PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES NEUROSCIENCE

Breakdown of utilitarian moral judgement after basolateral amygdala damage

Jack van Honk[a,b,c,1,2], David Terburg[a,c,2], Estrella R. Montoya[c], Jordan Grafman[d,e], Dan J. Stein[f], and Barak Morgan[g,h]

Edited by Walter Sinnott-Armstrong, Duke University, Durham, NC; received October 19, 2021; accepted June 7, 2022 by Editorial Board Member Michael S. Gazzaniga

Most of us would regard killing another person as morally wrong, but when the death of one saves multiple others, it can be morally permitted. According to a prominent computational dual-systems framework, in these life-and-death dilemmas, deontological (nonsacrificial) moral judgments stem from a model-free algorithm that emphasizes the intrinsic value of the sacrificial action, while utilitarian (sacrificial) moral judgments are derived from a model-based algorithm that emphasizes the outcome of the sacrificial action. Rodent decision-making research suggests that the model-based algorithm depends on the basolateral amygdala (BLA), but these findings have not yet been translated to human moral decision-making. Here, in five humans with selective, bilateral BLA damage, we show a breakdown of utilitarian sacrificial moral judgments, pointing at deficient model-based moral decision-making. Across an established set of moral dilemmas, healthy controls frequently sacrifice one person to save numerous others, but BLA-damaged humans withhold such sacrificial judgments even at the cost of thousands of lives. Our translational research confirms a neurocomputational hypothesis drawn from rodent decision-making research by indicating that the model-based algorithm which underlies outcome-based, utilitarian moral judgements in humans critically depends on the BLA.

moral judgement | basolateral amygdala | social decision-making | brain lesion | computational framework

Most of us would regard killing another person as morally wrong, particularly when this person is innocent of any wrongdoing. However, this intrinsic moral rule is rapidly breached when the death of one saves a number of others (1–4). In life-and-death moral dilemmas, the deontological principle of what is morally wrong conflicts with the utilitarian principle of maximizing outcomes (1, 2, 5). Importantly, decisions in life-and-death moral dilemmas, although hypothetical, are consistent with sacrificial moral decisions made in healthcare and warfare (6, 7).

In the classic trolley car dilemma, people are asked if they would flip a switch to make a runaway trolley that is rapidly approaching a fork in the tracks divert its direction. By performing this action, they kill an unfamiliar innocent person but save the lives of five others. Notwithstanding the collateral damage in terms of ending an innocent person's life, most people decide to flip the switch (8). Significantly fewer people opt for sacrifice in a footbridge variant of the trolley dilemma, in which the person needs to be pushed in front of the trolley. Not only direct physical contact but also the fact that the victim is used as an instrument or means to stop the train increases the emotional conflict in the push version of the trolley car dilemma (9–11). Nonetheless, in these sacrificial moral dilemmas, the inclination to sacrifice tends to rise conditionally upon the number of lives saved; thus, intrinsic moral rules can be rendered powerless when kill:save ratios decrease (12).

The principal theory of moral judgment is the dual-process model (DPM). This model of Greene and coworkers proposes two competing neural systems in the moral brain: an intuitive-emotional system and a controlled-cognitive system (13). Importantly, however, the utilitarian decision to sacrifice and the deontological decision to refrain from action are value-based social decisions that incorporate punishment and reward and are therefore inherently affect driven (9, 12). Interestingly, emphasizing reinforcement-learning models, research and theory in translational and computational neuroscience have recently advanced our understanding of such value-based learning and choice behaviors (14). According to this domain-general framework, organisms learn the value of actions and outcomes via punishment and reward (15). This framework also proposes two basic algorithms: one is model free, and the other is model based. The model-free (or action-based) algorithm rigidly assigns actions based upon the habitually learned value of the action. The model-based (or outcome-based) algorithm flexibly derives value from a causal model of the changing environment and is instrumentally guided by the expected value of the outcome (15, 17). Extending upon the DPM, this revised twofold algorithmic

Significance

Similar to real-life sacrificial decisions in healthcare and warfare, hypothetical moral dilemmas show that decisions to sacrifice depend on valuation of action (type of harm) vs. outcome (lives saved). Neurocomputational frameworks propose two valuation algorithms: a model-free one focused on action and a model-based one focused on outcome. Rodent research emphasizes that outcome-based decisions depend on the basolateral amygdala (BLA). Here, in humans with selective bilateral BLA damage, we show breakdown in outcome-based sacrificial moral judgements. Across dilemmas, healthy control subjects routinely opt for sacrifice, but BLA-damaged subjects rarely select the sacrificial option, even when thousands of lives can be saved. Our data suggest that value-based decisions to sacrifice another human for "the greater good" critically depend on the BLA.

Author contributions: J.V.H. designed research; J.V.H. and B.M. performed research; D.T., E.R.M., and D.J.S. analyzed data; and J.V.H., D.T., E.R.M., J.G., D.J.S., and B.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. W.S.-A. is a guest editor invited by the Editorial Board.

[1] To whom correspondence may be addressed. Email: jackvanh@gmail.com.

[2] J.V.H. and D.T. contributed equally to this work.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2119072119/-/DCSupplemental.

Published July 25, 2022.

system accounts for the fact that all moral decisions are value based and thus affect driven. Crucially, it predicts that sacrificial moral dilemmas, which pit intrinsic value vs. outcome (type of harm vs. lives saved), involve the strongest value-based conflict (16, 17). This intense conflict explains why sacrificial moral scenarios are highly effective in eliciting dissociations at both behavioral and brain levels (18–22).

Human lesion, neuroimaging, and intracranial electroencephalography studies have implicated the amygdala, the nucleus accumbens (NAc), and most prominently the ventral medial prefrontal cortex (vmPFC) as key structures in the neural network of moral decision-making (18, 19, 23–26). The guiding research model in human neuroscience is based upon seminal data from subjects with vmPFC lesions showing abnormally increased utilitarian moral judgments (25, 27, 28). This vmPFC-centered model fruitfully guided neuroimaging research (18, 19) and seems to hold promise for explaining increased utilitarian judgements seen across psychopathology (29–31). The vmPFC is considered to hold vital integrative-executive properties that are required for action-based, nonutilitarian, or deontological moral judgements (17, 18, 25).

Problematically, however, the neural mechanisms underlying outcome-based, utilitarian moral judgements are poorly understood (32). Abnormal decreases in utilitarian moral judgments are rarely observed in subjects with restricted brain damage and to our knowledge are absent in psychopathology. Notably, there is conflicting evidence for a role of the hippocampus in nonutilitarian moral judgements. First, while patients with brain lesions involving the hippocampus showed no abnormalities in moral judgements, a minority of these patients' utilitarian moral judgments were substantially increased (24). Contrariwise, a group of patients with more selective hippocampal lesions showed decreased utilitarian judgements (33). Further complicating matters, in patients with brain-volume reductions in both hippocampus and unilateral amygdala, decreased utilitarian moral judgements were observed, but a patient with hippocampal plus bilateral amygdala volume damage showed increased utilitarian judgements (34). Although these data do not provide definitive answers with respect to the role of the hippocampus in moral judgment, they add to recent evidence for a role of the hippocampus in social behavior (35). The hippocampus, however, often acts in synchrony with, and its gene expression and plasticity are regulated by, the basolateral subregion of the amygdala (36, 37). Thus, questions arise with respect to the exact role of the human amygdala in moral judgment.

The amygdala conveys a major translational obstacle with respect to the cross-species applicability of theories of value-based learning and choice. That is, the human amygdala is (whether unilateral or bilateral) typically researched and discussed as a single unit despite the fact that the mammalian amygdala consists of subregions different in structure and function (38). Most prominently, the basolateral amygdala (BLA) and the central-medial amygdala (CMA) consist of cortical-type and striatal-type neural structures, respectively (39). Rodent research has determined that by parallel actions on the NAc, the CMA subserves habitual (or action-based) choice behavior, while the BLA subserves goal-directed (or outcome-based) choice behavior (40, 41). Reconceptualized in terms of the computational dual-systems model, the CMA subserves the model-free algorithm, while the BLA subserves the model-based algorithm (14, 15). Furthermore, the BLA is the regulating hub: it not only acts on the NAc (42) (triggering the model-based algorithmic system) but also regulates inhibitory control over both the CMA (43, 44) and the vmPFC (45, 46) (controlling the model-free algorithmic system). Crucially, the vmPFC has no access to (and cannot learn and decide upon) the motivational value of outcomes without BLA input (45). It should be noted that rodent research mostly targets the orbital and medial regions of the medial prefrontal cortex (OMPFC), but the rodent OMPFC overlaps structurally and functionally with the human vmPFC (45, 47–51). For translational purposes, we use the term vmPFC throughout this paper.

Altogether, if we use these rodent data as a foundation for helping to explain human morality, the human BLA should govern the model-based algorithm and underlie outcome-based choice behaviors and therefore utilitarian sacrificial moral judgements. Fig. 1 gives our translational neural framework of moral judgement.

We tested this hypothesis in a group of South African subjects with Urbach-Wiethe disease (UWD), a genetic disorder caused by mutation of the extracellular matrix one (EMC1) gene. Previously, we showed that the variant of the ECM1 mutation found in South Africa (the Q276X mutation in exon 7) can produce selective bilateral BLA calcification while leaving CMA fully intact and functional (52–54). The selectivity of bilateral BLA damage seen in these South African UWD subjects is, to the best of our knowledge, unparalleled in human lesion research and provides the unique opportunity to translate detailed rodent amygdala models to the human case (52, 54). Indeed, in neuroeconomic research with these UWD subjects on tasks requiring social learning and decision-making, we successfully translated BLA rodent models to humans (55). Furthermore, we provided cross-species evidence for an evolutionary-conserved role of the BLA in escape behavior by studying both BLA-damaged subjects and rats with a chemogenetically silenced BLA (44).

For the present study, we first confirmed focal BLA damage in five UWD subjects using structural neuroimaging (Fig. 1 and *SI Appendix*). Our UWD subject sample furthermore had a normal IQ and no psychopathology. All UWD and control subjects were recruited from the Namaqualand area in South Africa. Testing took place in the rural Namaqualand area with a local experimenter who spoke the same Afrikaans dialect as the subjects. They were tested on moral decision-making, and their performance was compared with a group of 11 neurologically normal, healthy controls (HCs) matched on age, IQ, socioeconomic status, ethnicity/demography, and religion (*SI Appendix*). BLA-damaged and HC subjects made decisions on a range of moral dilemmas wherein exclusively a utilitarian or a nonutilitarian decision was permitted. Key for this discerning decision is the choice of whether to sacrifice an innocent person to save the lives of others. We used the set of nonmoral and moral scenarios from the seminal paper by Koenigs and coworkers (28), who reported abnormally increased utilitarian moral judgments in humans with medial prefrontal cortex (PFC) damage, which in all cases

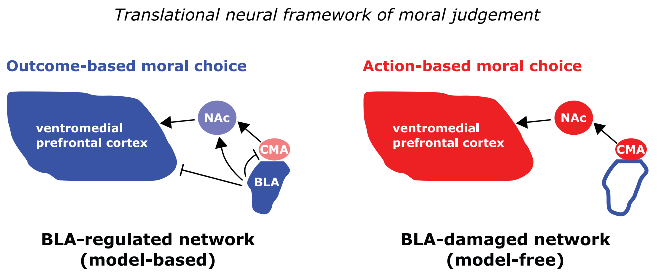
Translational neural framework of moral judgement
Outcome-based moral choice
ventromedial prefrontal cortex
NAc
CMA
BLA
BLA-regulated network (model-based)
Action-based moral choice
ventromedial prefrontal cortex
NAc
CMA
BLA-damaged network (model-free)

Fig. 1. This rodent-human translational framework predicts breakdown of outcome-based, utilitarian moral judgements in BLA-damaged subjects. This breakdown of model-based choice behavior is caused by loss of regulatory action of the BLA on the NAc and loss of the BLA's inhibitory control of the CMA and the vmPFC (41–46).

involved the vmPFC. We hypothesized that humans with selective bilateral BLA damage would show decreased outcome-based, utilitarian moral judgments (Fig. 1).

The moral judgment task is part of a set developed by Greene and colleagues (13) and consists of 50 dilemmas, of which 18 are nonmoral dilemmas, 21 are personal moral dilemmas, and 11 are impersonal moral dilemmas. In personal moral dilemmas, proposed actions include harm or sacrifice through direct physical contact (e.g., pushing someone), whereas in impersonal dilemmas, this harm is done indirectly (e.g., flipping a switch). In the set we used, subjects were repeatedly asked over 15 trials if they would sacrifice someone to save others by either personal or impersonal action. The task was self-paced, and subjects could always receive an explanation from the experimenter if necessary. The dilemmas were displayed on a computer screen: the first screen consisted of a short introduction to the scenario, the second screen consisted of the dilemma, and on the final screen, subjects were asked if they would endorse the proposed action. Specifically, the question was “Would you [action] in order to [gain of the action]?” and they could respond to it with “yes” or “no.”

Results

We first obtained high-resolution transverse relaxation time (T2) weighted magnetic resonance images (MRIs) of each of the five UWD subjects included in this study. Using an MRI probability-mapping method described by Eickhoff and colleagues (56), we were able to quantify the overlap of each calcification with cytoarchitectonic structure-probability maps of the amygdala subregions developed by Amunts and colleagues (57). In line with our previous findings in this group (44, 53, 55, 58), in each of the five UWD subjects, we found bilateral calcifications that were localized to the BLA without affecting other amygdala subregions (Fig. 2).

We evaluated our hypotheses using binary-logistic generalized estimating equation modeling with a robust estimator of the covariance matrix, a working correlation matrix with an exchangeable structure, and Bonferroni-corrected post hoc comparisons (data for the individual dilemmas can be found in *SI Appendix*, Fig. S1).

We first compared the BLA-damaged subjects and controls on their decisions for moral compared to nonmoral dilemmas (*SI Appendix*, Model 1), and as expected, both groups gave fewer yes responses to moral dilemmas (Wald $X^2 = 76.4$, $P < 0.001$). Crucially, grouping by dilemma-type interaction (Wald $X^2 = 32.8$, $P < 0.001$) indicated that UWD subjects gave significantly fewer yes responses to moral dilemmas than controls (estimated marginal mean difference (EMM-diff) = −27%, confidence interval (CI) = ±9%, $P < 0.001$), while their decisions were similar on nonmoral scenarios (EMM-diff = 3%, CI = ±11%, $P = 0.451$; Fig. 3*A*). Thus, BLA-damaged subjects were significantly less likely to approve of harmful actions compared to controls.

Decisions in moral dilemmas generally are modulated by the personal-impersonal action factor (direct vs. indirect harm), with disapproving of causing harm when there is direct personal action (9–11). Furthermore, the model-based/model-free algorithmic system should be most conflicted in sacrificial moral dilemmas that pit value vs. outcome (type of harm vs. lives saved) (16, 17). Accordingly, we investigated whether the breakdown of utilitarian moral judgement is most pronounced in the dilemmas with direct personal action and those that involve sacrifice of life. We compared the decisions to all moral dilemmas of BLA-damaged subjects and controls with factors representing personal vs. impersonal dilemmas and sacrificial vs. nonsacrificial dilemmas (*SI Appendix*, Model 2).

The direct personal factor (grouped by dilemma-type interaction, Wald $X^2 = 0.1$, $P = 0.736$; Fig. 3*B*) did not reveal a group difference, but the sacrifice factor did (grouped by dilemma-type interaction, Wald $X^2 = 20.0$, $P < 0.001$; Fig. 3*C*). Compared to controls, BLA-damaged subjects gave fewer yes responses to impersonal (EMM-diff = −39%, CI = ±13%, $P < 0.001$) and personal (EMM-diff = −28%, CI = ±11%, $P < 0.001$) dilemmas. In contrast, compared to controls, the BLA-damaged subjects did not give fewer yes responses to nonsacrificial dilemmas (EMM-diff = −8%, CI = ±11%, $P = 0.340$), but they did give fewer yes (that is, utilitarian) responses to sacrificial dilemmas (EMM-diff = −61%, CI = ±15%, $P < 0.001$). Note that a yes judgement in these sacrificial dilemmas is utilitarian because the harmful sacrificial action saves more lives.

Altogether, BLA-damaged compared to control subjects show significant reductions in utilitarian moral judgements and most dramatically in sacrificial moral dilemmas such as the trolley car dilemma. Crucially, this is not caused by any intellectual impairments in BLA-damaged subjects, as their IQs perfectly matched those of the controls. Moreover, we confirmed their ability to understand the inherent properties and implications of their moral judgements (*SI Appendix*), and in earlier socioeconomic research, we established that they can also understand more complex probabilities (55, 58).

Finally, we investigated whether this breakdown of utilitarian moral judgement in sacrificial dilemmas depends on egoistic concerns (i.e., saving one’s own life). When comparing the sacrificial dilemmas on this factor (*SI Appendix*, Model 3), we find that egoistic concern did not interact with the group difference in yes responses (Wald $X^2 = 0.2$, $P = 0.637$). Based upon the conventional high- vs. low-conflict ordering of personal scenarios (28), we also investigated whether the breakdown of utilitarian judgment in the personal dilemmas depends on the high- vs. low-conflict order (*SI Appendix*, Model 4). Yes responses indeed are more frequently made in the low-conflict scenarios (Wald $X^2 = 3.8$, $P = 0.051$), but there is no statistical interaction with groups (BLA-damaged vs. HC subjects, Wald $X^2 = 0.4$, $P = 0.551$).

Discussion

In a group of humans with bilateral damage restricted to the BLA and an intact and functional CMA compared to otherwise closely matched HC subjects, we show highly significant reductions in utilitarian moral judgments. Moreover, this effect was driven by a dramatic breakdown of utilitarian judgement in scenarios that involve sacrifice of human life. BLA-damaged subjects were unwilling to sacrifice innocent individuals to save multiple others, irrespective of a) whether the sacrificial action was indirect or direct (pushing a handle or pushing a person), b) egoistic concerns (saving one’s own life), and c) kill:save ratios (the relative number of lives saved). As can be seen in Fig. 3, HC subjects routinely sacrificed innocent individuals to save multiple others, whereas BLA-damaged subjects almost never did, with zero overlap between any of the BLA-damaged and HC subjects.

The difference between BLA-damaged and HC subjects in the standard trolley car dilemma is remarkable. Here, all HC subjects vs. none of the BLA-damaged subjects decided to flip the switch and sacrifice the innocent person to save five others (*SI Appendix*, Fig. S1). The decision seen in our HC subjects to sacrifice abundantly in the standard trolley car dilemma corresponds to the scientific literature (8), whereas the decision by our BLA-damaged subjects never to sacrifice is unprecedented. Moreover, in the

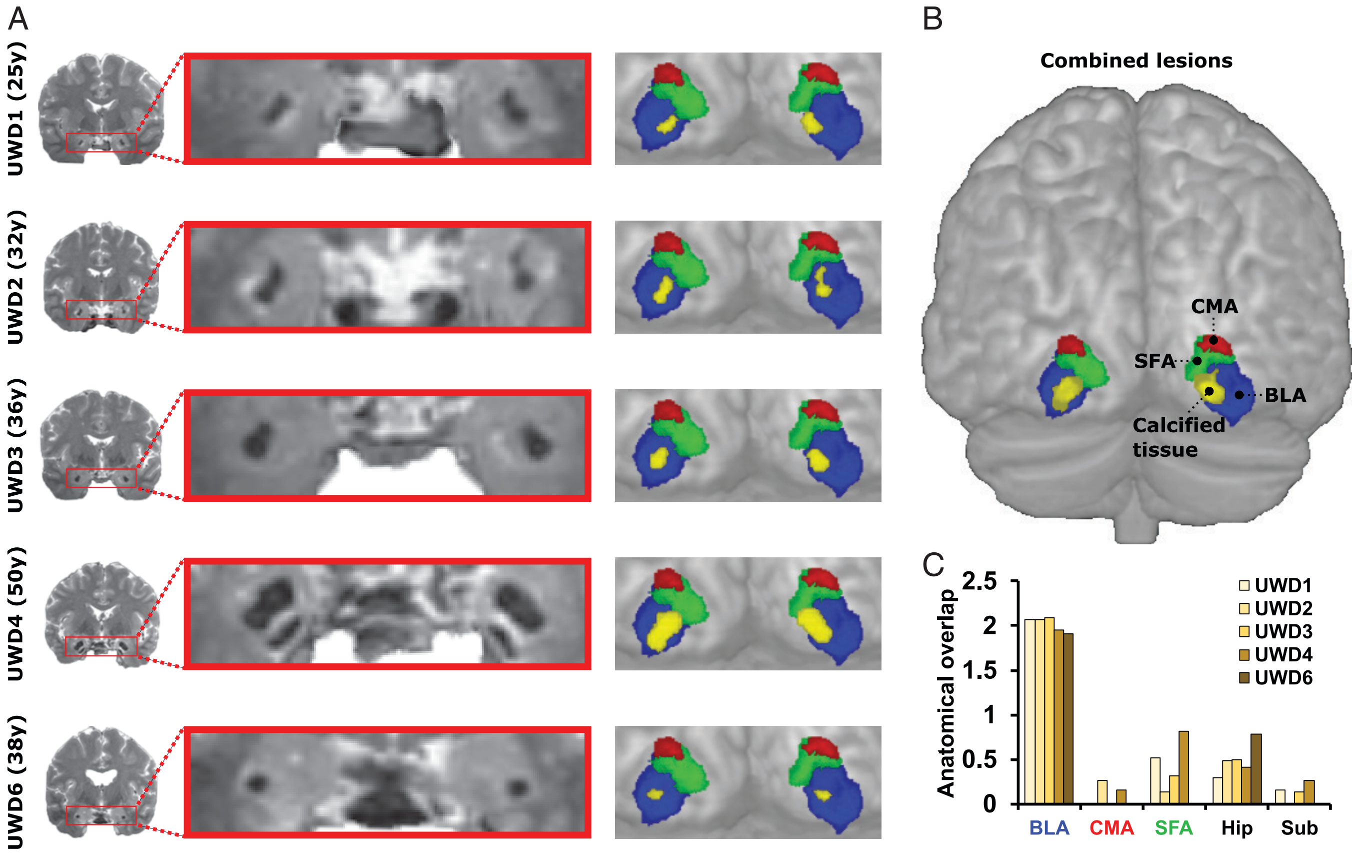

Fig. 2. Calcifications in the BLA-damaged subjects are bilateral and focal to the BLA. (*A*) Coronal slices from each individual's T2-weighted MRI scan, age at time of scanning, and in MNI space estimated lesion volumes plotted within the amygdala subregions' probability maps (voxel defined as voxels with subregion probability > 50%; *SI Appendix*). (*B*) Combined lesions image showing all five lesion volumes together. (*C*) Bar graph representing anatomical overlap quantified using bilateral excess probability (P_{excess}) values (*SI Appendix*) of the lesion volumes, whereby values > 1 indicate a reliable match of volume and anatomical location of the following: BLA, SFA, CMA, Hip = hippocampus, Sub = subiculum.

vaccine test dilemma, wherein one person needs to be sacrificed to save millions, the decisions of BLA-damaged subjects appear similarly irrational: All except one of our control subjects chose to sacrifice one person to save millions of others, but none of our BLA-damaged subjects makes this sacrificial decision (*SI Appendix*, Fig. S1). The vaccine policy dilemma, on the other hand, asks subjects merely to encourage the use of a vaccine to bring immunity from a deadly disease to many people, even though a small number will get ill from the vaccine. In this dilemma, more than 50% of the BLA-damaged subjects opt for the sacrificial decision, which is also the case in control subjects. This dilemma is notable in our sacrificial category for several reasons. First, there is a freedom to choose to take the vaccine or not. Second, the conditions are strongly impersonal, as no identifiable people are intentionally harmed and no physical action is required. Third, the precise numbers of people both sacrificed and saved are unknown. In other words, the action-based algorithm has no affective value to act upon, while the outcome-based algorithm has ambiguous data on action-outcome probabilities. Combined, these factors strongly decrease the emotional conflict of this sacrificial dilemma, which may explain the lack of difference between the decisions of BLA-damaged and those of control subjects.

Interestingly, with respect to the irrational moral judgements above, in these same BLA-damaged subjects, we previously observed irrational socioeconomic decisions (58). In that research, BLA-damaged subjects invested generously (100% more than HC subjects) in unfamiliar others in a trust game without expecting a fair return. Taken together, our socioeconomic and moral judgement data agree with psychological and neuroimaging research showing overlapping neural mechanisms for value-based economic and moral choice (19, 59). This cross-domain breakdown in outcome-based choice behaviors in our BLA-damaged subjects suggests a failure to implement a model-based algorithm: in value-based social decision-making, they effectively act model free.

In support of this notion, a retrospective interview focusing on the trolley car dilemma revealed that the BLA-damaged subjects understood both action and outcome but decided against sacrificing the innocent person because it was too upsetting or (socially) too painful. In their words, the sacrificial action "hurts too much" (*SI Appendix*). This heightened social pain sensitivity, together with an absence of egoistic concerns shown in the current study and in our past socioeconomic research (58), suggests that our BLA-damaged subjects—in their model-free choice behaviors—do not prioritize the self over the other. In support of this conclusion, recent research suggests that the model-free system underlies learning to avoid harming others (relative to the self) (60). Noteworthy in that respect, neurons in the primate BLA can signal the value of rewards either for the self or for others (61). In lacking these neurons, we speculate that our BLA-damaged subjects do not routinely differentiate between self and others in their value-based choice behaviors. This speculation in turn suggests a host of studies to further examine the importance of the BLA in self-consciousness.

In conclusion, we show a breakdown in utilitarian moral judgements in subjects with selective bilateral damage to the BLA. The rigid nonutilitarian moral judgements of our BLA-damaged subjects are fully independent of outcome and therefore ostensibly model free (16, 17). These nonsacrificial moral

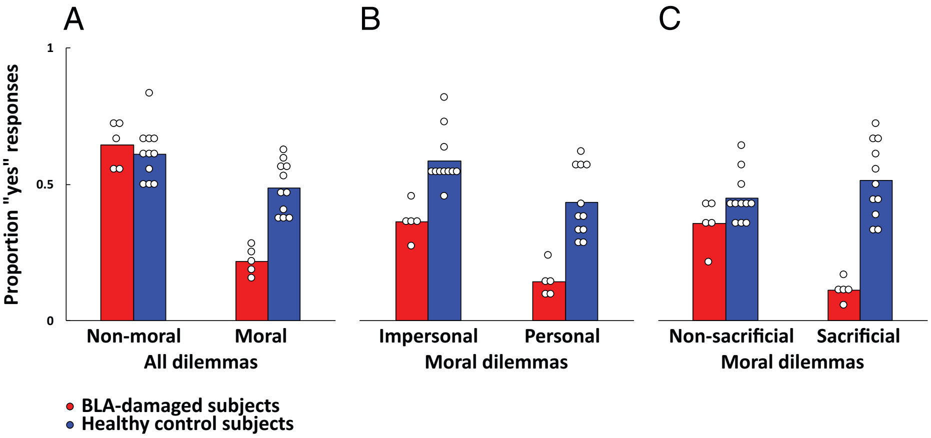

Fig. 3. Individual and average proportions of yes responses. (*A*) Yes responses to all dilemmas, plotted separately for dilemmas that involve nonmoral or moral decisions. (*B*) Yes responses to all dilemmas that involve a moral decision, plotted separately for dilemmas that require impersonal or personal actions. (*C*) Yes responses to all dilemmas that involve a moral decision, plotted separately for dilemmas that require nonsacrificial or sacrificial actions.

judgements cost thousands of lives and are therefore difficult to comprehend. However, typically, the model-free and model-based algorithmic systems operate in synchrony, depending on the environment, and both systems are highly adaptive (15). The current study reveals that in our BLA-damaged subjects synchrony is lost, and their relentless nonsacrificial moral judgments indicate a model-free system operating in isolation (16, 17). According to our translational neural framework (Fig. 1), isolated action of the model-free algorithmic system after BLA damage is caused by a) absence of regulatory action of the BLA on the NAc (the model-based algorithm is not triggered) (41), b) loss of inhibitory control of BLA on the level of the CMA (disinhibition of the model-free algorithm) (43, 44), c) loss of inhibitory control of the BLA on the level of the vmPFC (disinhibition of model-free integration) (46), and d) loss of access of the vmPFC to the BLA-encoded motivational outcome values (model-based algorithm fully disabled) (45).

The human moral brain recruits brain regions and processes other than those comprising the amygdala, including intricate neurobiological mechanisms involving neurochemicals (62, 63) and important brain network hubs such as the anterior cingulate, the insula, and the hippocampus (33, 64–66). Nonetheless, based upon a wealth of rodent and primate research (40, 41, 45, 61, 67), including these and our earlier translational data (44, 54, 55, 58), we propose that the human BLA is the vital model-based regulating neural hub in the network of social decision-making. Indeed, recent neurogenetic cross-species research has revealed that throughout 450 million years of vertebrate evolution, the BLA stands out as a highly conserved neural hub crucial to the network of social decision-making (68). The present data show that in the social decision-making network, the human BLA operates a model-based algorithm and is indispensable for outcome-based moral choice behaviors that lead to sacrificing the life of one person to save many others. For deeper translational and computational insights, further translational research in these BLA-damaged subjects across the domains of value-based learning and decision-making is necessary.

Methods and Materials

Participants. Five subjects with UWD with normal IQ and no psychopathology were compared with a group of 11 neurologically normal HCs matched on age, IQ, socioeconomic status, ethnicity/demography, and religion (*SI Appendix*, Table S1). All subjects provided informed consent before the beginning of the test session. The Health Sciences Faculty Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa, approved the study (HREC 639/2016).

Stimuli. We used the morality scenarios as previously described in Koenigs and colleagues (28). Three scenario categories are included in this set:

- Nonmoral scenarios, which involve decisions without moral judgements
- Impersonal moral scenarios, which involve moral judgements that require indirect and nonphysical actions toward the victim
- Personal moral scenarios, which involve moral judgements that require direct and physical interaction with the victim

Scenarios were translated to Afrikaans and backtranslated to English by separate native speakers. Adjustments were made when the backtranslation was not consistent with the original.

To test our hypothesis with regard to the influence of sacrifice, computability, and egoistic concern, we categorized all moral scenarios based on whether the victim would be at risk for death (sacrificial moral dilemma), whether the kill: save ratio had an economically positive outcome (computable sacrificial dilemma), and whether the decider herself was at risk for death (egoistic concern). *SI Appendix* gives all original English scenarios, their final Afrikaans translation, and their categorization.

Moral Reasoning and Neuropsychological Assessment. All participants live and were tested in the South African Northern Cape mountain-desert area of Namaqualand. Namaqualand is an economically impoverished region, and quality of school education is far below Western norms. Participants were tested in their local environment by a local psychologist using the Wechsler Abbreviated Scale of Intelligence (WASI, which provides a reliable IQ estimate) (69). A local research assistant assisted in testing them on the moral scenarios. The psychologist and the research assistant spoke the Afrikaans dialect which was the first language of the participants. The WASI verbal tests were translated by a local linguist into this Afrikaans dialect spoken in Namaqualand. All subjects scored in the low-normal range for Western standards; however, their scores cannot be compared to Western standards given education and socioeconomic environment, as well as because, for instance, for the synonyms test in WASI has a strong language bias. That is, the Afrikaans language has few synonyms compared to English. Furthermore, of relevance to morality and moral reasoning is the fact that all the participants in the experiment identified their religion as Dutch Reformed.

MRI Analyses Lesions. MRI scans were acquired with a Siemens Magnetom Allegra 3-Tesla head-only scanner at the Cape Universities Brain Imaging Centre in Cape Town, South Africa. For lesion analysis, we obtained whole-brain T2-weighted images with 1-mm isotropic resolution, repetition-time = 3,500 ms, and echo-time = 354 ms.

To estimate extent and anatomical location of the lesions, T2-weighted scans were normalized to Montreal Neurological Institute (MNI) space using unified segmentation, which is optimized for normalization of lesioned brains (70). Lesion volumes were defined using the 3D volume-of-interest feature implemented in

MRIcroN (https://people.cas.sc.edu/rorden/mricron/index.HTML). Based on MRIs, the precise borders between amygdalae and neighboring structures, or between the subregions of the amygdala, cannot be established (57, 71). To determine the precise location of the lesions in our UWD subjects, we therefore assigned the lesion volumes to cytoarchitectonic probability maps according to the method described by Eickhoff and colleagues (56). In this method, which is implemented in the SPM8 anatomy toolbox (https://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox), a volume of interest is superimposed onto a cytoarchitectonic probability map of the medial-temporal lobe (57). This map is based on microscopic analyses of 10 postmortem human brains and follows a generally accepted division of the human amygdala in three subregions. The first is the CMA, which consists of the central and medial nuclei. The second is the BLA, which includes the lateral, basolateral, basomedial, and paralaminar nuclei, and the third is the superficial (or corticoid) amygdala (SFA), which includes the anterior amygdaloid area, amygdala-piriform transition area, amygdaloid-hippocampal area, and cortical nucleus (57). This method assigns to any given voxel a value representing the probability that it belongs to an underlying structure. These are derived from an overlap analysis of 10 postmortem brains and are therefore divided in 10 separate probability classes ranging from 10 to 100% probability.

To estimate how well the lesion volumes fit to the underlying structure, P_{excess} values are computed using the following equation:

$$P_{\mathrm{excess}} = P_{\mathrm{lesion}} / P_{\mathrm{map}},$$

whereby P_{lesion} represents the average cytoarchitectonic probability of the voxels that are shared by the lesion and the cytoarchitectonic probability map and P_{map} represents the average probability of the whole structure's cytoarchitectonic map. These values thus represent how much the average probability of the overlapping voxels exceeds the overall probability distribution of that particular structure and thus indicate whether the lesion overlaps with relatively high or low probability classes of that structure. In other words, P_{excess} represents how central the location of the lesion is relative to that structure's cytoarchitectonic map, whereby $P_{\mathrm{excess}} > 1$ indicates a more central and $P_{\mathrm{excess}} < 1$ indicates a more peripheral location (56).

Short Qualitative Retrospective Interview. We interviewed the BLA-damaged subjects and the controls, focusing on the trolley car scenario. We asked if they understood the consequence of their decision and why they made the decision to sacrifice or not. All the controls understood the consequences of their decision, and all made the decision to sacrifice the one because of the five of lives that would be saved. The BLA-damaged subjects also understood what the consequences of their decision would be. However, all but one of the BLA-damaged subjects answered that they nonetheless could not make the sacrificial decision because it was distressing, upsetting, or painful; that is, it "hurts too much." The remaining BLA-damaged subject was not able to give an explanation for her decision but said it felt like the best thing to do. Although these qualitative data should be considered with some caution, they support the hypothesis that bilateral damage to the BLA in our subjects impairs their ability to apply the model-based algorithm and make outcome-based decisions in value-based moral decision-making.

Data Availability. Data that can be linked to the age of the participants (MRI scans) are available upon request. All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by Hope For Depression Research Foundation (RGA 9-015) and the Netherlands Society of Scientific Research (Brain and Cognition 056-24-010).

Author affiliations: [a]Department of Psychiatry, University of Cape Town, Cape Town, 7925, South Africa; [b]Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, 7925, South Africa; [c]Department of Psychology, Utrecht University, Utrecht, 3584CS, The Netherlands; [d]Feinberg School of Medicine, Departments of Physical Medicine, Rehabilitation, Neurology, Psychiatry, and Behavioral Sciences, Northwestern University, Evanston, IL 60208; [e]Shirley Ryan AbilityLab, Chicago, IL 60611; [f]Medical Research Council Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, 7925, South Africa; [g]Institute for Safety Governance and Criminology, Law Faculty, University of Cape Town, Cape Town, 7700, South Africa; and [h]Centre of Excellence in Human Development, University of Witwatersrand, Johannesburg, 2193, South Africa

1. P. Conway, B. Gawronski, Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *J. Pers. Soc. Psychol.* **104**, 216–235 (2013).
2. S. Nichols, R. Mallon, Moral dilemmas and moral rules. *Cognition* **100**, 530–542 (2006).
3. C. D. Navarrete, M. M. McDonald, M. L. Mott, B. Asher, Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion* **12**, 364–370 (2012).
4. J. Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (Atlantic Press, 2013).
5. J. M. Paxton, J. D. Greene, Moral reasoning: Hints and allegations. *Top. Cogn. Sci.* **2**, 511–527 (2010).
6. G. J. Annas, Military medical ethics–physician first, last, always. *N. Engl. J. Med.* **359**, 1087–1090 (2008).
7. E. J. Johnson, D. Goldstein, Do defaults save lives? *Science* **302**, 1338–1339 (2003).
8. A. Bleske-Rechek, L. A. Nelson, J. P. Baker, M. W. Remiker, S. J. Brandt, Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner. *J. Soc. Evol. Cult. Psychol.* **4**, 115–127 (2010).
9. F. Cushman, L. Young, M. Hauser, The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychol. Sci.* **17**, 1082–1089 (2006).
10. J. D. Greene *et al.*, Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* **111**, 364–371 (2009).
11. B. Bago *et al.*, Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. *Nat. Hum. Behav.* 10.1038/s41562-022-01319-5. (2022).
12. B. Trémolière, J.-F. Bonnefon, Efficient Kill-Save Ratios Ease Up the Cognitive Demands on Counterintuitive Moral Utilitarianism. *Pers. Soc. Psychol. Bull.* **40**, 923–930 (2014).
13. J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, J. D. Cohen, An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108 (2001).
14. A. Olsson, E. Knapska, B. Lindström, The neural and computational systems of social learning. *Nat. Rev. Neurosci.* **21**, 197–212 (2020).
15. N. Drummond, Y. Niv, Model-based decision making and model-free learning. *Curr. Biol.* **30**, R860–R865 (2020).
16. F. Cushman, Action, outcome, and value: A dual-system framework for morality. *Pers. Soc. Psychol. Rev.* **17**, 273–292 (2013).
17. M. J. Crockett, Models of morality. *Trends Cogn. Sci.* **17**, 363–366 (2013).
18. A. Shenhav, J. D. Greene, Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *J. Neurosci.* **34**, 4741–4749 (2014).
19. A. Shenhav, J. D. Greene, Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron* **67**, 667–677 (2010).
20. B. Pastötter, S. Gleixner, T. Neuhauser, K.-H. T. Bäuml, To push or not to push? Affective influences on moral judgment depend on decision frame. *Cognition* **126**, 373–377 (2013).
21. M. J. Crockett, L. Clark, M. D. Hauser, T. W. Robbins, Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17433–17438 (2010).
22. E. Awad, S. Dsouza, A. Shariff, I. Rahwan, J.-F. Bonnefon, Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2332–2337 (2020).
23. J. D. Greene, L. E. Nystrom, A. D. Engell, J. M. Darley, J. D. Cohen, The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).
24. C. F. Craver *et al.*, Moral judgment in episodic amnesia. *Hippocampus* **26**, 975–979 (2016).
25. B. C. Taber-Thomas *et al.*, Arrested development: Early prefrontal lesions impair the maturation of moral judgement. *Brain* **137**, 1254–1261 (2014).
26. E. Hesse *et al.*, Early detection of intentional harm in the human amygdala. *Brain* **139**, 54–61 (2016).
27. B. C. Thomas, K. E. Croft, D. Tranel, Harming kin to save strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *J. Cogn. Neurosci.* **23**, 2186–2196 (2011).
28. M. Koenigs *et al.*, Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* **446**, 908–911 (2007).
29. S. H. Kim *et al.*, Manic patients exhibit more utilitarian moral judgments in comparison with euthymic bipolar and healthy persons. *Compr. Psychiatry* **58**, 37–44 (2015).
30. M. Koenigs, M. Kruepke, J. Zeier, J. P. Newman, Utilitarian moral judgment in psychopathy. *Soc. Cogn. Affect. Neurosci.* **7**, 708–714 (2012).
31. L. Khemiri, J. Guterstam, J. Franck, N. Jayaram-Lindström, Alcohol dependence associated with increased utilitarian moral judgment: A case control study. *PLoS One* **7**, e39882 (2012).
32. B. Garrigan, A. L. R. Adlam, P. E. Langdon, The neural correlates of moral decision-making: A systematic review and meta-analysis of moral evaluations and response decision judgements. *Brain Cogn.* **108**, 88–97 (2016).
33. C. McCormick, C. R. Rosenthal, T. D. Miller, E. A. Maguire, Hippocampal damage increases deontological responses during moral decision making. *J. Neurosci.* **36**, 12157–12167 (2016).
34. M. Verfaellie, R. Hunsberger, M. M. Keane, Episodic processes in moral decisions: Evidence from medial temporal lobe amnesia. *Hippocampus* **31**, 569–579 (2021).
35. A. Montagrin, C. Saiote, D. Schiller, The social hippocampus. *Hippocampus* **28**, 672–679 (2018).
36. A. C. Felix-Ortiz, K. M. Tye, Amygdala inputs to the ventral hippocampus bidirectionally modulate social behavior. *J. Neurosci.* **34**, 586–595 (2014).
37. Y. Yang, J. Z. Wang, From structure to behavior in basolateral amygdala-hippocampus circuits. *Front. Neural Circuits* **11**, 86 (2017).
38. T. Hennessey, E. Andari, D. G. Rainnie, RDoC-based categorization of amygdala functions and its implications in autism. *Neurosci. Biobehav. Rev.* **90**, 115–129 (2018).
39. P. H. Janak, K. M. Tye, From circuits to behaviour in the amygdala. *Nature* **517**, 284–292 (2015).
40. A. G. Phillips, S. Ahn, J. G. Howland, Amygdalar control of the mesocorticolimbic dopamine system: Parallel pathways to motivated behavior. *Neurosci. Biobehav. Rev.* **27**, 543–554 (2003).
41. B. W. Balleine, S. Killcross, Parallel incentive processing: An integrated view of amygdala function. *Trends Neurosci.* **29**, 272–279 (2006).
42. B. M. Sharp, Basolateral amygdala and stress-induced hyperexcitability affect motivated behaviors and addiction. *Transl. Psychiatry* **7**, e1194 (2017).
43. P. Blaesse *et al.*, μ-Opioid receptor-mediated inhibition of intercalated neurons and effect on synaptic transmission to the central amygdala. *J. Neurosci.* **35**, 7317–7325 (2015).
44. D. Terburg *et al.*, The basolateral amygdala is essential for rapid escape: A human and rodent study. *Cell* **175**, 723–735.e16 (2018).

45. G. Schoenbaum, B. Setlow, M. P. Saddoris, M. Gallagher, Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867 (2003).
46. J. Dilgen, H. A. Tejeda, P. O'Donnell, Amygdala inputs drive feedforward inhibition in the medial prefrontal cortex. *J. Neurophysiol.* **110**, 221–229 (2013).
47. D. H. Zald, C. Andreotti, Neuropsychological assessment of the orbital and ventromedial prefrontal cortex. *Neuropsychologia* **48**, 3377–3391 (2010).
48. D. Ongür, J. L. Price, The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).
49. M. L. Gross, Military medical ethics. *Camb. Q. Healthc. Ethics* **22**, 92–109 (2013).
50. J. Hiser, M. Koenigs, The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biol. Psychiatry* **83**, 638–647 (2018).
51. C. C. Ruff, E. Fehr, The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* **15**, 549–562 (2014).
52. N. Koen *et al.*, Translational neuroscience of basolateral amygdala lesions: Studies of Urbach-Wiethe disease. *J. Neurosci. Res.* **94**, 504–512 (2016).
53. D. Terburg *et al.*, Hypervigilance for fear after basolateral amygdala damage in humans. *Transl. Psychiatry* **2**, e115 (2012).
54. J. van Honk, D. Terburg, H. Thornton, D. Stein, B. Morgan, "Consquences of bilateral lesions to the basolateral amygdala in humans" in *Living Without an Amygdala*, D. Amaral, R. Adolphs, Eds. (Guilford Press, 2016), pp. 334–363.
55. L. A. Rosenberger *et al.*, The human basolateral amygdala is indispensable for social experiential learning. *Curr. Biol.* **29**, 3532–3537.e3 (2019).
56. S. B. Eickhoff *et al.*, A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325–1335 (2005).
57. K. Amunts *et al.*, Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat. Embryol. (Berl.)* **210**, 343–352 (2005).
58. J. van Honk, C. Eisenegger, D. Terburg, D. J. Stein, B. Morgan, Generous economic investments after basolateral amygdala damage. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2506–2510 (2013).
59. D. J. Cohen, A. R. Cromley, K. E. Freda, M. White, Psychological value theory: The psychological value of human lives and economic goods. *J. Exp. Psychol. Learn. Mem. Cogn.* 10.1037/xlm0001047. (2021).
60. P. L. Lockwood, M. C. Klein-Flügge, A. Abdurahman, M. J. Crockett, Model-free decision making is prioritized when learning to avoid harming others. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 27719–27730 (2020).
61. S. W. C. Chang *et al.*, Neural mechanisms of social decision-making in the primate amygdala. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 16012–16017 (2015).
62. S. M. Brannon, S. Carr, E. S. Jin, R. A. Josephs, B. Gawronski, Exogenous testosterone increases sensitivity to moral norms in moral dilemma judgements. *Nat. Hum. Behav.* **3**, 856–866 (2019).
63. D. Scheele *et al.*, Opposing effects of oxytocin on moral judgment in males and females. *Hum. Brain Mapp.* **35**, 6067–6076 (2014).
64. M. Stallen *et al.*, Neurobiological mechanisms of responding to injustice. *J. Neurosci.* **38**, 2944–2954 (2018).
65. M. Boccia *et al.*, Neural foundation of human moral reasoning: An ALE meta-analysis about the role of personal perspective. *Brain Imaging Behav.* **11**, 278–292 (2017).
66. J. Moll, R. Zahn, R. de Oliveira-Souza, F. Krueger, J. Grafman, Opinion: The neural basis of human moral cognition. *Nat. Rev. Neurosci.* **6**, 799–809 (2005).
67. R. L. Jenison, A. Rangel, H. Oya, H. Kawasaki, M. A. Howard, Value encoding in single neurons in the human amygdala during decision making. *J. Neurosci.* **31**, 331–338 (2011).
68. L. A. O'Connell, H. A. Hofmann, Evolution of a vertebrate social decision-making network. *Science* **336**, 1154–1157 (2012).
69. D. Wechsler, *Abbreviated Scale of Intelligence* (Psychological Corporation, 1999).
70. J. Crinion *et al.*, Spatial normalization of lesioned brains: Performance evaluation and impact on fMRI analyses. *Neuroimage* **37**, 866–875 (2007).
71. E. Solano-Castiella *et al.*, Diffusion tensor imaging segments the human amygdala in vivo. *Neuroimage* **49**, 2958–2965 (2010).