The human basolateral amygdala is indispensable for social experiential learning

Lisa A. Rosenberger[1]*, Christoph Eisenegger[1], Michael Naef[2], David Terburg[3,4], Jorique Fourie[4], Dan J. Stein[4], and Jack van Honk[3,4,5]*

Affiliations:

[1]Neuropsychopharmacology & Biopsychology Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, 1010 Vienna, Austria.

[2]Experimental Economics Laboratory, Department of Economics, Royal Holloway, University of London, Egham, TW20 0EX, United Kingdom.

[3]Department of Psychology, Utrecht University, 3584 CS Utrecht, the Netherlands.

[4]Department of Psychiatry and Mental Health, MRC Unit on Risk & Resilience in Mental Disorders, University of Cape Town, Observatory, 7925 Cape Town, South Africa.

[5]Institute of Infectious Disease and Molecular Medicine (IDM), University of Cape Town, Observatory, 7925 Cape Town, South Africa.

*Correspondence to: Lisa A. Rosenberger (lisa.anna.rosenberger@univie.ac.at) or Jack van Honk (j.vanhonk@uu.nl, lead contact).

https://doi.org/10.1016/j.cub.2019.08.078

Summary

Trust and betrayal are central to our social world, and adaptive responses are crucial to our economic and social well-being [1]. We learn about others´ trustworthiness through trial-and-error during repeated interactions [2]. Rodent research has established a crucial role for the basolateral amygdala (BLA) in social experiential learning [3], by respectively reinforcing and suppressing behavior during positive and negative interactions with conspecifics [4]. The human BLA has undergone a reorganization with massive expansion relative to other amygdala nuclei [5], and there is no translational research on its role in experiential learning. The human amygdala is traditionally researched as a single structure [6], neglecting the sub-nuclei's structural und functional differences [7], which might explain inconsistent findings in research on social interactions [8, 9]. Here we study whether the human BLA is necessary for social and non-social experiential learning, by testing a group of five humans with selective bilateral damage to the BLA. We compared their learning behavior in a repeated trust game, and a non-social control task, to healthy, matched controls. Crucially, BLA-damaged subjects, unlike control subjects, completely failed to adapt their investments when interacting with a trustworthy and an untrustworthy partner. In the non-social task, BLA-damaged subjects learned from positive outcomes, but differed from the controls by not learning from negative outcomes. Our data extend findings in rodent research by showing that the human BLA is essential for social experiential learning and provide confirmatory evidence of divergent mechanisms for differentially-valenced outcomes in non-social learning.

Keywords

Basolateral amygdala, Trust, Trust Game, social learning, Decision making, Urbach-Wiethe disease, Brain lesion, Neuroimaging, Neuroeconomics.

Results and Discussion

To directly examine the role of the human basolateral amygdala (BLA) in social and non-social experiential learning, we investigated behavioral adaptation in 5 humans with selective bilateral calcifications of the BLA, and an intact and functional central-medial amygdala (CMA), and 17 healthy matched control subjects during a repeated trust game and a non-social control task (the Raffle task). In the repeated trust game two players make sequential decisions over multiple rounds of monetary transfers on whether to trust each other (Figure 1) [10]. Subjects interacted in turn with a generous trustee, who acted in a trustworthy way, and a selfish trustee, who acted in an untrustworthy way, for 52 rounds. A change in the subject's investments reflects the extent to which she learned about the trustee's trustworthiness [2]. In a cross-species study with BLA-damaged humans and BLA-silenced rats, we recently demonstrated that rodent amygdala models of defensive behavior directly translate to humans [11]. Correspondingly, if the human BLA, like the rodent BLA [3, 4], is vital to social learning from experience, then BLA-damaged subjects should not be able to learn about the trustee´s trustworthiness, and therefore should show no adaptive behavioral updating in the repeated trust game. In contrast, in non-social experiential learning, on the basis of rodent research [12], the involvement of the BLA might diverge for positive and negative outcomes. We used the Raffle task to explore whether human BLA involvement in experiential learning is specific to social context [13]. In this task subjects bought tickets for two different raffle boxes during 52 rounds. Outcomes and outcome probabilities exactly matched those used in the repeated trust game. Subjects had to learn from which raffle box they could win (the positive box) and from which they could lose money (the negative box), and adjust their ticket buying behavior accordingly.

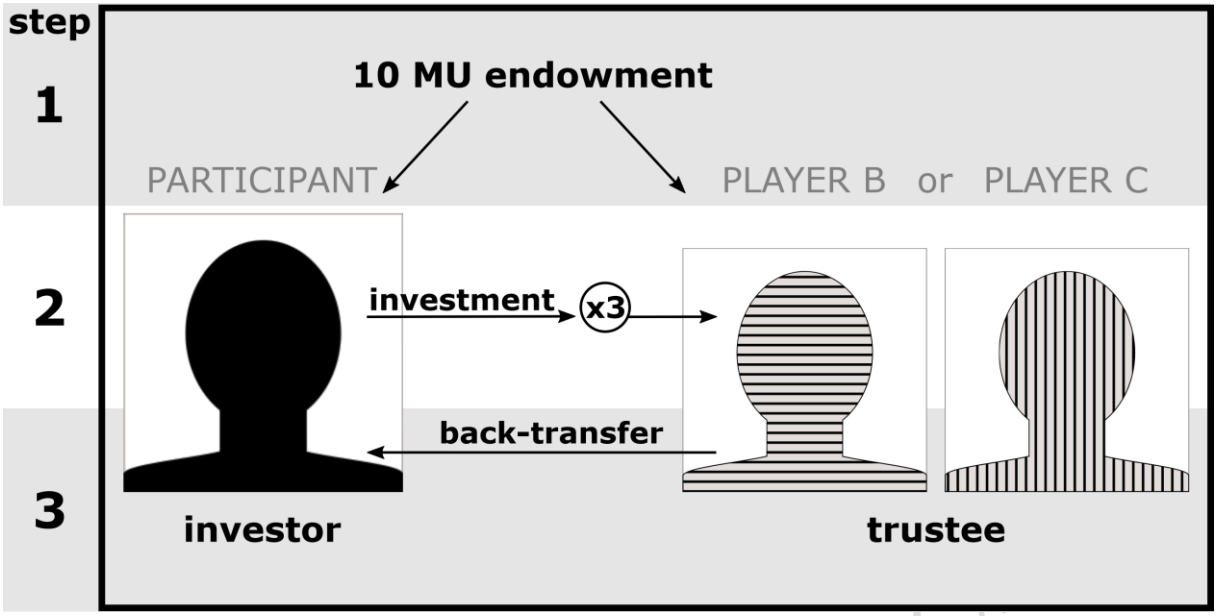

Figure 1. Repeated trust game design with original avatar pictures

Step 1: both players receive an endowment of 10 monetary units (MU) at the beginning of each round. Step 2: the subject makes an investment (between 1 – 10 MU), which gets tripled and send to the trustee (player b or c). Step 3: the trustee then makes a variable back-transfer. One trustee displayed generous behavior (back-transfers equal or higher than investments), the other displayed selfish behavior (back-transfers equal or lower than investments) [29].

In the repeated trust game, healthy control subjects learned that the two trustees were different and adapted their behavior accordingly (Figure 2A). They transferred significantly more money to the generous trustee than to the selfish trustee (Wilcoxon signed rank test z = 3.29, P = 0.001, n = 17), with average transfers of 7.51 MU (SE = 0.14) to the generous trustee and 4.14 MU (SE = 0.14) to the selfish trustee. The BLA-damaged subjects, on the other hand, did not learn to distinguish between the two trustees (Wilcoxon signed rank test z = 0.13, P = 0.89, n = 5), with average transfers of 5.33 MU (SE = 0.28) to the generous trustee and 5.15 MU (SE = 0.27) to the selfish trustee. The crucial difference between the transfers to the generous and selfish trustee was also significantly higher (3.37 vs 0.18 MU) in control subjects compared to BLA-damaged subjects (Mann-Whitney U-test, P = 0.019, n = 22). These comparisons are also significant when using a linear mixed model with random intercept per subject (Model 1 (M1) in Table S1).

Figure 2A (and Figure S1 with added individual data points) shows that the healthy subjects gradually learned that the generous trustee was more trustworthy than the selfish trustee, while the BLA-damaged subjects fluctuated around 5 MU without noticeable change in the two trustees. To test for differences in learning, we estimated a linear mixed model with random intercept per subject (Model 2 (M2) in Table S1). The matched controls clearly learned about the trustee´s trustworthiness, as they adjusted their investment behavior in an adaptive manner (trustee type x round number: b = -0.201, p < 0.001). They steadily increased investments in the generous trustee (b = 0.074, p < 0.001), and steadily decreased investments in the selfish trustee (b = -0.126, p < 0.001). These slopes differed from those of the BLA-damaged group (group x trustee type x round number: b = 0.16, p = 0.001). BLA-damaged subjects did not increase or decrease their investment behavior in the two trustees ($b_{generous}$ = 0.044, p = 0.155; $b_{selfish}$ = 0.002, p = 0.940; trustee type x round number: b = -0.042, p = 0.341), pointing to absent social experiential learning. This resulted in the control subjects

earning more money than BLA-damaged subjects (Mann-Whitney U-test, P = 0.034, n = 22, Figure 2B).

We measured the subjects' self-reported trust in the trustees, reflecting their perception of the trustees' trustworthiness. During the last two rounds of the task we prompted the subjects after their investments, and before they saw any back-transfer feedback, to indicate on a VAS scale (ranging from 0 to 100) how much they trusted the trustee in this round. Generous trustees ($m_{controls}$ = 77 (SE = 7), $m_{BLA\text{-}damage}$ = 63 (SE = 16)) were rated as significantly more trustworthy than selfish trustees ($m_{controls}$ = 24 (SE = 7), $m_{BLA\text{-}damage}$ = 35 (SE = 18)), Wilcoxon signed rank test: z = 2.841, p = 0.005, n = 22. The groups did not differ in their ratings (Mann-Whitney U-test: p = 0.967, n = 22), and the difference in the ratings of the trustees did not differ between the two groups (Mann-Whitney U-test: p = 0.319, n = 22). This suggests that the BLA-damaged subjects have potentially formed an explicit memory of the trustworthiness of the trustees. However, the variability of their ratings was high, and they were also not associated with their investments in that round (b = 0.00, p = 0.837). In the healthy subjects, the association between their ratings and investments differed significantly from the BLA-damaged subjects (trustworthiness rating x group interaction: b = -6.38, p = 0.011, Table S2), such that their ratings were positively associated with the investments in that round (b = 0.07, p < 0.001). Thus, while the ratings point towards explicit memory formation in the BLA-damaged subjects, they were unable to use this memory to adjust their investment behavior. While this interpretation requires caution, it fits with the function of the rodent BLA, which is not part of the declarative memory system [14, 15], but rather is involved in non-declarative, emotionally significant memory formation [16, 17].

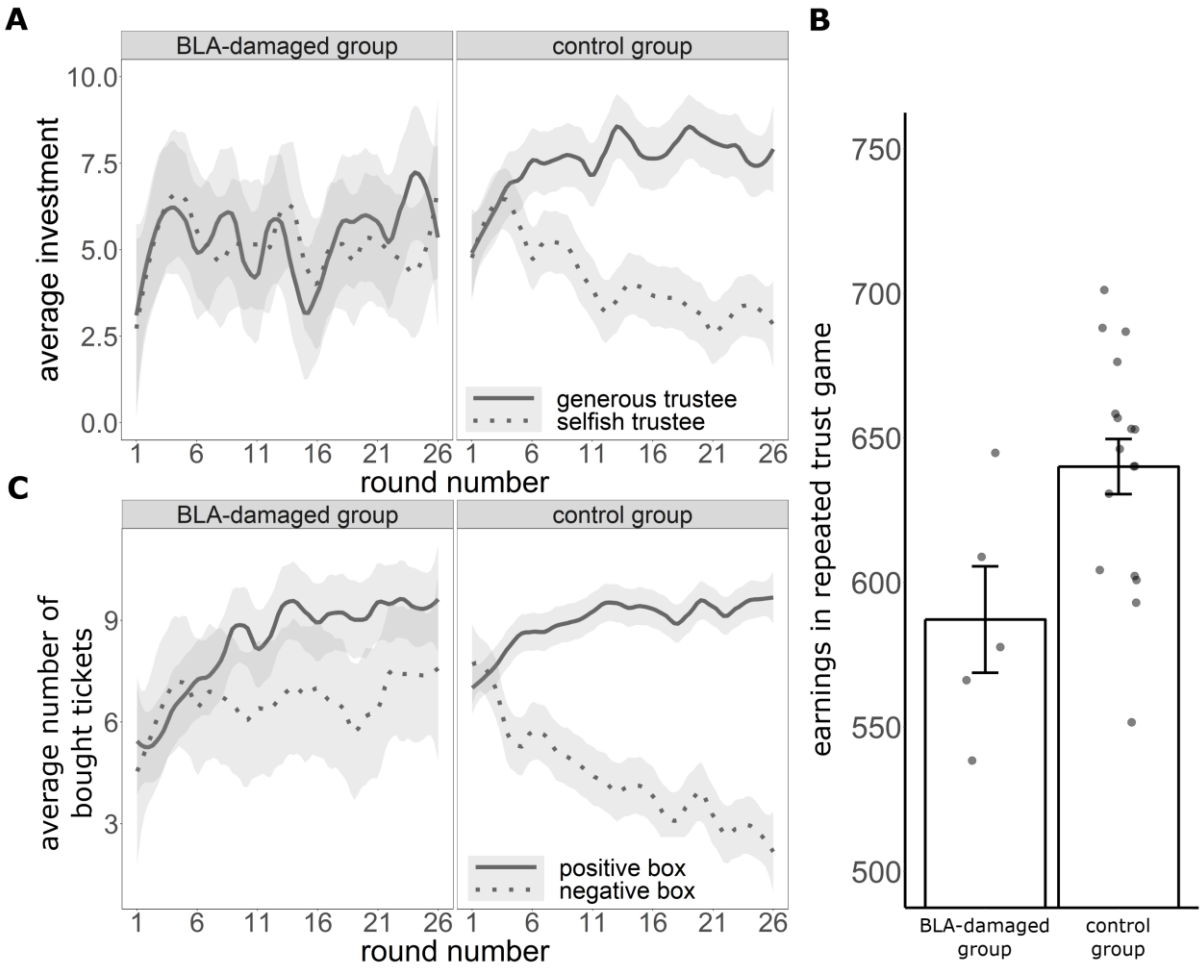

Figure 2. Learning in the repeated trust game and the raffle task

(A) Investments in the repeated trust game. Rounds are split into consecutive trustee interactions (N = 26 per trustee), plot is divided by group. Group data is smoothed with local regression (LOESS), shading represents 95% confidence intervals. (B) Total earnings (in MU) in repeated trust game, split by group. Dots represent individual subjects ($N_{control\ group}$ = 17, $N_{Bla\text{-}damaged\ group}$ = 5). (C) Bought tickets in the Raffle task. Rounds are split into consecutive box raffles (N = 26 per box), plot is divided by group. Group data are smoothed with local regression (LOESS), shading represents 95% confidence intervals. Individual data relating to panel A can be found in Figure S1. Analyses details can be found in Table S1 (for repeated trust game) and S3 (for Raffle task).

In the Raffle task, BLA-damaged subjects learned about the outcomes of the positive box and increased the amount of bought tickets, whereas healthy control subjects learned about the outcomes of both boxes and adapted their ticket buying behavior accordingly (see Figure 2C). To test the groups' learning behavior for the two boxes, we estimated a linear mixed model with random intercept per subject (Table S3). BLA-damaged subjects (N = 5) increased the amount of bought tickets for the positive box from an average of 5.6 (SE = 1.36) to 9.8 (SE = 0.2), b = 0.08, p < .001, but did not decrease the amount of bought tickets for the negative box ($m_{round\ 1}$ = 4.6 (SE = 1.4), $m_{round\ 52}$ = 7.6 (SE = 1.69)), b = 0.02, p = .066 over the rounds (box x round number interaction: b = 0.06, p = .004). This differed from the control subjects' learning behavior for the two boxes (group x box x round number interaction: b = -0.07, p < .001). Control subjects (N = 18) adjusted their buying behavior both for the positive and the negative box over the rounds (box x round number interaction: b = 0.13, p < .001): they increased the amount of bought tickets for the positive box from an average of 6.94 (SE = 0.61) to 9.72 (SE = 0.14), b = 0.04, p < .001, and decreased the amount of bought tickets for the negative box from an average of 7.72 (SE = 0.53) to 2.17 (SE = 0.2), b = -0.09, p < .001 over the rounds. Thus, in the Raffle, BLA-damaged subjects were able to learn from positive outcomes but had impaired learning from negative outcomes where they could lose money.

During forced-choice control questions after the Raffle task, BLA-damaged subjects were able to discriminate between different reward magnitudes. They were as accurate as control subjects in choosing the box with which they earned more money during the task (Fisher's exact test: p = .217, N = 23), as well as the box which would earn them a higher reward when the reward options in the two boxes partly overlapped (Fisher's exact test: p > .99, N = 23), and when they did not overlap (Fisher's exact test: p = .539, N = 23).

Using the Raffle task, combined with control questions, we show that the experiential learning impairments in BLA-damaged subjects are specific to the social context. Furthermore, these learning impairments evidently are not due to problems in discrimination

of reward magnitude, which is in line with spared reward discrimination abilities in BLA-lesioned rats [4]. The BLA-damaged subjects' inability to learn from negative outcomes in the Raffle task suggests different positive and negative non-social experiential learning mechanisms and related underlying neurocircuitry. In this respect our data are consistent with the theory of LeDoux and Daw [12], which holds that rodent instrumental responses in a positive non-social context can be learned directly through stimulus-response associations [18], rather than relying on BLA-dependent response–outcome associations [19]. Negative non-social experiential learning, in contrast, relies on BLA-dependent response-outcome associations in rodents [20]. In sum, we show that in humans experiential learning in a social context is fully dependent on the BLA, while in a non-social context only negative experiential learning is dependent on the BLA. The latter finding supports the model of Ledoux and Daw [12], and is consistent with impairments in Pavlovian threat conditioning in BLA-lesioned subjects [21].

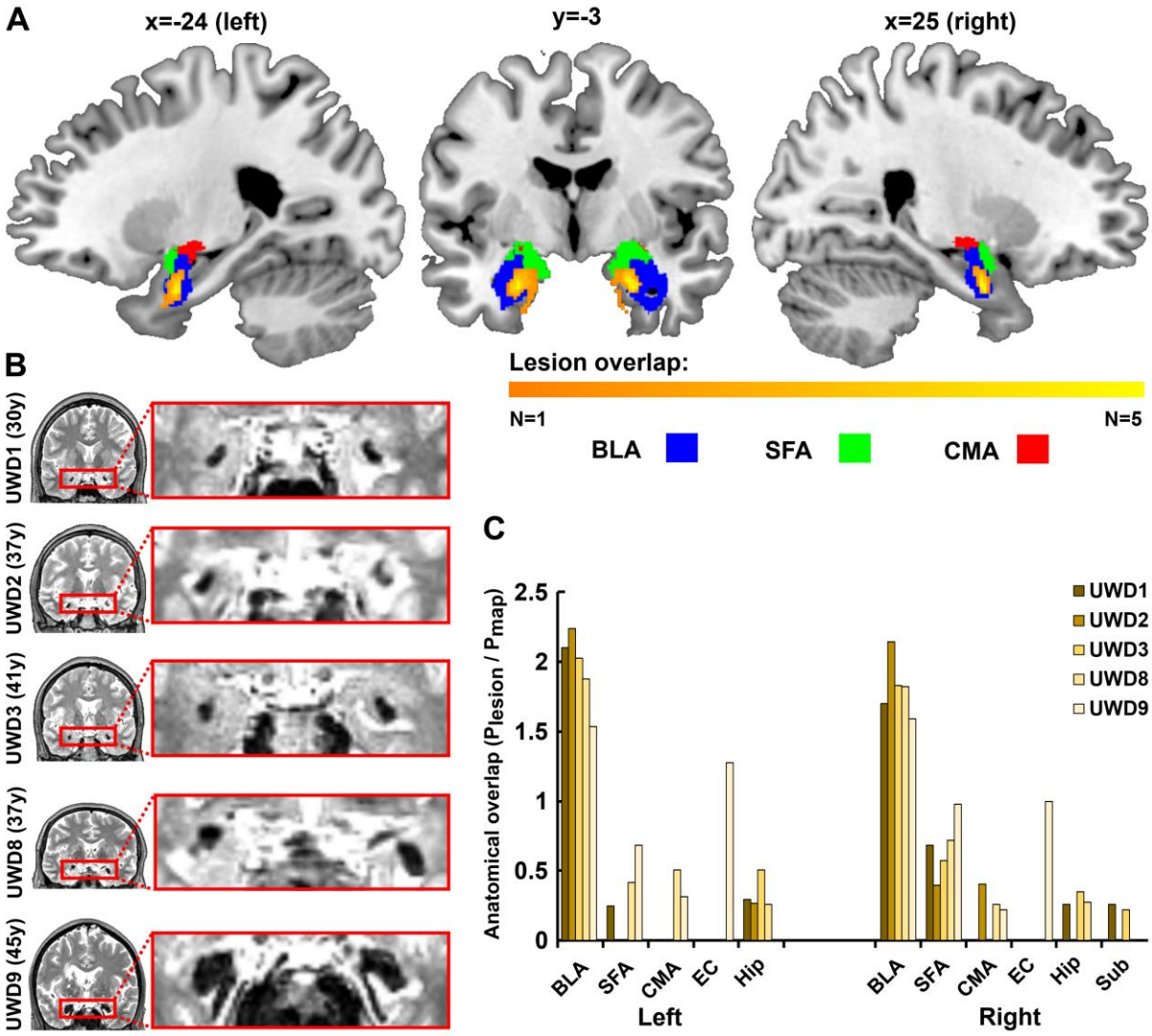

Figure 3. Bilateral and focal calcifications to the BLA

(A) Lesion-overlap image in MNI space plotted within the amygdala sub-regions defined as voxels with sub-region probability >50% (see Methods). (B) Coronal slices from each individual's raw T2-weighted MRI scan and age at time of scanning. (C) Bar-graph representing bilateral anatomical overlap with excess probability (P(lesion)/P(map)) values of the lesion volumes, whereby values >1 indicate a reliable match of volume and anatomical location of: BLA = Basolateral Amygdala, SFA = Superficial Amygdala, CMA = Central-Medial Amygdala complex, EC = Entorhinal Cortex, Hip = Hippocampus, Sub = Subiculum. UWD <X> (Urbach-Wiethe disease) represent the BLA-damaged subjects' identifiers.

Our results cannot be attributed to impairments in IQ (see Table 1), working memory [22], or emotion perception [23], as BLA-damaged subjects perform similar to and in some instances even better than healthy control subjects in these areas. Additionally, all subjects were instructed by a local native speaker who ensured that everyone fully understood the task. Lastly, previously we showed generous investments in BLA-damaged subjects during a one-shot trust game [24], where three BLA-damaged subjects invested 100% more than matched healthy controls. This generosity was not driven by increased trust in others, but arguably by altruistic motives, because the BLA-damaged subjects had similarly low expectations about their partners´ back-transfers as the control subjects. In the present task we did not observe such generosity, defensibly because different behavioral mechanisms are present in the repeated trust game. That is, the repeated trust game we use not only requests continuous adjustments in response to feedback, but even before feedback there are reputational concerns that strongly influence decisions about investments.

In line with rodent models [3, 7], we demonstrate that the human BLA is necessary for experiential learning in a social context, that is, continuously adjusting behavior to distinct feedback from selfish and generous partners. In rodents the BLA–OMPFC (orbital medial prefrontal cortex) network seems to underlie learning that stimuli are associated with distinct outcomes, and subsequent behavioral updating [25]. Our data suggest that the function and neural connectivity of the human BLA in this learning depends on an interaction between context (social or non-social) and valence (positive or negative) of the outcomes. These findings translate to rodent models of BLA function in social reinforcement learning [3] and in learning from non-social positive and negative outcomes [12].

Our study furthermore highlights the importance of studying the functionality of human amygdala sub-nuclei separately, which is the typical approach in rodent research. Since amygdala sub-nuclei have different functionality and connectivity within the brain [26],

behavioral outcomes from human subjects with damage to the whole amygdala [27] may be less informative for cross-species translation. Given the antagonistic properties of the BLA and CMA in decision-making in rodents, effects of damage on one sub-region might even be abolished by lesions to both sub-regions [7, 28]. We cannot exclude the possibility that selective CMA lesions in humans would result in similar behavioral patterns. There is no possibility to test this, as there are no humans with selective CMA lesions, as far as we know. Crucially however, our findings in the BLA-damaged subjects reported here and earlier [11, 21, 24] continuously translate to rodent models of the BLA. On that basis, we expect that the trust learning deficit in our BLA-damaged subjects depends on their selective BLA damage (see Figure 3), and, also given our findings in the non-social Raffle task, that the different role for the rodent CMA in learning [7] might also apply to humans.

The BLA-damaged subjects' impairments in Pavlovian fear learning [21] are consistent with their impairments in experiential learning from negative outcomes in the Raffle task. In contrast, and most essentially, our findings with the repeated trust game show that the human BLA is necessary for all social experiential learning, that is, for both learning from negative and positive social outcomes. The extent to which the human BLA is involved in other aspects of reward processing [25] that are important for social decision-making remains to be established. In the interim, our study is the first that addresses the gap in cross-species translational research on the role of the human BLA in social experiential learning. We have established that the human BLA plays a crucial role in social learning, and specifically in deciding who can and cannot be trusted after repeated social-economic interactions, and in adjusting investment behavior accordingly. This parallels results and predictions from rodent research, which emphasize the role of the BLA in social learning [3, 7], but the relevance of these findings for humans was hitherto unresolved.

Acknowledgments

In loving memory of Christoph Eisenegger, who was the mind and soul behind this project. Chris died in Namibia, 27 February 2017. We thank Marijn Moerbeek for creating Figure 1 and Claus Lamm for valuable feedback on the manuscript. The work in this paper was supported by the National Research Foundation (NRF) South Africa, the South African Medical Research Council, the University of Cape Town (UCT), Utrecht University (UU), Royal Holloway, the Vienna Science and Technology Fund (WWTF VRG13-007), and the Marietta-Blau grant (OeAD-GmbH and the Austrian Ministry for Science, Research, and Economics (BMWFW)).

Author Contributions

Conceptualization and Methodology, L.A.R., C.E., M.N., and J.v.H.; Investigation, C.E., M.N., J.C., L.A.R., and J.v.H.; Formal Analysis, L.A.R., M.N. and D.T.; Writing – original draft, L.A.R., D.T., M.N., and J.v.H.; Writing – review & editing, L.A.R., D.T., M.N., D.J.S. and J.v.H.; Funding Acquisition, C.E., L.A.R., and J.v.H.; Supervision, C.E., D.J.S, and J.v.H.

Declaration of Interests

D.J.S. has received research grants and/or consultancy honoraria from Lundbeck, and Sun. All other authors declare no competing interests.

Tables with titles and legends

Table 1. Age and IQ from the Wechsler Abbreviated Scale of Intelligence (WASI) for the BLA-lesioned subjects and healthy controls.

VIQ, verbal IQ; PIQ, performance IQ; FSIQ, full-scale IQ; SD, standard deviation; UWD, Urbach-Wiethe disease.

	BLA-damaged subjects							Controls	
	UWD 1	UWD 2	UWD 3	UWD 8	UWD 9	Mean	SD	Mean	SD
Age	30	37	41	37	45	38.00	5.57	36.41	6.86
VIQ	97	84	93	89	87	90.00	5.10	89.71	3.77
PIQ	99	87	85	95	89	91.00	5.83	88.65	4.73
FSIQ	98	84	87	91	86	89.20	5.54	87.76	3.03

STAR Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Jack van Honk (j.vanhonk@uu.nl). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We tested 5 female subjects with focal, bilateral damage to the BLA, due to a genetic disease (Urbach-Wiethe disease) where parts of the brain calcify over time [30]. Previous research in this population showed that the calcifications are bilateral and focal to the BLA, while the CMA remains functional [23]. The control group, consisting of 17 healthy subjects, were matched for age, gender, socio-economic status, cultural-ethnic status, religion, and IQ (measured with the Wechsler Abbreviated Scale of Intelligence (WASI) [31]), see Table 1. All subjects were recruited from the same rural Namaqualand area in South Africa, and all had similar socio-economic backgrounds. None of the subjects had any history of secondary psychopathology. Testing took place in the rural Namaqualand area with a local experimenter who spoke the same Afrikaans language as the subjects. Subjects provided informed consent before the beginning of the test session. The Health Sciences Faculty Human Research Ethics Committee of the University of Cape Town, South Africa, approved the study (HREC # 639/2016).

METHOD DETAILS

<u>Repeated Trust Game</u>

The Repeated Trust Game was adapted from Fett and colleagues [29]. Subjects played 52 rounds in the role of the investor, with two randomly alternating trustees. Subjects did not receive any priors about the trustees and had to learn through trial-and-error how to best

adjust their investments (i.e. high investments with generous trustee, low investments with selfish trustee). All players were depicted with a neutral avatar picture (see Figure 1) and a neutral name (player B/ C), which were randomly assigned to the trustees. At the beginning of every round both players received an endowment of 10 monetary units (=MU). The investor had to invest at least 1 MU. Investments were tripled and transferred to the trustee. Trustees could make a variable back-transfer. In the first four rounds of the task back-transfers of the trustees were 100% of the subject´s investment, so that there were no immediate ceiling or floor effects on the investment and a learning curve could be measured easily. After the fourth-round back-transfers for the generous trustee were either 100%, 150%, or 200% of the investment (with an equal probability of every option occurring). When the subject increased their investment relative to the previous round with the generous trustee, the probability of receiving a back-transfer of 200% of the investment increased with 10% (and thus the other two options decreased with 5% each). Back-transfer options for the selfish trustee were either 100%, 75%, or 50% of the investment. When the subject decreased their investment relative to the previous round with the selfish trustee, the probability of receiving a back-transfer of 50% of the investment increased with 10%. Because of the semi-random nature of the back-transfer options, it is possible that some subjects received ambivalent back-transfers of 100% of their investments in consecutive rounds, making it difficult to discern the trustees and thus to learn about their trustworthiness. Analyses of the back-transfer options in rounds 5 - 16 [43] did not reveal a group difference in the amount of ambivalent back-transfer options. In the last two rounds of the task, subjects were prompted to indicate the trustworthiness (with a slider on a scale from 0 – 100) of the trustee they were playing with in that round. This prompt appeared after the investment and before the back-transfer decision. Analyses of the association between the investments and the ratings in these two rounds are described in Table S2. Points earned throughout the task were exchanged to South African Rand. The task was played in z-tree (version 3.42 [32]). The experimenter explained the task verbally in

Afrikaans until subjects understood the task. In addition, subjects played 10 practice rounds in z-tree, while the experimenter was present to familiarize themselves with the software and the task. During the practice rounds the experimenter gave additional explanations of the task.

Raffle Task

This task was designed to match exactly the outcomes, the outcome probabilities, as well as the adjustment of the outcome probabilities of the repeated trust game. The task was programmed in z-tree (version 3.42). In this task subjects played a raffle for 52 rounds. In the raffle, subjects could buy raffle tickets for two different boxes, delineated by different colors (blue and green) and patterns (horizontal stripes, no stripes), in consecutive rounds (order of box presentation was random and was restricted to two consecutive presentations per box). Subjects did not have any priors about the boxes. At the beginning of a round, one of the two closed raffle boxes was presented, indicating for which box subjects would buy raffle tickets in that round. Then in step 1, subjects indicated how many raffle tickets (between 1 and 10) they want to buy for the box. In step 2, the box was opened and the amount of money the subject won was presented. Earnings were paid out at the end of the task. Subjects played 5 practice rounds with different raffle boxes. After the task, subjects were presented with three forced-choice control questions. In the first question the two closed boxes from the Raffle task were shown side by side and subjects had to choose the one with which they earned more money during the task. In the other questions two new boxes were shown with each containing three possible outcomes. In question two these outcomes represented the outcome probabilities of the Raffle task (box 1: 5, 7, 10 MU; box 2: 10, 15, 20 MU), and in question three the possible outcomes were non-overlapping (box 1: 12, 15, 18 MU; box 2: 8, 6, 4 MU). Subjects were instructed to choose one of the two presented boxes, and informed that one of the possible outcomes from the chosen box would be paid out at the end of the task. All instructions and in-game text were displayed in the native language of the subjects (Afrikaans).

Lesion methods

MRI-scans were acquired with a Siemens Magnetom Skyra 3-Tesla scanner at the Cape Universities Brain Imaging Centre (CUBIC) in Cape Town, South Africa. For lesion analysis, we obtained whole brain T2-weighted images using an iPAT acceleration-factor of 3, voxel-size 0.9x0.9x1mm, TR = 3200 ms, and TE = 410 ms.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

We analyzed investment behavior both with parametric and with non-parametric two-tailed tests with alpha set to 0.05. For the parametric tests we computed linear mixed models with the lme4 package (version 1.1-15 [33]) in R (version 3.4.3 [34]), which included a random intercept per subject. Effects were calculated with a treatment contrast and p-values were computed via Wald-statistics approximation. The results table in the supplements was produced with the SjPlot package (version 2.4.1 [35]), and the graphs with the ggplot2 package (version 2.2.1 [36]). For the non-parametric tests, we used the Mann-Whitney U test for between-subject comparisons, the Wilcoxon-signed rank test for within-subject comparisons, and Kendall´s tau for the correlation between final rounds investment and trustworthiness rating.

Lesion analysis

To estimate extent and anatomical location of the lesions, T2-weighted scans were normalized to MNI-space using unified segmentation, which is optimized for normalization of lesioned brains [37]. Lesion volumes were defined using the 3D volume-of-interest featured implemented in MRIcron [38]. Based on MR-images the precise borders between amygdalae and neighboring structures, or between the sub-regions of the amygdala, cannot be established [39, 40]. To determine the precise location of the lesions in our BLA-damaged subjects, we therefore assigned the lesion volumes to cytoarchitectonic probability maps according to the

method described by Eickhoff and colleagues [41]. In this method, which is implemented in the SPM8 anatomy toolbox [42], a volume of interest is superimposed onto a cytoarchitectonic probability map of the medial-temporal lobe [39]. This map is based on microscopic analyses of ten post-mortem human brains and follows a generally accepted division of the human amygdala in three sub-regions. The first is the central-medial amygdala (CMA), which consists of the central and medial nuclei. The second is the basolateral amygdala (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, amygdalopyrifom transition area, amygdaloid-hippocampal area, and the cortical nucleus [39]. This method assigns to any given voxel a value representing the probability that it belongs to an underlying structure. These structures are derived from an overlap analysis of ten post-mortem brains and are therefore divided in ten 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: $\mathrm{P(excess)} = \frac{\mathrm{P(lesion)}}{\mathrm{P(map)}}$ 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 exceed 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(excess) > 1 indicates a more central, and P(excess) < 1 a more peripheral location [41]

As depicted in Figure 3A and C, calcified brain-tissue of the BLA-damaged subjects is localized in the BLA and the CMA seems unaffected. In a quantitative analysis these results

are confirmed. Figure 3C shows P(excess) values for the individual lesions and these lesions are, bilaterally, most central to the BLA as P(excess) values exceed 1 for each individual and hemisphere. Since this method is purely based on probability distributions, it is impossible to fully exclude the possibility that structures other than the BLA are affected by the calcifications. The fact that the lesion-volumes largely overlap with high probability classes in the bilateral BLA, and that P(excess) values greatly exceed the value of 1, can however be seen as strong support for our claim that these subjects have bilateral damage limited to the BLA. In subject UWD9 the calcifications might extend into neighboring structures, namely the right superficial amygdala (SFA) and bilateral entorhinal cortex (all main results reported above remain significant when excluding subject UWD9). We can however safely conclude that the CMA is unaffected by the bilateral calcifications found in all of these BLA-damaged subjects.

DATA AND CODE AVAILABILITY

Behavioral data and analyses scripts are deposited here: osf.io/hwc5q [43]. Data relating to lesion analyses will only be made available upon request from the lead contact due to the de-anonymized nature of the data.

References

1. Axelrod, R., and Hamilton, W.D. (1981). The evolution of cooperation. Science *211*, 1390-1396.
2. Chang, L.J., Doll, B.B., van 't Wout, M., Frank, M.J., and Sanfey, A.G. (2010). Seeing is believing: trustworthiness as a dynamic belief. Cogn Psychol *61*, 87-105.
3. Hernandez-Lallement, J., van Wingerden, M., Schable, S., and Kalenscher, T. (2017). A Social Reinforcement Learning Hypothesis of Mutual Reward Preferences in Rats. Curr Top Behav Neurosci *30*, 159-176.
4. Hernandez-Lallement, J., van Wingerden, M., Schable, S., and Kalenscher, T. (2016). Basolateral amygdala lesions abolish mutual reward preferences in rats. Neurobiol. Learn. Mem. *127*, 1-9.
5. Janak, P.H., and Tye, K.M. (2015). From circuits to behaviour in the amygdala. Nature *517*, 284-292.
6. Gupta, R., Koscik, T.R., Bechara, A., and Tranel, D. (2011). The amygdala and decision-making. Neuropsychologia *49*, 760-766.
7. Balleine, B.W., and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. Trends Neurosci. *29*, 272-279.
8. Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron *58*, 639-650.
9. FeldmanHall, O., Dunsmoor, J.E., Tompary, A., Hunter, L.E., Todorov, A., and Phelps, E.A. (2018). Stimulus generalization as a mechanism for learning to trust. Proc. Natl. Acad. Sci. U. S. A. *115*, E1690-E1697.
10. Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. GEB *10*, 122 - 142.
11. Terburg, D., Scheggia, D., Triana Del Rio, R., Klumpers, F., Ciobanu, A.C., Morgan, B., Montoya, E.R., Bos, P.A., Giobellina, G., van den Burg, E.H., et al. (2018). The Basolateral Amygdala Is Essential for Rapid Escape: A Human and Rodent Study. Cell *175*, 723-735 e716.
12. LeDoux, J., and Daw, N.D. (2018). Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. Nature Reviews Neuroscience *19*, 269-282.
13. Fareri, D.S., Chang, L.J., and Delgado, M.R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. Frontiers in neuroscience *6*, 148.
14. Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. Nature reviews. Neuroscience *1*, 41-50.
15. Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. Science *269*, 1115-1118.
16. McDonald, R.J., and White, N.M. (1993). A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behav. Neurosci. *107*, 3-22.
17. Roozendaal, B., McEwen, B.S., and Chattarji, S. (2009). Stress, memory and the amygdala. Nature reviews. Neuroscience *10*, 423-433.
18. Lingawi, N.W., and Balleine, B.W. (2012). Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. J. Neurosci. *32*, 1073-1081.
19. Balleine, B., Killcross, S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. J. Neurosci. *23*, 666 - 675.
20. Killcross, S., Robbins, T.W., and Everitt, B.J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature *388*.
21. Klumpers, F., Morgan, B., Terburg, D., Stein, D.J., and van Honk, J. (2015). Impaired acquisition of classically conditioned fear-potentiated startle reflexes in humans with focal bilateral basolateral amygdala damage. Social cognitive and affective neuroscience *10*, 1161-1168.

22. Morgan, B., Terburg, D., Thornton, H.B., Stein, D.J., and van Honk, J. (2012). Paradoxical facilitation of working memory after basolateral amygdala damage. PLoS One *7*, e38116.
23. Terburg, D., Morgan, B.E., Montoya, E.R., Hooge, I.T., Thornton, H.B., Hariri, A.R., Panksepp, J., Stein, D.J., and van Honk, J. (2012). Hypervigilance for fear after basolateral amygdala damage in humans. Transl Psychiatry *2*, e115.
24. van Honk, J., Eisenegger, C., Terburg, D., Stein, D.J., and Morgan, B. (2013). Generous economic investments after basolateral amygdala damage. Proc. Natl. Acad. Sci. U. S. A. *110*, 2506-2510.
25. Wassum, K.M., and Izquierdo, A. (2015). The basolateral amygdala in reward learning and addiction. Neurosci. Biobehav. Rev. *57*, 271-283.
26. McDonald, A.J. (1998). Cortical pathways to the mammalian amygdala. Prog. Neurobiol. *55*, 257-332.
27. Bhatt, M.A., Lohrenz, T., Camerer, C.F., and Montague, P.R. (2012). Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game. Proc. Natl. Acad. Sci. U. S. A. *109*, 8728-8733.
28. Stalnaker, T.A., Franz, T.M., Singh, T., and Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. Neuron *54*, 51-58.
29. Fett, A.K., Gromann, P.M., Giampietro, V., Shergill, S.S., and Krabbendam, L. (2012). Default distrust? An fMRI investigation of the neural development of trust and cooperation. Social cognitive and affective neuroscience *9*, 395-402.
30. Koen, N., Fourie, J., Terburg, D., Stoop, R., Morgan, B., Stein, D.J., and Van Honk, J. (2016). Translational neuroscience of basolateral amygdala lesions: studies of Urbach-Wiethe Disease. J. Neurosci. Res. *94*, 504 - 512.
31. Wechsler, D. (1999). Wechsler abbreviated scale of intelligence, (San Antonio, Texas: Psychological Corporation).
32. Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental economics *10*, 171-178.
33. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software *67*, 1 - 48.
34. R Core Team (2017). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).
35. Ludecke, D., and Schwemmer, C. (2017). sjPlot: Data visualization for statistics in social science (version 2.3. 1)(software).
36. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis., (New York: Springer).
37. Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., and Friston, K. (2007). Spatial normalization of lesioned brains: Performance evaluation and impact on fMRI analyses. Neuroimage *37*, 866-875.
38. Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. Behavioural Neurology *12*, 191 - 200.
39. Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N.J., Habel, U., Schneider, F., and Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. Anatomy and Embryology *210*, 343-352.
40. Solano-Castiella, E., Anwander, A., Lohmann, G., Weiss, M., Docherty, C., Geyer, S., Reimer, E., Friederici, A.D., and Turner, R. (2010). Diffusion tensor imaging segments the human amygdala in vivo. Neuroimage *49*, 2958-2965.
41. Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., and Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage *36*, 511-521.
42. Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., and Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage *25*, 1325-1335.

43. Rosenberger, L.A., Eisenegger, C., Naef, M., Terburg, D., Fourie, J., Stein, D.J., and Van Honk, J. (2019). Data and analyses scripts accompanying manuscript: The human basolateral amygdala is indispensable for social experiential learning. (osf.io/hwc5q).