counter-intuitive response
WASI verbal IQ				
r	.104	.069	−.329*	.269*
p	.441	.609	.013	.043
BCa 95% CI	−0.134, 0.367	−0.123, 0.281	−0.533, −0.042	0.056, 0.451
WASI performance IQ				
r	−.097	.099	−.359**	.154
p	.474	.463	.006	.251
BCa 95% CI	−0.279, 0.110	−0.109, 0.358	−0.654, −0.125	−0.128, 0.483
WASI full-scale IQ				
r	.057	.143	−.482***	.320*
p	.674	.29	.000	.015
BCa 95% CI	−0.175, 0.300	−0.048, 0.345	−0.636, −0.293	0.101, 0.505
Faux Pas Intentions Score				
r	.086	.208	−.357**	.198
p	.526	.121	.006	.139
BCa 95% CI	−0.122, 0.259	−0.077, 0.444	−0.536, −0.151	0.013, 0.407
Faux Pas Belief Score				
r	.026	.214	−.379**	.26
p	.845	.11	.004	.051
BCa 95% CI	−0.250, 0.313	−0.117, 0.486	−0.579, −0.135	0.015, 0.474
RME				
r	−.056	.321*	−.266*	.303*
p	.676	.015	.046	.022
BCa 95% CI	−0.319, 0.230	0.096, 0.527	−0.465, −0.049	0.119, 0.468

Notes. *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$. BCa bootstrap 95% CI's reported.

and ToM variables showed significant, generally moderate, correlations with the proportion of counter-intuitive judgements (see Table 3).

TBI group

The first model in the regression equation, containing affective ToM variables, significantly predicted 20.6% of the variance, $F(2, 27) = 3.492$, $p = .045$. In this model, only the RME contributed uniquely to prediction of counter-intuitive judgements ($\beta = -.520$, $p = .022$). The second model, containing both cognitive and affective ToM variables, accounted for only 3.6% of additional variance in counter-intuitive judgements (R^2 change = .036) and did not attain statistical significance, $F(3, 26) = 2.755$, $p = .063$.

Discussion

Previous research has demonstrated that several neurological patient populations, including TBI, produce utilitarian judgements to moral dilemmas (Ciaramelli *et al.*, 2007; Koenigs *et al.*, 2007; Martins *et al.*, 2012; Mendez *et al.*, 2005) although how best to interpret the data was unclear. The present study adapted moral dilemmas from previous research (Kahane *et al.*, 2011), which allowed the intuitiveness of moral judgement to be controlled for, and applied these new dilemmas to participants with TBI for the first time.

Characterizing moral judgement in TBI

Overall, our TBI participants made similar proportions of utilitarian judgements to controls – but they made substantially more counter-intuitive judgements. On closer analysis, our TBI group did in fact show an atypical preference for utilitarian judgements under limited circumstances; disproportionately selecting utilitarian judgements in extreme moral dilemmas where the utilitarian option was counter-intuitive (i.e., DI dilemmas similar to those used in previous research). However, in more everyday dilemmas where utilitarianism was intuitive (i.e., UI dilemmas), our TBI participants were *less* likely than controls to endorse the utilitarian option, again favouring the counter-intuitive (and incidentally, deontological) response. On this evidence then, TBI causes a generalized bias towards the counter-intuitive option, not a specific bias towards utilitarianism.

These findings support the hypothesis that the distributed neural systems damaged by TBI are not sensitive to the deontological or utilitarian content of a judgement, but rather to how psychologically intuitive these judgements are. They speak directly against the assertion that TBI gives rise to atypically utilitarian judgements (Martins *et al.*, 2012), and cast doubt more broadly on the generalizability of similar conclusions in other neurological populations (e.g., Ciaramelli *et al.*, 2007; Koenigs *et al.*, 2007; Mendez *et al.*, 2005). Such studies may have been biased by the limited range of dilemmas they employed; our TBI participants made more counter-intuitive judgements regardless of utilitarian or deontological content. Previous research has focussed exclusively on extreme dilemmas where a utilitarian response was counter-intuitive. As a consequence, counter-intuitive judgements were able to masquerade as a tendency towards utilitarian judgements.

The generalized pattern of counter-intuitive judgements reported here deviates somewhat from recent evidence that higher psychoticism correlates selectively with increased levels of counter-intuitive utilitarian judgements, but not counter-intuitive deontological judgements (Wiech *et al.*, 2013). However, in the present study, 87% of the TBI group had suffered a severe or very severe TBI. Injuries of this type are known to cause extensive cortical and subcortical pathophysiology resulting in chronic and severe disturbances in executive functions, social cognition, judgement and decision-making, and a host of supportive cognitive functions (Cicerone & Tanenbaum, 1997; Lezak *et al.*, 2012; Mathias & Wheaton, 2007; Newcombe *et al.*, 2011; Rao & Lyketsos, 2000). Given this level of impairment, it is perhaps unsurprising that judgement disturbances were apparent across extreme and more everyday moral dilemmas.

Moral judgement and social cognition in TBI

Neither the TBI nor control group demonstrated a significant difficulty cost when selecting the utilitarian response over the deontological response, supporting previous findings that utilitarian judgements are not more difficult than deontological judgements (Kahane *et al.*, 2011). Our controls exhibited a substantial difficulty cost when making counter-intuitive judgements over intuitive judgements, but the TBI group showed a complete absence of this effect, indicating that they arrived at these counter-intuitive judgements with ease relative to controls. These data support the hypothesis that a strongly reduced aversion to harm underlies counter-intuitive judgements following TBI.

This is consistent with neuroimaging and behavioural evidence which implicates social-cognitive processes in moral judgement (Avramova & Inbar, 2013; Greene *et al.*, 2001). It is striking that our TBI group were able to make counter-intuitive judgements in the complete absence of a difficulty cost, and this is consistent with evidence that VMPFC patients show a total absence of SCR when making counter-intuitive utilitarian judgements involving highly aversive emotional content (Moretto *et al.*, 2010). This absence of a difficulty cost was evident across both DI and UI dilemmas, indicating that aversion to harm is relevant across the spectrum of moral dilemmas. Indeed, although harms were more extreme in DI dilemmas, UI dilemmas still involved significant harms, where negative outcomes included serious social consequences such as the breakdown of a friend's marriage. Nonetheless, the use of objective physiological measures of affect would be beneficial in evaluating this view in future research.

In our whole sample, affective and cognitive ToM correlated moderately with the proportion of counter-intuitive judgements, although general intelligence was the strongest correlate. Affective ToM, as measured by the RME, captured significant variance in counter-intuitive judgements after TBI, but the FP test failed to add significant predictive value to the regression model. As such, our regression model indicates that better performance on the affective ToM task predicts more intuitive moral judgements (and thus, fewer counter-intuitive judgements) following TBI. Such a particular role for affective ToM is consistent with the literature suggesting that intuitive judgements (including deontological judgements in extreme dilemmas) are computed by a reflexive 'system 1' involving visual, pre-motor, and VMPFC activity at the neural level, which is thought to correspond to emotion processing, empathy and affective ToM at the cognitive level (Kahane *et al.*, 2011; Lamm *et al.*, 2007; Nummenmaa *et al.*, 2008; Shamay-Tsoory *et al.*, 2006).

Indeed, affective and cognitive ToM are supported by partially dissociable pre-frontal networks (Shamay-Tsoory & Aharon-Peretz, 2007), with affective ToM relying specifically on the VMPFC, and cognitive ToM recruiting the pre-frontal cortex more broadly. On a somewhat speculative note, this suggests that cognitive ToM may be a more computationally complex process, and as such more likely a higher order, conscious, and deliberative 'system 2' process. As such, its lack of contribution to the prediction of intuitive moral judgements in our study is not surprising. Irrespective of this issue, our findings provide general support for the involvement of socio-cognitive processes and harm aversion in counter-intuitive moral judgement following TBI.

Finally, the combined observations that TBI results in a bias towards counter-intuitive moral judgement, and that these judgements tend to be arrived at with relatively little effort, may go some way to explaining the clinical and familial observations that TBI survivors are often impulsive in their decision-making and make judgements that are hard for others to understand (Bechara & Van Der Linden, 2005). Indeed, when TBI participants responded in a counter-intuitive way, our data indicate that they did so *as though the judgement had come to them intuitively*. This is likely to be disconcerting to others and could certainly contribute to post-injury social and communication difficulties.

Conclusion

Our study presents behavioural evidence that intuitive and counter-intuitive moral judgements are perturbed in TBI, but utilitarian judgements are not. This evidence is in accordance with recent neuroimaging data (Kahane *et al.*, 2011) and indicates that the neural systems involved in moral judgement are sensitive to the properties of psychologically generated intuitions, but not to the tensions between competing normative philosophical doctrines. Our difficulty rating and social cognition data further suggest that atypical moral judgement in TBI is attributable, at least in part, to an impaired ability to mentalize about the emotional experiences of others, and ultimately an absence of emotional aversion to harming others.

These disturbances in moral judgement held across a wide range of dilemmas, including extreme 'killing' scenarios which are unlikely to ever occur to a person, as well as more 'everyday' dilemmas regarding marital infidelity, stealing, and conflict resolution. It is likely that investigation of these everyday dilemmas will show the most promise in enhancing the clinical impact of this research, which has been identified as an objective for the area (Rosas & Koenigs, 2014).

References

Apperly, I. (2010). *Mindreaders: The cognitive basis of "theory of mind"*. East Sussex, UK: Psychology Press.

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 169–178. doi:10.1002/wcs.1216

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251. doi:10.1111/1469-7610.00715

Bechara, A., & Van Der Linden, M. (2005). Decision-making and impulse control after frontal lobe injuries. *Current Opinion in Neurology*, *18*, 734–739. doi:10.1097/01.wco.0000194141.56429.3c

Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience*, *18*, 871–879. doi:10.1162/jocn.2006.18.6.871

Bibby, H., & McDonald, S. (2005). Theory of mind after traumatic brain injury. *Neuropsychologia*, *43*(1), 99–114. doi:10.1016/j.neuropsychologia.2004.04.027

Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, *6*, 580–592.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92. doi:10.1093/scan/nsm001

Cicerone, K. D., & Tanenbaum, L. N. (1997). Disturbance of social cognition after traumatic orbitofrontal brain injury. *Archives of Clinical Neuropsychology*, *12*(2), 173–188. doi:10.1016/S0887-6177(96)00022-4

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 17433–17438. doi:10.1073/pnas.1009396107

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2–7. doi:10.1037/a0025071

Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marenco, V., Roca, M., & Manes, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience*, *8*, 780–788. doi:10.1093/scan/nss067

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One*, *8*, e60418. doi:10.1371/journal.pone.0060418

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154. doi:10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400. doi:10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108. doi:10.1126/science.1062872

Hayes, J. P., Bigler, E. D., & Verfaellie, M. (2016). Traumatic brain injury as a disorder of brain connectivity. *Journal of the International Neuropsychological Society*, *22*(2), 120–137. doi:10.1017/S1355617715000740

Jones, C. (1979). Glasgow coma scale. *The American Journal of Nursing*, *79*, 1551–1557.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2011). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*, 393–402. doi:10.1093/scan/nsr005

Kant, I. (1959). *Foundation of the metaphysics of morals* (l W. Beck, Trans.). Indianapolis, IN: Bobbs-Merrill (Original work published 1785).

Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, *7*, 708–714. doi:10.1093/scan/nsr048

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–911. doi:10.1038/nature05631

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42–58. doi:10.1162/jocn.2007.19.1.42

Lezak, M. D., Howeison, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford, UK & New York, NY: Oxford University Press.

Lipton, M. L., Gulko, E., Zimmerman, M. E., Friedman, B. W., Kim, M., Gellella, E., . . . Branch, C. A. (2009). Diffusion-tensor imaging implicates prefrontal axonal injury in executive function impairment following very mild traumatic brain injury. *Radiology*, *252*, 816–824. doi:10.1148/radiol.2523081584

Martín-Rodríguez, J. F., & León-Carrión, J. (2010). Theory of mind deficits in patients with acquired brain injury: A quantitative review. *Neuropsychologia*, *48*, 1181–1191. doi:10.1016/j.neuropsychologia.2010.02.009

Martins, A. T., Faísca, L. M., Esteves, F., Muresan, A., & Reis, A. (2012). Atypical moral judgment following traumatic brain injury. *Judgment and Decision Making*, *7*, 478–487.

Mathias, J. L., & Wheaton, P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: A meta-analytic review. *Neuropsychology*, *21*(2), 212–223. doi:10.1037/0894-4105.21.2.212

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*, 193–197. doi:10.1097/01.wnn.0000191292.17964.bb

Mitchell, D. G., Avny, S. B., & Blair, R. J. R. (2006). Divergent patterns of aggressive and neurocognitive characteristics in acquired versus developmental psychopathy. *Neurocase*, *12*, 164–178. doi:10.1080/13554790600611288

Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*, 1888–1899. doi:10.1162/jocn.2009.21367

Müller, J. L., Schuierer, G., Marienhagen, J., Putzhammer, A., & Klein, H. E. (2003). “Acquired psychopathy” and the neurobiology of emotion and violence. *Psychiatrische Praxis*, *30*, S221–S225. doi:10.1055/s-2003-39769

Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J. M., Barat, M., & Joseph, P. A. (2010). Exploring theory of mind after severe traumatic brain injury. *Cortex*, *46*, 1088–1099. doi:10.1016/j.cortex.2009.08.014

Newcombe, V. F., Outtrim, J. G., Chatfield, D. A., Manktelow, A., Hutchinson, P. J., Coles, J. P., . . . Menon, D. K. (2011). Parcellating the neuroanatomical basis of impaired decision-making in traumatic brain injury. *Brain*, *1343*), 759–768. doi:10.1093/brain/awq388

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *NeuroImage*, *43*, 571–580. doi:10.1016/j.neuroimage.2008.08.014

Paxton, J. M., Bruni, T., & Greene, J. D. (2014). Are "counter-intuitive" deontological judgments really counter-intuitive?: An empirical reply to Kahane et al.(2012). *Social Cognitive and Affective Neuroscience*, *9*, 1368–1371. doi:10.1093/scan/nst102

Pitman, I., Haddlesey, C., Ramos, S. D., Oddy, M., & Fortescue, D. (2014). The association between neuropsychological performance and self-reported traumatic brain injury in a sample of adult male prisoners in the UK. *Neuropsychological Rehabilitation*, *25*, 763–779. doi:10.1080/09602011.2014.973887

Rao, V., & Lyketsos, C. (2000). Neuropsychiatric sequelae of traumatic brain injury. *Psychosomatics*, *41*(2), 95–103. doi:10.1176/appi.psy.41.2.95

Rosas, A., & Koenigs, M. (2014). Beyond "utilitarianism": Maximizing the clinical impact of moral judgment research. *Social Neuroscience*, *9*, 661–667. doi:10.1080/17470919.2014.937506

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*, 3054–3067. doi:10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Tibi-Elhanany, Y., & Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. *Social Neuroscience*, *1*(3–4), 149–166. doi:10.1080/17470910600985589

Siegal, M., & Varley, R. (2002). Neural systems involved in 'theory of mind'. *Nature Reviews Neuroscience*, *3*, 463–471. doi:10.1038/nrn844

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656. doi:10.1162/089892998562942

Stuss, D. T. (2011). Traumatic brain injury: Relation to executive dysfunction and the frontal lobes. *Current Opinion in Neurology*, *24*, 584–589. doi:10.1097/WCO.0b013e32834c7eb9

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124–129. doi:10.1016/j.cognition.2008.10.008

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonia, TX: Harcourt Assessment.

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition*, *126*, 364–372. doi:10.1016/j.cognition.2012.11.002

World Medical Association (2013). World Medical Association Declaration of Helsinki ethical principles for medical research involving human subjects. *Journal of American Medical Association*, *310*(20), 2191–2194.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*, 361–370. doi:10.1111/j.1600-0447.1983.tb09716.x

Received 14 September 2016; revised version received 6 January 2017

Supporting Information

The following supporting information may be found in the online edition of the article:

Table S1. Moral dilemma pilot data.

Appendix S1. Moral dilemma illustrations.