Dorsolateral prefrontal neurons mediate subjective decisions and their variation in humans

Mohsen Jamali[1,*], **Ben Grannan**[1,*], **Keren Haroush**[2,*], **Ziev B. Moses**[3], **Emad N Eskandar**[4], **Todd Herrington**[5], **Shaun Patel**[5], and **Ziv M Williams**[1,6,7]

[1.] Department of Neurosurgery, Harvard Medical School, Massachusetts General Hospital, Boston MA

[2.] Department of Neurobiology, Stanford University School of Medicine, Palo Alto CA

[3.] Department of Neurosurgery, Brigham and Women's Hospital, Boston MA

[4.] Department of Neurosurgery, Albert Einstein University, Bronx NY

[5.] Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston MA

[6.] Harvard-MIT Division of Health Sciences and Technology, Cambridge MA

[7.] Harvard Medical School, Program in Neuroscience, Boston MA

Abstract

Subjective decisions play a vital role in human behavior because, while often grounded in fact, they are inherently based on personal beliefs that can vary broadly within and between individuals. While these properties set subjective decisions apart from many other sensorimotor processes and are of wide sociological impact, their single-neuronal basis in humans is unknown. Here, we find cells in the dorsolateral prefrontal cortex (dlPFC) that reflected variations in the subjective decisions of humans when performing naturalistic opinion-based tasks. These neurons changed their activities gradually as the participants transitioned between choice options but also reflected their unique point of transition at equipoise. Focal disruption of the dlPFC, by contrast, diminished gradation between opposing decisions but had little effect on sensory perceptual choices or their motor report. These findings suggest that the human dlPFC plays an important role in subjective decisions and propose mechanism for mediating their variation during opinion formation.

Subjective decisions play an important role in human behavior – from how we vote in political elections to how we socially interact with others or what products we buy [1–3]. Prior theoretical and observational work has suggested that subjective decision making provides a

vital function in human behavior because it allows us to consider options under conditions in which there are no intrinsically correct choices or rubric to guide our selections [1, 2, 4]. Therefore, compared to other sensorimotor processes that are based on trained contingencies between sensory input, action and outcome; subjective decisions are inherently based on personal beliefs that can vary broadly within or between individuals even when presented with the same information. The single-neuronal mechanisms that underlie subjective decisions or their variation in humans, however, remain largely unknown.

Animal studies have provided critical insight into the likely role frontal cortical areas such as the dorsolateral prefrontal cortex (dlPFC) play in action selection [5–7]. For example, studies using sensory stimuli such as ambiguous apparent motion or auditory cues have offered an important understanding of how choices are made based on sensory percepts and evidence accumulation [8–13]. They have also provided an improved understanding of how neurons encode information related to economic value and reward [14, 15]. How humans make non-perceptual decisions or how subjective decisions are made in situations under which there are no explicitly correct choices (e.g., when forming opinions about a political policy, the morality of a situation or the trustworthiness of others) [16–21], however, has remained less well understood. Furthermore, while there has been a growing understanding of how distributed frontal injuries from as trauma or stroke affects human decision making [22, 23], little is known about the contribution of localized areas such as the dlFPC.

In this study, we aimed to evaluate whether and what role neurons in the dlPFC may play in subjective decision making in humans by acutely following the activity of individual neurons in the dlPFC of participants undergoing planned operative neurophysiology (Fig. 1a) [24]. To further evaluate the causal contribution of the dlPFC to this process, we also examined the performances of individuals who had previously undergone clinically targeted resection of focal lesions in this same area. Using these combined approaches, we find that neurons in the dlPFC reflected differences in the participant's subjective decisions as they performed naturalistic opinion-based tasks, revealing spiking activities that changed gradually as they transitioned from one choice option to another and which reflected their unique point of transition (i.e., 'tipping point') at equipoise. These activity changes were most prominent prior to motor selection and were not observed when making sensory perceptual choices. Focal disruption of the dlPFC, on the other hand, led to a diminished gradation between the opposing decisions and was consistently observed across two different behavioral tasks. Disruption of the dlPFC, however, had little effect on the participants' sensory perceptual choices or their motor report. Based on these findings, we propose a prospective role for the human dlPFC in subjective decision making and its relation to opinion formation.

RESULTS

Situational assessment task

A common framework for evaluating human subjective decisions is through situational assessments. In these tasks, for example, a participant may be given a naturalistic scene of a jogger running near a busy street or a person filling a cup of hot tea near its brim, and then be asked whether they believe the situation to be safe or unsafe. These and similar

approaches have often been used in sociological and psychological studies [20, 21] because they allow one to follow the specific points at which individuals transition from one choice option to another as they form their opinions of the various scenes, and because they can be used to test the participant's subjective decisions under conditions in which there is no explicit correct or 'optimal' choice [21, 25]. Therefore, in our study, for each trial, the participants were presented with a visual scene followed by a brief delay period. They were then given two randomly positioned choice options used to indicate whether they believed the situation to be safe or unsafe. For generalizability and to further evaluate the point of equipoise at which the participants transitioned from one choice to the other, they were presented with seven prototypic scenes over the course of the session, each with its own unique situation and theme. For example, the participant may see a jogger running near a busy street in one scene but may see a worker standing in a construction yard in another. Further, each of the prototypic scenes consisted of five different situations in which the relation between the subject and object within the scenes were covertly manipulated. For example, the position of the jogger (subject) in relation to the car (object) may vary from one scene to another (Fig. 1b, Fig. 2a). Second, all scenes were given in random order throughout the session and no feedback was given to the participants at any point, such that there was no explicit correct choice, and the point of transition between choice options could intrinsically vary both within and between individuals. Finally, to confirm that the participants were making appropriate choices and to provide a control comparison, some of the scenes included only an object that could itself be objectively categorized as safe or unsafe (e.g., a 'gun' vs. a 'shirt'). Additional control tasks that did not involve elements of safety or that involved sensory perceptual choices are described further below.

In total, we recruited 39 participants. Eleven participants underwent single-neuronal recordings, 4 previously underwent focal dlPFC lesion resection and 3 previously underwent anterior temporal resection for of the situational assessment task (Fig 1a and Supplementary Table 1). An additional 4 participants underwent neuronal recordings, 5 had focal dlPFC lesion resections and 12 served as healthy participants for control task comparisons (Supplementary Table 2).

Behavioral performance

We evaluated the subjective decisions of the participants by first quantifying their probability of selecting one choice over the other across the different situations and scenes [14, 15]. For example, the probability of making the same choice (e.g., of safe) may gradually change as individuals transition, across equipoise, to the choice of unsafe. Therefore, when aligning trials in our task to the point at which the participants changed their choices from safe to unsafe *per* scene, choices aligned to the 'left' of equipoise would reflect situations that the participants subjectively believed to be progressively safer (ordinal positions from −1 to −4) whereas those aligned to the 'right' of equipoise would reflect situations that they believed to be progressively unsafe (ordinal positions from +1 to +4; Fig. 2d).

As expected, the participants displayed a gradual change in their probability of selecting one choice option over the other as they transitioned across equipoise (Lilliefors test, n=11 participants, p<0.05, Fig. 2b-d). Overall, when considered together across trials, the

participants made the same choice selections 90.0±3.3% of the time (Chi-square=127.6, p<0.0001). However, when considering their choices in relation to equipoise, the participants made the same choice selection 84.4±5.4% of the time when near equipoise (ordinal positions +1 and −1) and the same choice selection 94.4±2.9% of the time when further away (ordinal positions +2 and −2 and above; Chi-square=54.36, p<0.0001).

We also find that the participants gave the same choices when presented with similar situations. For instance, if the position of the jogger differed only slightly (e.g., comparing manipulations 1 and 2 in the middle panel of Fig. 1b), one would expect the participant to hold a similar opinion of them. Overall, the participants were significantly more likely than chance to make the same choice selections when the situations were similar to each other across all ordinal positions (82.9±3.5%; Chi-square=60.50, p<0.0001; Fig. 2c,d).

Finally, to confirm that the participants were indeed making appropriate choices, we evaluated their selections when asked to categorize individual objects as safe or unsafe in the absence of a situational assessment. Here, the participants agreed in their selection on 94.2±1.3% of trials (Chi-square=142.3, p<0.0001), meaning that essentially all participants agreed in their selections. Taken together, these findings suggested that the participant's understood the task and that their choices accurately reflected their subjective evaluations of the scenes.

Neuronal responses when making subjective decisions

We initially recorded from 119 well-isolated units in Brodmann's area 9 of the dlPFC. Here, a motorized micro-drive and a temporary biodegradable buffer were used to provide stable unit isolation (Fig 1a). Only units with a high degree of signal-to-noise, an adequate refractory period and stable waveform morphology over the course of recordings were included (Fig. 3a and Supplementary Fig. 1). This amounted to an average of 10.8±2.8 units *per* participant across the 11 participants that performed the situational assessment task.

We first searched for neurons whose activities may be modulated by the task. Here, we centered our analyses on the time period 500–2000 ms from scene onset to take into account the sensory response-delay normally observed in the prefrontal cortex as well as the timespan over which the scenes were presented. Of the recorded population, we find that 26% of cells (*n*=31) displayed a difference in response on trials in which the participants subsequently reported the scenes to be safe compared to unsafe (two- sided t-test, p<0.05; Fig. 3b,c and Supplementary Fig. 2). Consistent with prior studies [26, 27], only few cells displayed a difference in response to the prototypic scenes themselves (*n*=3; two- sided t-test, p<0.05; Bonferroni corrected for scene) or to the specific left/right movement (*n*=4; two- sided t-test, p<0.05).

To next examine whether and to what degree neuronal activity in the dlPFC may correlate with the participant's subjective decisions as they varied from trial-to-trial, we aligned each of the trials to the points of equipoise at which the participants transitioned from making the choice of safe to unsafe (see Methods). As described above, trials aligned to the 'left' of equipoise reflected situations that the participants believed to be progressively safe (ordinal positions from −1 to −4) whereas those aligned to the 'right' of equipoise reflected situations

that they believed to be progressively unsafe (ordinal positions from +1 to +4; Fig. 2d). Together, these psychometric curves constituted the participant's 'voting profile' [19,20]. Overall, 58% of the 31 neurons displayed responses that were positively correlated with the participants' voting profiles, meaning that they displayed a gradual increase in their spiking rate as the participants transitioned from viewing the scenes from safe-to-unsafe (Fig. 3d,e). Forty-two percent displayed activity which was negatively correlated. Figure 3d illustrates two such representative dlPFC cells recorded from the same participant. The overall degree of correlation across neurons was r = 0.49±0.03 (Supplementary Fig. 3). Neurons in the dlPFC therefore demonstrated gradual changes in their activity as the participants transitioned from one choice to the other across equipoise.

Individually, the spiking activities of these neurons were best described by a logistic function. Using a modeling approach to evaluate the activity profile of each neuron, we fit their firing rates to either a constant, linear, binary, logistic or u-shaped function (Fig. 4a). Thus, for example, whereas neuronal responses that are best fit by a binary function can be used to describe differences in the participants' binary choices, a logistic function (i.e., one that gradually changes from one choice to the other) can also be used to additionally describe the relation between their choices and equipoise. Calculating the residuals of each fit, we find that neural responses were best explained by a logistic fit (1.9-fold in AIC, n=31; Supplementary Table 3), with activities being significantly better explained by a logistic compared to binary function (variance accounted for; two-sided paired t-test, t(30)=5.35, p<0.0001; Fig. 4c).

These changes in activity did not reflect differences in choice uncertainty that peaked or fell as the participants approached equipoise. For example, an increase in spiking activity near equipoise could be explained by a nonspecific increase in uncertainty [11, 28–30] or change in 'value' (e.g., when based on reward [14, 31]) assigned to a choice. Here, however, we find that the spiking activities of the neurons were significantly better explained by a logistic compared to u-shaped function (variance accounted for; two-sided paired t-test, t(30)=9.05, p<0.0001, Fig. 4d). Consistent with these observations, there was also no difference in averaged population activity for trials lying near vs. far from equipoise (1.06±0.05 sp/s vs. 1.07±0.04 sp/s, respectively; two-sided paired t-test, t(30)=0.18, p=0.87; Supplementary Fig. 4). We also find no behavioral difference in reaction times on trials lying near equipoise compared to those that did not (1206±46 vs. 1283±43 ms; two-sided t-test, t(1023)= −1.17, p=0.24), and no correlation between the participants' reaction times and firing rate on a trial-by-trial basis (Pearson correlation; n=4723, r = 0.004, p = 0.78; Supplementary Fig. 5) to suggest a difficulty- or demand-related effect.

As noted above, a choice selection made near equipoise (e.g., ordinal position +1) and a choice selection made further away (e.g., ordinal position +2 or above) reflected differences in the degree to which the participants believed the scenes to be unsafe (i.e., even though the choice selections themselves are the same). Therefore, to further evaluate for neural modulation by these variations across the population, we used a two-way analysis of variance (2-way ANOVA, p<0.01) that considered the interaction between terms describing the participant's choices (safe vs. unsafe) and their relation equipoise (near vs. far). Using this evaluation, we find that the interaction between terms describing the participants choice

report and their relation to equipoise was significant (2-way ANOVA, $F_{(1,120)}=24.5$, $p<1\times10^{-5}$; choice $\times$ equipoise; Supplementary Table 4). Further *post-hoc* analyses demonstrated a significant effect for both variables (choice; $F_{(1,120)}=35.6$, 1×10^{-7} and equipoise; $F_{(1,120)}=12.3$, $p<0.001$; Fig. 4b). Therefore, taken together, the activities of these neurons appeared to reflect the participant's subjective decisions and their variation when considered across scenes.

Lack of neuronal modulation during sensory perceptual choices

Unlike neuronal responses observed during the situational assessment task, neurons in the dlPFC displayed little modulation during performance of an analogous sensory perceptual task [11, 12, 32, 33]. For example, it is possible that similar results may have been observed under any condition that required selection between two binary choice options. Therefore, to test for this, we next used a perceptual control task that involved small changes in the orientation of a Gabor [32] (Fig. 5a). The Gabor's orientation varied from −3 to 3 degrees across vertical and the participants were allowed to make two-forced choices on whether the Gabor was oriented to the left or the right. Here, we used this task because, similar to the situational assessment task, it was based on a two force-choice selection and because small variations in Gabor's orientation produced psychometric profiles that gradually transitioned from one choice to the other (Fig. 5c).

Of the 71 neurons recorded from the dlPFC during performance of this task, we find 32% (n = 23) neurons that responded to the Gabor presentation, meaning that they displayed either an increase or decrease in firing activity compared to baseline (i.e., prior to image presentation; paired t-test, $p<0.05$). Figure 5b illustrates an example of such a neuron. Of this population, however, only one cell displays a difference in their firing activity when the participants perceived the Gabor as tilting to the right (clockwise) vs. left (counterclockwise; Fig. 5b, middle and bottom panels). In other words, only 4.2% (n = 3) of the neurons display a difference in response to the Gabor as the participant's transitioned from one choice to the other across equipoise (Fig. 5c). This difference between the situational assessment task and the sensory perceptual task was significant (3/71 vs. 31/119 modulated cells respectively; Chi-square = 9.4, p = 0.0012).

Next, we also considered whether neuronal responses may reflect the participant's upcoming motor report. Subjective decisions are thought to be formulated early on during opinion formation; and prior to making the motor selection [34–36]. Using a linear discriminant, we find that decoding accuracy for the population (n=31) was highest during scene presentation, with an overall decoding accuracy of 69.2±1.2% (H_0=50%, z-value=4.3; two-sided Wilcoxon signrank test; p <0.0001). The mean time point of maximum decoding accuracy was approximately 1700±170 ms from image onset, and ranged between 1670±170 ms for trials lying near equipoise (ordinal position −1/+1) to 1770±170 ms for trials lying furthest away (two-sided paired *t*-test, t(30)=−0.41, p=0.69; Fig. 6).

We also examined the object control trials in which the same motor selections of safe/unsafe were made, but in which no contextual scene or situation was given. During object presentation, 19% (n=22) of the recorded neurons (Supplementary Fig. 6) were modulated by the participant's upcoming choice report (two- sided t-test, p<0.05) but, of these, only 9

cells responded to the participant's choices during both scene and object presentation (two-sided t-test, p<0.05). Moreover, as a population, a similar number of cells displayed a net increase *vs.* decrease in activity for items selected as unsafe (12 vs. 10 cells, Chi-square = 0.18, p = 0.67).

Variations in subjective decision between individuals

A second core feature of subjective decisions is that they often vary from person-to-person, even when presented with the same sensory information. For example, one participant may view a particular situation as safe, whereas another may view that same precise situation as unsafe (Fig. 2c) [1, 2, 4, 37]. Therefore, we next examined how closely neuronal activity may correlate with the participant's own unique voting profile when compared to that of the other participants. To do this, we performed a permutation procedure, similar to that described previously [14], in which we compared the same choice selections of safe and unsafe, but now substituted the recorded participant's voting profile with that of another participant (Fig. 7a). In other words, the same choice selections were evaluated, but their unique relation to equipoise within the voting profiles varied.

Here, we find that, when neuronal activity was taken from the participant recorded from but correlated with the voting profile of the other 10 participants, the fitted coefficients that described the regression were significantly lower (absolute 35% lower; permutation test; p=0.005; Fig. 7b). We also examined the choice selections themselves in which the trials were permuted across the participants. This maintained the same net choice comparisons of safe/unsafe, but produced a mismatch between the participant's subjective decision and the particular situation and scene in which it was made. Using this procedure, we similarly find that the difference in firing activity between the choices was 21% lower following permutation (permutation test, p=0.043; Fig. 7a). Therefore, when considering the same choice selections, the spiking activities of these neurons accounted for much of the variation in choice between individuals.

Effect of localized dlPFC disruption on subjective decision making

While the above observations suggested that neural responses in the dlPFC may reflect the participants' subjective decisions and their variation during situational assessments, they did not reveal whether or what causal contribution the dlPFC may have. Therefore, to next test this, we searched for additional subjects who had previously undergone focal resection of frontal cortical lesions for clinical reasons [38], and whose resection cavities were localized to Brodmann's area 9 of the dlPFC. Using these criteria, we identified 4 such participants (Supplementary Table 1). Their resection cavities measured 11.1 ± 2.7 cm^2 and were centered within the mid-portion of Brodmann's area 9 (Fig. 8a). We then compared performance of these 'lesioned' participants to that of the 11 control participants who had previously undergone neurosurgical access to the dlPFC by neuronal recordings but no lesion resection.

Individuals with focal dlPFC resections made largely similar choice selections to that of control. Specifically, the lesioned participants reported 42% of the scenes as safe whereas the controls reported 41% of the scenes as safe (1% difference in choice probability, with both groups having a slight bias towards reporting the scenes as unsafe, Chi-square=0.041,

p=0.83; Fig. 8b). Moreover, when comparing choices during the object presentations, the lesioned participants made the same selections as the control on almost all trials (95%, Chi-square=0.66, p=0.42). In other words, essentially all the participants from both groups agreed that a 'gun' is unsafe and that a 'shirt' is safe. The lesioned participants therefore appeared to understand the task and make largely appropriate choice selections.

Next, we examined the participants voting profiles. Here, we focused on the two main parameters that describe the participant's voting profile – the slope of the curve (i.e., at its point of inflection) and plateaus (i.e., its base and asymptote; see Methods). We then fit each of the participant's choices to a logistic function. We confirmed the choices of the two groups were indeed normally distributed and could therefore be appropriately fit (Lilliefors test, p>0.05). We also confirmed that the quality of the fits for the two groups were not significantly different (two-sided t-test comparing fit residuals, t(13)=0.56, p=0.58). Lastly, we demonstrated that similar results were obtained when evaluating the raw un-fitted behavioral data (Supplementary Fig. 7).

Individuals with dlPFC lesions displayed a diminished gradation between the opposing decisions. Overall, the control population displayed a range of voting profiles, with some individuals having a steeper slope than others (38 to 120 %). However, when compared to the lesioned participants (n=4), the average slope of the curve for the control group (n=11) was 72±12 % and significantly lower than 117±4 % for the lesioned participants (two-sided Wilcoxon test, p=0.026; Fig. 8b,c). In other words, the slopes for the lesioned participants were significantly higher. Similar results were obtained when calculating the slopes of the curves directly from the non-fitted behavioral data (two-sided Wilcoxon test, p=0.0015). Second, and largely in line with these findings, increase in the steepness of the curve was significantly positively correlated with the participant performance at the plateaus (Pearson correlation; n=15, r=0.81, p=0.00024; Fig. 8d). Disruption of the dlPFC therefore appeared to reduce variability of the participant's choices in relation to equipoise and to a diminish gradation in their voting profile.

The selectivity of dlPFC disruption on behavior

Although these findings suggested that lesions in the dlPFC led to reduced variability of the participant's choices in relation to equipoise, it was unclear whether similar observations may be found when performing other opinion-based tasks. For example, it is possible that these findings could have reflected processes specific to the evaluation of safety or perceived danger not observed in other situations. To test for this, we therefore included a second control task that involved evaluation of another's subjective trustworthiness. Here, small variations in face features taken from a validated database [39] produced psychometric profiles that gradually transitioned from the choice selection of trustworthy to untrustworthy across equipoise (Supplementary Fig. 8a). Therefore, like the situational assessment task, there were no explicitly correct choices under any condition. Unlike it, however, this task did not involve elements of safety or danger.

Here, we tested 5 participants with dlPFC lesions and 14 individuals with no lesions (Supplementary Table 2). Similar to our original task, we find that the participants with dlPFC lesions displayed a significant increase in the slope of their psychometric curves

when compared to the healthy participants (one-sided Wilcoxon test, z-value = 1.74, p = 0.04). We also find a significant increase in slope when compared to the control participants who had undergone neuronal recordings but not lesioning (one-sided Wilcoxon test, z-value = 2.25, p = 0.01; Supplementary Fig. 8b). Lesions in the dlPFC therefore appeared to lead to a consistently diminish gradation in the participant's voting profile when comparing these two tasks.

Unlike the opinion-based tasks, lesions in the dlFPC did not appear to affect the participants' choices when performing a sensory perceptual task. Using the same Gabor task described above [32], we find that the 5 participants with dlPFC lesions displayed no difference in the slope of their psychometric curves when compared to that of the participants with no lesion (two-sided Wilcoxon test, p = 0.96; Fig. 5d). Therefore, unlike for the opinion-based tasks, focal lesions in the dlPFC did not appear to affect the participants' ability to make sensory perceptual choices or enact their motor report.

Controls for motor-related behavior and neuropathology

Next, we also considered whether these findings could have been explained by other, more generalized, sensorimotor deficits related to surgical resection or underlying disease state. Overall, the lesioned participants demonstrated that they understood the task and made similar choice selections when compared to participants without dlPFC lesions. They also displayed no difference in performance on the object evaluations or when performing the sensory perceptual task suggesting that the effect of lesions in this area was selective. Nonetheless, to cast a wide net and to consider other possible deficits, we further performed four additional controls.

First, we examined the participants' choices at the extremes of the curve, at which the situations were most obviously safe or unsafe (e.g., ordinal positions +/−2 or above). If differences between the lesioned and control groups were due to simple unintended 'errors' or difference in choice consistency, we would expect selections for these manipulations to differ as well. However, we find no difference in selection on these trials between the participants (96.8±1.4% vs. 94.4±2.9%, two-sided t-test, t(46)=1.06, p = 0.29). We also find no difference in performances between the population that had undergone intraoperative neuronal recordings and healthy controls during these trials (95.2±1.5% vs. 94.4±2.9%, two-sided t-test, t(66)=0.58, p = 0.56). In other words, the patients displayed little variation in report when presented with the same scene on 'obvious' trials.

Second, we searched for differences that could be potentially explained by non-specific variations in task difficulty or demand. As mentioned before, the point of transition at equipoise was defined by the participant's own subjective evaluation of the scenes rather than by an extrinsic change in required response or attended scene feature (e.g., such as in a conflict monitoring- or countermand-type task) [29, 30, 40]. Nonetheless, to test for possible differences related to task difficulty, we examined whether the lesioned participants may be taking more time to make their selections and, therefore, displaying evidence of a demand-related changes near equipoise. However, we find no difference in response times between the lesioned participants and controls (two-sided t-test, t(13) = 0.74, p=0.45). We also find

no difference in response times on the *n*+1 trial after equipoise to suggest a potential difficulty-related Gratton effect [40] (two-sided *t*-test, t(21)=1.01, p=0.31).

Third, we considered cortical areas other than the dlPFC. Specifically, we examined whether resection in an upstream area that is broadly interconnected with the dlPFC and associated with higher-level sensory perceptual processing [41, 42], could produce a similar effect. Here, we searched for subjects who had previously undergone wide resection of the anterior temporal lobe (inferior/middle/superior gyri including amygdala and anterior hippocampus; ATL) and found 3 such participants (Supplementary Table 1). As with the control participants, individuals with ATL lesions displayed a range of voting profiles (Fig. 8e). Yet, unlike those with dlPFC lesions, the participants with ATL lesions displayed no difference in the slope of their voting profiles compared to control (68 ± 12 % vs. 72 ± 12 %; two-sided Wilcoxon test, p=0.23). The slopes of the voting profiles for the ATL lesioned participants were also lower than that of participants who had undergone dlPFC lesion resection (i.e., compared to 117 ± 4%; one-sided Wilcoxon test, p=0.03).

Last, we considered whether differences in the underlying disease state itself could have contributed to these results. Of the 11 patients who had undergone neuronal recordings during the situational assessment task, 2 had Parkinson's disease (PD), 7 had essential tremor (ET) and 2 had primary dystonia or chronic pain (OTHER). While the number of patients may be too small to find definitive trends, we observed here no significant difference in psychometric slopes between patients with PD and ET (slopes of 85 ± 17 % vs. 57 ± 15 %, respectively, two-sided Wilcoxon test, p = 0.50). We also find a largely similar proportion of neurons that were modulated by the task across the three participant groups (27%, 26% and 33% of cells, for PD, ET and OTHER respectively; Kruskal-Wallis: $H(2,8) = 0.778$, p = 0.68; Supplementary Fig. 9). Comparisons between individuals with different underlying dlPFC pathologies before resection also revealed similar results (e.g., glial vs. non-glial lesions; see Supplementary Fig. 10).

DISCUSSION

What neuronal processes underlie subjective decisions or what makes us change from one choice to another when forming opinions have long been intriguing questions in cognitive neuroscience, psychology and sociology [1, 4, 37, 43]. Here, we find neurons in the dlPFC that displayed spiking activities which changed gradually as the participants transitioned from one choice to another and that varied non-linearly, in a logistic manner, as the participants' transitions between choice options. In other words, the absolute difference in their activity for choices made near equipoise was proportionally larger than for choices made further away. Together, these observations are notable because they appear to be consistent with behavioral observations of how humans likely form their opinions, and describe how even small differences in a situation can produce a precipitous changes in choice [37, 44]. For instance, observing your friend having a small piece of cake after a large dinner may dramatically tip your opinion of them (e.g., as gluttonous) even though, calorically, adding a small piece of desert accounts for only a minor proportion of the meal. Here, changes in the spiking activities of these neurons appeared to reflect both this subjective point of transition as well as the specific decision being made.

Supporting these observations, individuals with focal lesions of the dlPFC displayed a loss of gradation in their psychometric profiles and near-unitary transition across equipoise. By comparison, they displayed little difference in their choice selections or motor responses when compared to non-lesioned participants, suggesting that these changes were not explained by a non-specific sensory or motor deficit. This dissociation was also not explained by difficulty with task switching or in adaptively changing between choices [45, 46], as these transitions occurred randomly throughout the task and were not instructed.

Such behavior, in turn, whereby there is little gradation between opposing choice options, is often referred to in decision theory as false dichotomous or black-and-white thinking. Individuals with such tendency commonly consider choice options unambiguously as 'right and wrong' or 'good and evil' and have difficulty in considering intermediate possibilities; especially under circumstances in which there is no explicit correct choice [20, 47, 48]. While such a deficit would not be apparent when examining individual choices, they are apparent when evaluating the participant's complete voting profiles. The ability of humans to consider choice options in a continuous or non-dichotomous manner and to allow for variability in their decisions are important features that define humans decision making and the process by which opinions are likely formed [1, 2, 4]. Here, our findings suggest that the human dlPFC plays an important role in this process and explains how it may contribute to this choice variability.

These findings also reveal an important distinction between how subjective decisions and perceptual choices are likely processed in the human dlPFC. Prior animal studies, for instance, have shown differences in the activities of neurons in frontal and parietal areas when making sensory perceptual choices on the apparent motion of stimuli, their tactile discrimination or source location [8–13, 49]. They have also shown differences in activity that correlate with the proportion of dots moving coherently in one direction over another or the amount of auditory clicks coming from a particular location; as predicted from signal-detection-theory [50]. Opinion-based tasks such as the ones tested here, by comparison, test for variations in subjective judgment or valuation [16–21], and for which there is no intrinsic correct choice for any situation; whether the participant's choices lie near equipoise or not [1–3]. Consistent with this, we find that neurons in the human dlPFC gradually changed their activities in relation to equipoise during the situational assessments even though the choice reports were the same. Moreover, they displayed a notable difference in response and lesion effects when comparing the situational assessment to the sensory perceptual Gabor task.

Collectively, our data suggest a neuronal process in the dlPFC that may support subjective decision making in humans. In particular, they suggest a process that could allow single neurons to encode both a graded evaluation of the available choice options as well as their unique point of transition at equipoise and, therefore, allow them to reflect an individual's complete voting profile. Disrupting this process, in turn, would lead to a loss of gradation between opposing choice options but would not impact the primary mechanisms by which sensory information is processed or motor selections are made. It would also not impact more generalized cognitive process, such as demand adaptation or sensory perceptual processing. These observations on how subjective decisions are made during opinion formation could help improve our understanding of the mechanisms that underlie human

behavior and its variability [5, 6, 40]. They could also help place more informed constraints on operational models that aim to explain human behavior and their disruption in psychological and social disorders.

METHODS

Study subjects and enrollment

All aspects of the study were approved by the Massachusetts General Hospital Institutional Review Board (IRB) and were held in strict accordance with Harvard Medical School ethical guidelines. For neuronal recordings, participants underwent intraoperative neurophysiology as part of their planned deep brain stimulator (DBS) placement [51–53]. Prior to consideration, candidates for the study were evaluated by a multidisciplinary team of neurologists, neurosurgeons, and neuropsychologists [24, 40] and decisions for surgery were unrelated to study participation. After surgical consent was obtained and the participants were scheduled for surgery, an independent member of the research team approached the patient for study participation. They then filled a separate research consent if they wished to participate in the study. At any point in the study, including during the intraoperative phase, patients were freely able to withdraw from the study without any consequence to their clinical care. Participants included for lesion evaluation were identified retrospectively. Decision for lesion resection was made for clinical reasons and had no relation to the present study. None of the participants demonstrated a focal neurological deficit on neurological exam (e.g., paraplegia) or demonstrated a language or comprehension deficit. A total of 39 participants were included in the study (Supplementary Table 1 and Supplementary Table 2). Evaluation of the relationship between underlying pathology, task performance and neuronal recording results are further described in Supplementary Figures 7–10 further below.

Neurophysiology and dlPFC targeting

Acute neuronal recording—Subjects partaking in planned DBS at our institution normally also undergo standardized micro-electrode recording in order to optimize anatomical targeting [24, 51]. These micro-electrodes sometimes traverse the dorsal lateral surface of the prefrontal cortex on way to the target nucleus (e.g., the subthalamic nucleus or thalamus) and, therefore, provide the opportunity to briefly study neuronal responses in the area. These recordings do not perturb the planned operative approach or alter clinical care [24, 34, 40].

To perform single-neuronal recordings from the dlPFC, we first incorporated an FDA approved, biodegradable, fibrin sealant (Tisseal) between the cortical surface and the inner table of the skull [24]. The sealant is normally used after DBS placement but, in our setting, placement before micro-electrode targeting further allowed for cortical pulsations to be further locally mitigated (Fig. 1a). Second, we incrementally advanced the micro-electrodes along the cortical ribbon at 10–100 μm increments in order to identify and isolate individual units (i.e., rather than focusing on the subcortical structures alone). Here, we employed the same array of 5 tungsten microelectrodes (500–1,500kOhm; FHC) normally used for deep targeting (Alpha Omega Engineering, Nazareth, Israel). The micro-drive did not have independent control over each electrode. Once putative neurons were identified, the

microelectrodes were held in position for 4–5 minutes in order to confirm signal stability (we did not screen putative neurons for task responsiveness). Lastly, we used the intraoperative MER system and I/O DAQ that allowed us to precisely time-stamp task events (1kHz) and sample the neuronal data (44 kHz) at millisecond resolution. Neuronal signals were amplified, bandpass filtered (300 Hz and 6 kHz) and stored off-line (Alpha Omega Engineering, Nazareth, Israel). After recording from the dlPFC, subcortical neuronal recordings and DBS placement proceeded as scheduled (see below).

Single-unit isolation—Single-units were identified and sorted off-line through a Plexon work station (Plexon Inc. Dallas TX). To ensure the identification of single, well-isolated units, we first constructed a histogram of peak heights from the raw voltage tracings on each channel. A minimum threshold of three standard deviations was used to differentiate between neural signals from background noise. Next, template matching and principal component analyses (PCA) were used to classify action potentials and sort neurons. An analysis of variance performed on the first three PCA components was used to discriminate between waveform morphologies in which more than one unit was present. Candidate clusters of putative neurons needed to clearly separate from channel noise, display a voltage waveform consistent with that of a cortical neuron, and to have at least 99% of action potentials separated by an inter-spike interval of at least 1 ms (although most neurons displayed an ISI of 2 ms or more; Fig. 3a *inset*). Finally, any units that did not demonstrate waveform stability over the course of the trial were excluded from analysis (Supplementary Fig. 1a). No multi-units were used. Considering each patient underwent an average of 2 recording sessions, we recorded from an average of 1.6±0.2 well-isolated single units *per* microelectrode. Supplementary Fig. 1b illustrates two examples of spike waveform morphologies and associated PCA clusters. In the first example, only a single unit was isolated from the recorded trace. In the second example, two units were isolated (red and green, the baseline noise is displayed in gray). Any prospective units that displayed significant overlap in their PCA distributions by MANOVA ($p < 0.01$) or overlapped with the baseline signal/noise were considered multi-units and excluded from further analyses.

Recording and lesion site confirmation—All subjects involved in neuronal recordings underwent high-resolution 1.5 or 3-Telsa magnetic resonance imaging (MRI) scan and computed tomographic (CT) scans prior to surgery. Post-operatively, the same subjects underwent a repeat CT scan to confirm placement of the DBS electrodes. In order to confirm recording locations, we referenced the location of the burr hole and DBS lead. Recording sites for all participants in which neuronal recordings were made were within the middle frontal gyrus of Brodmann's area 9 of the dlPFC (Fig. 1).

Subjects with focal dlPFC lesions were identified by retrospectively reviewing patients at our institution who had previously undergone craniotomy for frontal or temporal tumors within two years of study participation. Post-operative, contrast-enhanced MRIs were used to identify the location and extent of the resection cavities. Exclusion criteria included resection cavities that were associated with sub-total resection (i.e., the lesion was not entirely removed), the resection cavities encroached on surrounding cortical territories or transcortical tracts (e.g., the medial prefrontal cortex) or the lesions were associated with

significant edema at the time of evaluation. Based on these criteria, and for participants performing the subjective situational assessment task, we identified a subset of 4 individuals with solitary, focal resection cavities localized to the dlPFC (Fig. 8a). The average area of the resection cavities at the cortical surface was 11.1 ± 2.7 cm^2. We also separately identified 3 additional individuals with resection cavities localized to the ATL. These individuals underwent complete resection of the entire anterior superior, middle and inferior temporal gyri. Description of lesion location and underlying pathologies for all participants can be found in Supplementary Tables 1 & 2.

Behavioral Tasks

Subjective situational assessment: Each trial began with a blank screen for 1 second (the baseline period). Next, a scene was presented for 2 seconds followed by blank screen delay for another 1 second. After this, two targets representing the choice options of 'safe' and 'unsafe' were presented in random location on each end of the screen (Fig. 2a). The subject had up to 10 seconds to select his or her choice by moving the joystick towards the safe or unsafe target option (Fig. 1b). For example, if the target option of safe was presented at the bottom of the screen, and the participant believed the scene to represent a safe situation, the participant would move the joystick downwards. However, if the target option of safe was presented at the top of the screen, he or she would move the joystick upwards. All scenes and target locations were given pseudo-randomly and in counterbalanced fashion over the course of the session. No feedback was given at any point during the trial.

A common pool of 70 images was used in the study and consisted of three main variations. First, 35 of the 70 images were divided into seven distinct prototypic scenes; each with their own unique situation and theme. For example, the participant may see a jogger running near a car in one scene but may see a person standing in a construction yard in another (Fig. 1b). Second, each of the prototypic scenes consisted of a subject that varied in relation to an object. For example, the relationship between the jogger (subject) and the car (object) on the street may vary from trial-to-trial for a total of five variations per prototypic scene (Fig. 1b). Finally, the other 35 images consisted of scenes in which the object itself was identifiable as safe or unsafe. For example, the participant may be given the image of a gun. In total, these variations were introduced to allow us to test the participant's choices across a variety of real-world scenarios, study the particular points at which the participant's transitioned from one choice to another across scenes, and compare variations in the choices of the different participants [22,23].

Subjective trustworthiness assessment: While the main task used situational safety assessments to evaluate for variations in opinion, it is possible that our results may have been specific to tasks that only involve safety or danger. Therefore, to further test for this possibility, we included an additional control task that required the participants to render an opinion on trustworthiness. Guided by prior literature, the participants were shown faces taken from a validated database [39], whereby small variations in the features of the faces produced psychometric profiles that gradually transitioned from trustworthy to untrustworthy (Supplementary Fig. 8a). The faces presented to the participants were taken from a common pool of 15 faces. These faces were created from four pairs of prototypic

faces that varied both in gender and facial features (see [39] for additional detail). Each pair was then manipulated to produce three additional intermediates for a total of five variations per prototypic face pair (three sets of five faces). Faces were presented pseudo-randomly and in counterbalanced fashion (~ 3 times repeated on average) over the course of the session. The trial sequence was also similar to the subjective situational assessment task. Here, each trial began with a blank screen for 1 second. A face was then presented for 2 seconds followed by a blank screen delay for another 1 second. After this, two targets representing the choice options of 'trustworthy' and 'non-trustworthy' were presented in random location on each end of the screen. The subject had up to 10 seconds to select his or her choice by moving the joystick towards one of the two options (Fig. 1b). No feedback was given at any point during the trial.

Sensory perceptual Gabor task: To further evaluate for selectivity of effect, we included an additional control task that required the participants to perform a sensory perceptual task. Here, the participants were presented with a Gabor image that was slightly rotated in leftward (counterclockwise) or rightward (clockwise) direction (Fig. 5a). The Gabor's orientation varied from −3 to 3 degrees across vertical, and the participants were allowed to make forced choices on whether the Gabor was rotated to the left or right. While this task produced similar psychometric profiles as in the situational assessment task (Fig. 5c), the participant's choices were explicitly correct or incorrect. Similar to the main task, each trial began with a blank screen for 1 second followed by a Gabor for 2 seconds and then a blank screen delay for another 1 second. Finally, two targets representing the choice options of 'left' and 'right' were randomly presented on each end of the screen. The subject then had up to 10 seconds to select his or her choice by moving a joystick. Overall, the participants were presented with 8 Gabors of different orientations (Figure 5a), which were given in pseudo-random fashion (each Gabor was presented 4 times on average) during the course of each session. These ranged from most ambiguous (+/− 3.0 degrees) to least ambiguous (+/− 0.2 degrees). No feedback was given at any point during the trial.

Participant instruction: For participants undergoing neuronal recordings, the subjects were familiarized with the task on the morning of surgery. For healthy participants or participants with dlPFC/ATL lesions, the subjects were familiarized with the task prior to testing. All subjects were given instruction prior to practice. For example, for the situational assessment task, they were told that: "1) You will be shown a series of images on the computer screen displaying a scene or an object. 2) You will need to decide whether each scene or object is safe or unsafe. 3) To indicate your choice, please wait for the target choice to appear at the end of each trial and move the joystick to the displayed 'safe' target if you believe the scene or object is safe and to the displayed 'unsafe' target if you believe it is unsafe." Each subject practiced for 5–10 minutes prior to performing the task. All behavioral tasks were performed and recorded using customized software written in MATLAB (MathWorks, Inc., Natick, MA, USA) [54].

Statistical Analysis

Psychometric curves and voting profile analysis—The voting profiles of the participants were constructed in three steps. For the situational assessment task, we

constructed a psychometric curve based on the choices of the participant for each specific scene and manipulation. For example, as can be seen in Figure 2b, the choice responses of one of the participants to prototypic scene #5 were *safe-safe-safe-unsafe-unsafe* across the five manipulations #1–5. Second, the choices of each participant across the 7 prototypic scenes were aligned to the point at which the participant's choices transitioned from safe to unsafe. This point of transition was determined through a 2-point moving average function (i.e., neighboring choices) in which the smoothed value was closest to 0.5. Essentially identical results were obtained through a formal state-space approach [8, 55]. Thus, for example, if the manipulations 1–5 for prototypic scene 1 lead to choices of *safe-safe-safe-unsafe-unsafe*, whereas the manipulations 1–5 for prototypic scene 2 lead to choices of *safe-unsafe-unsafe-unsafe-unsafe*, their point of equipoise would be between manipulations 3–4 and 1–2 respectively. For certain cases, there was variation in choice for adjacent manipulations (e.g., *safe-safe-unsafe-safe-unsafe*) or on control duplicate manipulations (e.g., *safe-safe-[safe/unsafe]-safe-unsafe*) within a particular prototypic scene. Here, the point of transition was calculated in the same fashion as above. In other words, the point of transition was based on the net 'weight' of all choice selections; such that the most equal set of selections laid on the 'left' and 'right' of equipoise. Last, we aligned these points of transition across all the prototypic scenes *per* participant and calculated their average in order to obtain the psychometric curve. This represented the participant's voting profile across the different manipulations; with ordinal positions above zero representing scenes that the participant believed were progressively safe and ordinal positions below zero representing scenes that the participant believed were progressively unsafe (Fig. 2d). For the figures, psychometric profiles are displayed with each of the participant's data points. The box-and-whisker plots display the median and quantiles.

In order to further evaluate for differences in the participants voting profiles, we fit their choices to;

$$y = B + \frac{A - B}{1 + e^{-k(x - x_0)}} \quad (1)$$

where *y* is the overall probability of the participant choosing 'unsafe'; *A* and *B* are the asymptotic values of *y* in the extreme 'unsafe' and 'safe' domains of the logistic curve, respectively; *k* is the steepness of the curve. Lastly, x_0 represents the point of inflection at equipoise. As done previously, these fits were optimized to minimize the root-mean-square error between the simulated and the experimental data for each participant [55].

For further validation, we also calculated *k* from the raw behavioral data. Here, we considered the participant's choice probability at ordinal positions −1 and +1 (i.e., the values closest to equipoise) and calculated the slope of the curve simply as the mean difference in their values. The values of *A* and *B* were calculated as the raw choice probability at the two plateaus of the curve.

Task modulation and neurometric analysis—Peri-stimulus histograms and rasters were constructed for all units that displayed stable waveform parameters, adequate refractory periods and morphology. To allow for comparisons between cells with different firing rates, trial activity was normalized (divided by) the average firing rate for the cells at baseline (1000 ms prior to image onset). Because of the known temporal delay of frontal neurons to sensory stimuli [34, 46], we focused on a 1500 ms window starting 500 ms from the onset of scene presentation and lasting for the duration of scene presentation. Differences in the firing rates of the cells to the choices of safe/unsafe were made through a standard two-sided Student's t-test (p<0.05; with Bonferroni correction). A two-way analysis of variance was used to examine differences in the firing rates of the cells in relation to choice (safe/unsafe) and equipoise (near/far; 2-way ANOVA, p<0.05). For simplicity and to provide an identical number of trials to compare, we defined 'near' equipoise as ordinal positions [−2,−1,+1+2] and 'far' from equipoise as ordinal positions [−4,−3,+3,+4] (although largely identical results were obtained by comparing [−1,+1] to the other positions).

To analyze differences in firing rate as they related to the participant's voting profile, we used two principal procedures. For the first, we fit three distinct functions to the neuronal activities: 1) a logistic function similar to what we used above for the psychometric curves where y is the neuronal activity, B and A represents the minimum and maximum firing rates respectively, k is the steepness of the curve, and x_0 represents the point of inflection, 2) a binary step function that only included B and A, representing the minimum and maximum neuronal activity, respectively, on each side of equipoise, and 3) a u-shaped (i.e., parabolic) function. Here, the u-shaped function was defined by:

$$y = c(x - x_0)^2 + m \quad (2)$$

where y is the neuronal activity, the parameter c determines the steepness of the parabola as well as its direction (upward if $c > 0$ and downward if $c < 0$), and m represents the participant's maximum (or minimum) probability of choosing unsafe depending on whether the parabola is upward or downward. Lastly, x_0 represents the neutral ordinal position (i.e., equipoise).

These fits were optimized to minimize the root-mean-square error between the simulated and the experimental data for each neuron. The goodness of fit was evaluated using the variance accounted for (*VAF*),

$$VAF = 1 - \frac{var(A - \widehat{A})}{var(A)} \quad (3)$$

Where *var* is the variance, *A* is the actual neuronal activity and $\widehat{A}$ is the estimated neuronal activity using either the logistic or the binary step function.

For the second approach, we performed a model selection procedure that progressively considered the neuronal activity, point of inflection and slope to fit the experimental data. These successively included a unitary fit (which was essentially $(A+B)/2$ in equation 1, a linear fit (which described the linear slope between A and B and contains a constant and a slope parameter), a logistic fit (which included $e^{-k(x-x_0)}$ in equation 1), and a u-shaped fit according to the equation 2. To determine which model best fits the data, we used both an Akaike Information Criterion (AIC) and a Bayesian Information Criterion (BIC) which penalize models containing more parameters (Supplementary Table 3). AIC was computed as: $AIC = 2k - 2\ln(L)$ where L is the maximum value of the likelihood function of a given model, and k is the number of free parameters in the model. Similarly, BIC was calculated as: $BIC = \ln(n)k - 2\ln(L)$ where L is the is maximum value of the likelihood function of a given model, k is the number of free parameters in the model, and n is the number of data points.

Substitution analysis—A substitution analysis was performed to quantify the degree to which neural activity correlated with each participant's own unique voting profile. This procedure was done in four parts. First, we would identify neuronal activity recorded in one participant (e.g., participant #1) when he or she was viewing a particular scene and then determine when, during the session, another participant (e.g., participant #2) was viewing the same precise scene. Second, we would maintain the trial-by-trial relation between the neuronal activity and the specific manipulation (i.e., manipulations 1–5 in Fig. 1b) *per* prototypic scene from the recorded participant #1, but the choice reports to those manipulations would now be assigned to the choice reports of participant #2. Third, in order to ensure an even sample across participants and an identical net proportion of safe/unsafe selections comparison, we pseudo-randomly repeated this procedure 1000 times (i.e., the choices were preserved on average across all 1000 shuffles). Thus for example, participant #1 may view manipulation 2 of scene 1 as safe whereas participant #2 may view that same manipulation/scene as unsafe. However, whereas participant #2 may view that manipulation/ scene as unsafe, participant #3 may view it as safe, *etc*. Therefore, even though we were comparing the same choice of 'safe' across shuffles, the participant's subjective evaluation of the scenes, as inferred from their psychometric profiles, differed. Lastly, to examine the degree to which neural activity correlated with the participant's unique voting profiles, we calculated the β logistic regression coefficients which described the correlation between neural activity and the original and substituted voting profiles. We also examined for differences in raw firing rates for the safe/unsafe choices between the original and substituted trials (permutation test, p<0.05).

Neuronal population decoding—A Fisher's discriminant was used to quantify the degree to which population activity was informative of the participant's upcoming choice. As described previously [55], we quantitatively measured the ratio of the variance in neuronal activity between the two group options of 'safe' and 'unsafe' to the variance within the groups based on; $S_w^{-1}S_B v = \lambda v$, whereby S_w and S_B are the within group scatter matrices and between group scatter matrices, respectively. The prediction vector v, corresponds to the largest eigenvalue of the matrix on the left-hand side of the equation. The prediction vector

defines a projection of the recorded activity into a scalar unit that is then compared to a threshold, θ, and if greater than θ, the trial choice was predicted to be safe and unsafe if less than θ. For validation, we divided the data into a training set consisting of 75% of the trials and tested the accuracy of the prediction on the remaining 25% of trials. This operation was repeated 1000 times using a random sampling of the total trials. Finally, in order to examine the timing of maximum decoding accuracy, we used a 1500 ms sliding window advanced in 100 ms increments.

Statistics—Throughout the paper, we report mean and standard error of the mean unless otherwise noted. Correlation analyses were conducted using Pearson's *r* value. Testing for differences of means of distributions was performed using non-parametric Wilcoxon ranksum test, unless parametric *t*-test or analysis of variance testing is specified. No data points were excluded from analyses unless otherwise stated. Parametric testing was used when data was assumed to be normally distribution. This was not formally tested, however. If data was not assumed to be normally distributed, non-parametric tests were used. When box-and-whisker plots are displayed, the central line represents the median of the distribution, box limits represent 25th and 75th quantiles, and whisker limits represent the full range of data. Sample sizes were not predetermined using power analysis or other statistical tests, but the sample sizes acquired are comparable to those in previously reported work [34, 40, 46]. Data collection and analysis were not performed in a blinded fashion.

REPORTING SUMMARY

Further information and details regarding the experimental design can be found in the Life Sciences Reporting Summary.

DATA AVAILABILITY:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY:

The primary codes used to analyze the data are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Zaller J The nature and origins of mass opinion. (Cambridge University Press, Cambridge England; New York, NY, USA; 1992).

2. Dryzek JS Discursive democracy : politics, policy, and political science. (Cambridge University Press, Cambridge; New York; 1990).

3. Mara GM Socrates' discursive democracy : logos and ergon in Platonic political philosophy. (State University of New York Press, Albany; 1997).

4. Turner CF, Martin E & National Research Council (U.S.). Panel on Survey Measurement of Subjective Phenomena Surveying subjective phenomena. (Russell Sage Foundation, New York; 1984).

5. Pesaran B, Nelson MJ & Andersen RA Free choice activates a decision circuit between frontal and parietal cortex. Nature 453, 406–409 (2008). [PubMed: 18418380]

6. Mante V, Sussillo D, Shenoy KV & Newsome WT Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84 (2013). [PubMed: 24201281]

7. Hanks TD et al. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature 520, 220–223 (2015). [PubMed: 25600270]

8. Williams ZM & Eskandar EN Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat Neurosci 9, 562–568 (2006). [PubMed: 16501567]

9. Yartsev MM, Hanks TD, Yoon AM & Brody CD Causal contribution and dynamical encoding in the striatum during evidence accumulation. Elife 7 (2018).

10. Kim JN & Shadlen MN Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci 2, 176–185 (1999). [PubMed: 10195203]

11. Odegaard B et al. Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. Proc Natl Acad Sci U S A 115, E1588–E1597 (2018). [PubMed: 29382765]

12. Romo R & de Lafuente V Conversion of sensory signals into perceptual decisions. Prog Neurobiol 103, 41–75 (2013). [PubMed: 22472964]

13. de Lafuente V & Romo R Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. Proc Natl Acad Sci U S A 108, 19767–19771 (2011). [PubMed: 22106310]

14. Padoa-Schioppa C & Assad JA Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226 (2006). [PubMed: 16633341]

15. Cai X & Padoa-Schioppa C Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. Neuron 81, 1140–1151 (2014). [PubMed: 24529981]

16. Koenigs M et al. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446, 908–911 (2007). [PubMed: 17377536]

17. Chung D, Christopoulos GI, King-Casas B, Ball SB & Chiu PH Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. Nat Neurosci 18, 912–916 (2015). [PubMed: 25984890]

18. Kable JW & Glimcher PW The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10, 1625–1633 (2007). [PubMed: 17982449]

19. Shenhav A & Greene JD Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. Neuron 67, 667–677 (2010). [PubMed: 20797542]

20. Sternberg RJ Cognitive psychology, Edn. 3rd (Thomson/Wadsworth, Australia Belmont, CA; 2003).

21. Weekley JA & Ployhart RE Situational judgment tests : theory, measurement, and application. (Lawrence Erlbaum Associates, Publishers, Mahwah, N.J.; 2006).

22. Nowinski WL, Gupta V, Qian G, Ambrosius W & Kazmierski R Population-based Stroke Atlas for outcome prediction: method and preliminary results for ischemic stroke from CT. PLoS One 9, e102048 (2014). [PubMed: 25121979]

23. Ndode-Ekane XE, Kharatishvili I & Pitkanen A Unfolded Maps for Quantitative Analysis of Cortical Lesion Location and Extent after Traumatic Brain Injury. J Neurotrauma (2016).

24. Patel SR et al. Studying task-related activity of individual neurons in the human brain. Nat Protoc 8, 949–957 (2013). [PubMed: 23598445]

25. Zedeck S & American Psychological Association. APA handbook of industrial and organizational psychology, Edn. 1st (American Psychological Association, Washington, DC; 2011).

26. Rainer G, Asaad WF & Miller EK Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature 393, 577–579 (1998). [PubMed: 9634233]

27. Haroush K & Williams ZM Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates. Cell (2015).

28. Kepecs A, Uchida N, Zariwala HA & Mainen ZF Neural correlates, computation and behavioural impact of decision confidence. Nature 455, 227–231 (2008). [PubMed: 18690210]

29. Hanes DP, Patterson WF 2nd, & Schall JD Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. J Neurophysiol 79, 817–834 (1998). [PubMed: 9463444]

30. Freeman E, Sagi D & Driver J Lateral interactions between targets and flankers in low-level vision depend on attention to the flankers. Nat Neurosci 4, 1032–1036 (2001). [PubMed: 11559851]

31. Padoa-Schioppa C & Assad JA The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat Neurosci 11, 95–102 (2008). [PubMed: 18066060]

32. Summerfield C, Behrens TE & Koechlin E Perceptual classification in a rapidly changing environment. Neuron 71, 725–736 (2011). [PubMed: 21867887]

33. Williams ZM, Elfar JC, Eskandar EN, Toth LJ & Assad JA Parietal activity and the perceived direction of ambiguous apparent motion. Nat Neurosci 6, 616–623 (2003). [PubMed: 12730699]

34. Mian MK et al. Encoding of Rules by Neurons in the Human Dorsolateral Prefrontal Cortex. Cereb Cortex (2012).

35. Rigotti M et al. The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590 (2013). [PubMed: 23685452]

36. Fusi S, Asaad WF, Miller EK & Wang XJ A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. Neuron 54, 319–333 (2007). [PubMed: 17442251]

37. Zuckerman M Psychobiology of personality, Edn. 2nd (Cambridge University Press, Cambridge, UK; New York, N.Y.; 2005).

38. Porter KR, McCarthy BJ, Freels S, Kim Y & Davis FG Prevalence estimates for primary brain tumors in the United States by age, gender, behavior, and histology. Neuro Oncol 12, 520–527 (2010). [PubMed: 20511189]

39. Sofer C, Dotsch R, Wigboldus DH & Todorov A What is typical is good: the influence of face typicality on perceived trustworthiness. Psychol Sci 26, 39–47 (2015). [PubMed: 25512052]

40. Sheth SA et al. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature 488, 218–221 (2012). [PubMed: 22722841]

41. Freiwald WA, Tsao DY & Livingstone MS A face feature space in the macaque temporal lobe. Nat Neurosci 12, 1187–1196 (2009). [PubMed: 19668199]

42. Hirabayashi T, Takeuchi D, Tamura K & Miyashita Y Microcircuits for hierarchical elaboration of object coding across primate temporal areas. Science 341, 191–195 (2013). [PubMed: 23846902]

43. Siddique Z, Anand S & Lewis-Greene H Situational judgment tests for dentists : the DF1 guidebook.

44. Eysenck HJ The psychology of politics. (Routledge and K. Paul, London,; 1954).

45. Rushworth MF, Hadland KA, Gaffan D & Passingham RE The effect of cingulate cortex lesions on task switching and working memory. J Cogn Neurosci 15, 338–353 (2003). [PubMed: 12729487]

46. Williams ZM, Bush G, Rauch SL, Cosgrove GR & Eskandar EN Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nat Neurosci 7, 1370–1375 (2004). [PubMed: 15558064]

47. Wenzel A, Chapman JE, Newman CF, Beck AT & Brown GK Hypothesized mechanisms of change in cognitive therapy for borderline personality disorder. J Clin Psychol 62, 503–516 (2006). [PubMed: 16470716]

48. Eysenck HJ The psychology of politics. (Transaction Publishers, New Brunswick, N.J., U.S.A.; 1999).

49. Asaad WF, Rainer G & Miller EK Neural activity in the primate prefrontal cortex during associative learning. Neuron 21, 1399–1407 (1998). [PubMed: 9883732]

50. Green DM & Swets JA Signal detection theory and psychophysics. (R. E. Krieger Pub. Co., Huntington, N.Y.,; 1974).

METHODS REFERENCES

51. Amirnovin R, Williams ZM, Cosgrove GR & Eskandar EN Experience with microelectrode guided subthalamic nucleus deep brain stimulation. Neurosurgery 58, ONS96–102; discussion ONS196–102 (2006). [PubMed: 16543878]

52. Gross RE, Krack P, Rodriguez-Oroz MC, Rezai AR & Benabid AL Electrophysiological mapping for the implantation of deep brain stimulators for Parkinson's disease and tremor. Mov Disord 21 Suppl 14, S259–283 (2006). [PubMed: 16810720]

53. Theodosopoulos PV, Marks WJ Jr., Christine C & Starr PA Locations of movement-related cells in the human subthalamic nucleus in Parkinson's disease. Mov Disord 18, 791–798 (2003). [PubMed: 12815658]

54. Asaad WF & Eskandar EN Achieving behavioral control with millisecond resolution in a high-level programming environment. J Neurosci Methods 173, 235–240 (2008). [PubMed: 18606188]

55. Wasserman L All of statistics: a concise course in statistical inference. (Springer Texts, New York, NY; 2005).
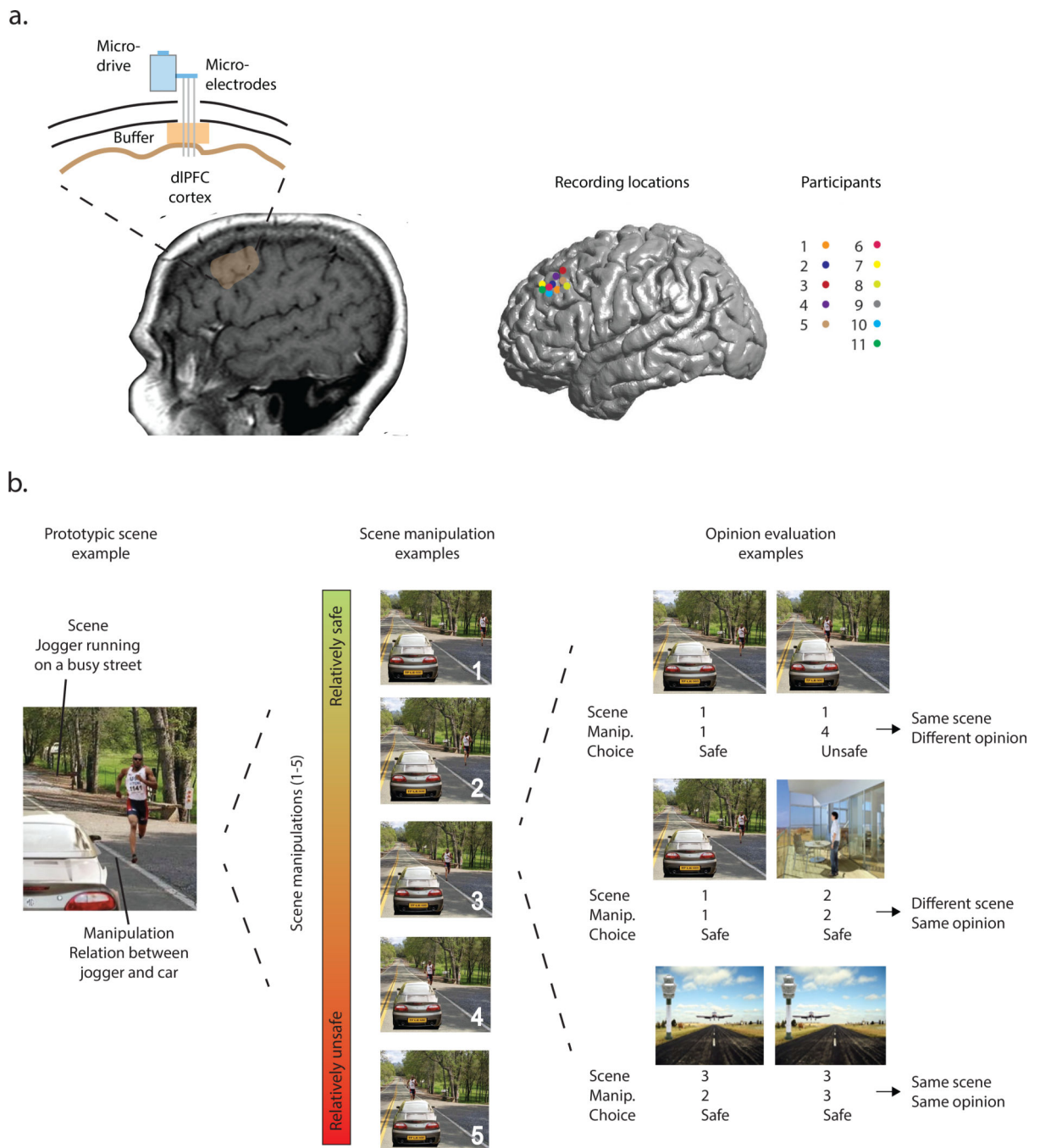
Figure 1. Single-neuronal recordings from the dlPFC during performance of a naturalistic situational assessment task.
(**a**) Individual units were acutely recorded from the dlPFC of subjects undergoing planned intraoperative neurophysiology (see Methods). To the *right* are the specific recording locations obtained across the eleven participants. (**b**) Situational assessments were used to evaluate for variations in the participants subjective decisions and their point of transition. During the task, the participants were presented with scenes depicting a real-world scenario and were required to indicate whether they believed the scenes to be safe or unsafe. Seven

prototypic scenes were used, each with their own unique theme. Each of the prototypic scenes was manipulated so as to vary the degree to which it represented a safe or unsafe situation (e.g., the relation between a jogger running on a street and a car passing near would vary from trial-to-trial). Trials were given randomly throughout the session and no feedback was given at any time to indicate which choice was 'correct'. On the far *right* are examples of different scenes, manipulations and choices made by a participant. As discussed in the main text, additional comparisons included scenes in which only an object was presented.
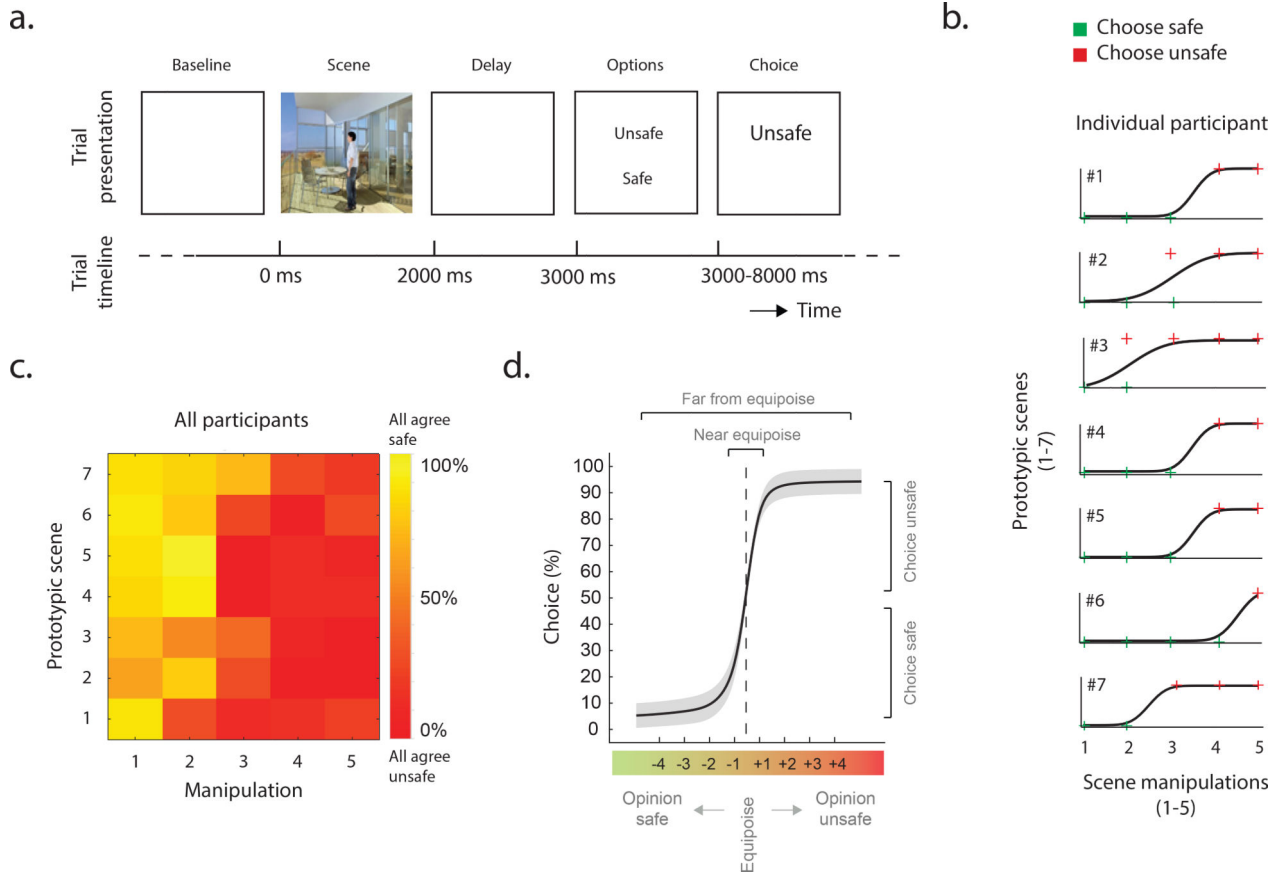
Figure 2. Trial structure, psychometric performances, and variation in subjective decisions within/between the participants.

(**a**) On each trial, after scene presentation, the participants would be given two randomly displayed choice options. (**b**) Examples of choices made by one of the participants (n = 11) for prototypic scenes 1–7. The x-axis reflects manipulations 1–5, and the y-axis reflects the participant's reported choice of safe or unsafe (colored green and red, respectively). Here, points in which a safe and unsafe choice are vertically opposed (i.e., one above another) represent manipulations in which the participant's made a different choice on separate trials. (**c**) The choices of all the participants across all prototypic scenes and manipulations. (**d**) The voting profile (black line) of the eleven participants and their corresponding s.e.m. (shaded grey band). Here, the x-axis represents the participant's choices aligned to equipoise. Ordinal values below zero represent scenes that the participant's believed were progressively more safe and values above zero represent scenes that the participant's believed were progressively more unsafe. The y-axis reflects the overall probability of reporting the scene as unsafe. The point of transition at equipoise was based on the participant's own subjective evaluation of the different, randomly-presented scenes.
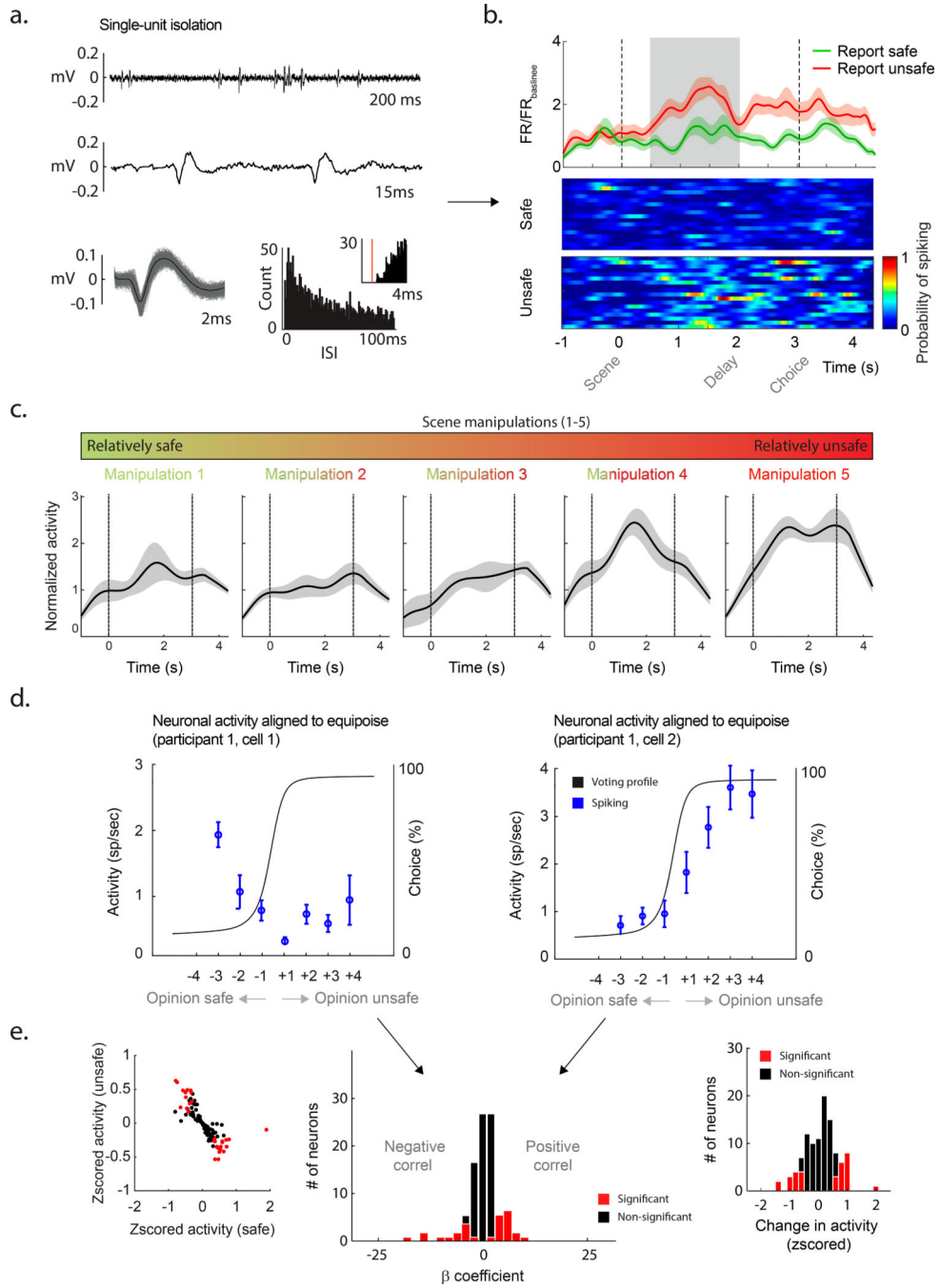
Figure 3. Neuronal responses in the dlPFC correlate with variations in the participant's subjective decisions.

(**a**) A representative neuron recorded from the dlPFC (total population n = 119). The raw voltage tracings are displayed *above* and the waveform morphology and inter-spike-intervals are displayed below (the *inset* is in log scale to emphasize the refractory period; the red line is at 1 ms). Supplementary Fig. 1 provides additional illustration of waveform stability over time and their isolations. (**b**) Peri-stimulus spike histogram and trial heatmap of the same cell during task performance for trials reported as safe (green line) versus unsafe (red line).

The shaded bands represent standard error of the mean. The first vertical dashed line (0 sec) represents scene presentation onset and the second vertical dashed line (3 sec) represents the choice option presentation. The grayed area represents the time period considered for neuronal analysis and was chosen to reflect the corresponding stimulus-response delay and scene presentation times. See Supplementary Fig. 2 for additional examples. (**c**) Peri-stimulus histogram of neural activity ± s.e.m. over time across different scene manipulations for the same cell displayed in Fig. 3b. The five figures represent the five main manipulations from 1–5 (43 trials). These figures do not take into account the participant's specific choices or the point of equipoise. (**d**) Average spiking activity ± s.e.m of two representative cells aligned to equipoise. The corresponding voting profile of the participant is shown in black. For the x-axis, ordinal values below zero represent scenes that the participant's believed were progressively safe and values above zero represent scenes that the participant's believed were progressively unsafe (43 trials). Neural activity of the cell on the *left* is negatively correlated with the participant's voting profile (logistic β regression coefficient of −3.40) and the cell on the *right* is positively correlated (+13.21). (**e**) The β coefficients describing the correlations for all recorded cells. *Left inset*: scatter plot of activity during reported safe versus unsafe trials. *Right inset*: distribution of changes in neuronal activity.
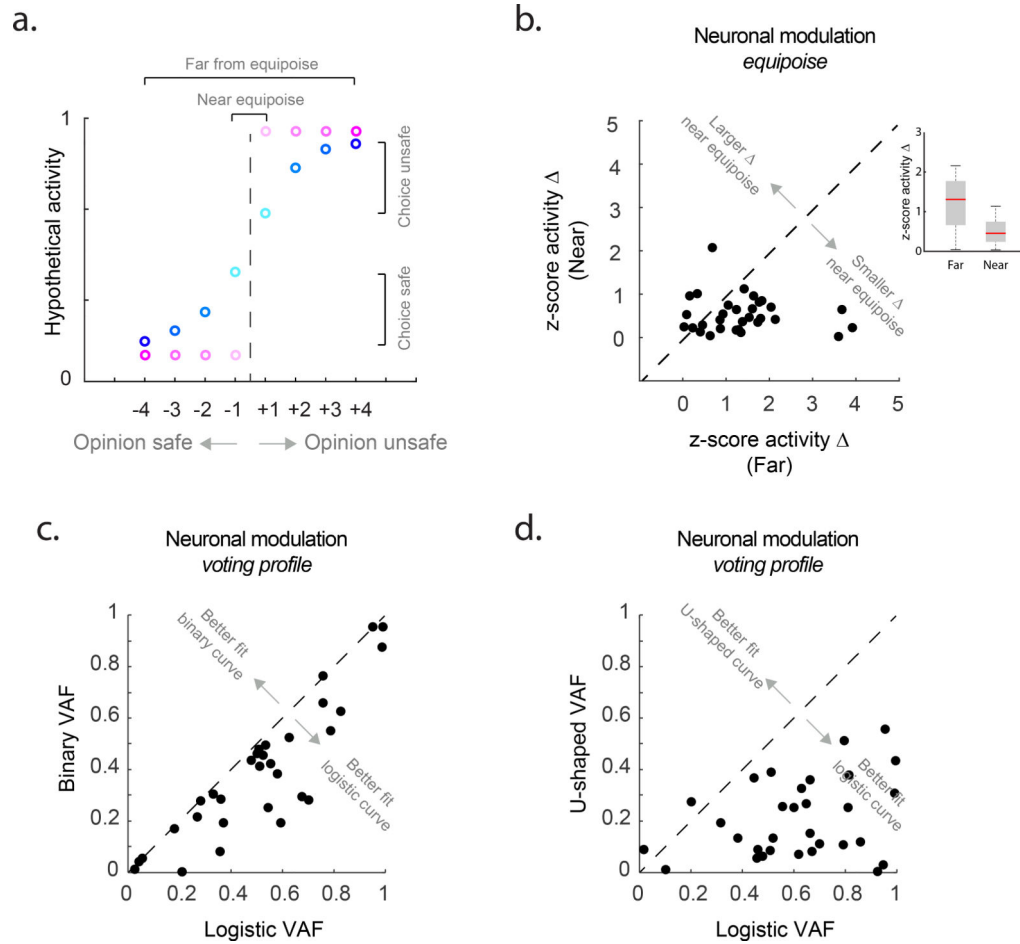
Figure 4. Neuronal modulation in relation to equipoise and voting profile.
(**a**) Three hypothetical neurons illustrating the prospective relations between spiking activity, choice and equipoise. These follow binary (purple), logistic (blue) or u-type (brown) profiles. (**b**) A scatter plot comparing the difference in z-scored activity for safe/unsafe choices near versus far from the equipoise. Circles below the diagonal depict cells in which the choices of safe vs. unsafe were associated with a smaller firing difference when choices were made near equipoise compared to when they were not. The inset shows the median (red line), quartiles (boxes) and the range (whiskers) of the changes in activity for near versus far (n = 31 neurons). **See** Supplementary Fig. 4 for additional details on the firing rate distributions. (**c, d**) Goodness of fit when fitting neuronal activity to a binary (c) or u-shaped (d) vs. logistic function. Differences in the variance accounted for (VAF) indicate that activity of most cells was better explained by a logistic compared to other fits. While a few cells were equally fitted by either a binary or a logistic fit, only two cells displayed a better u-shaped fit than the logistic fit.
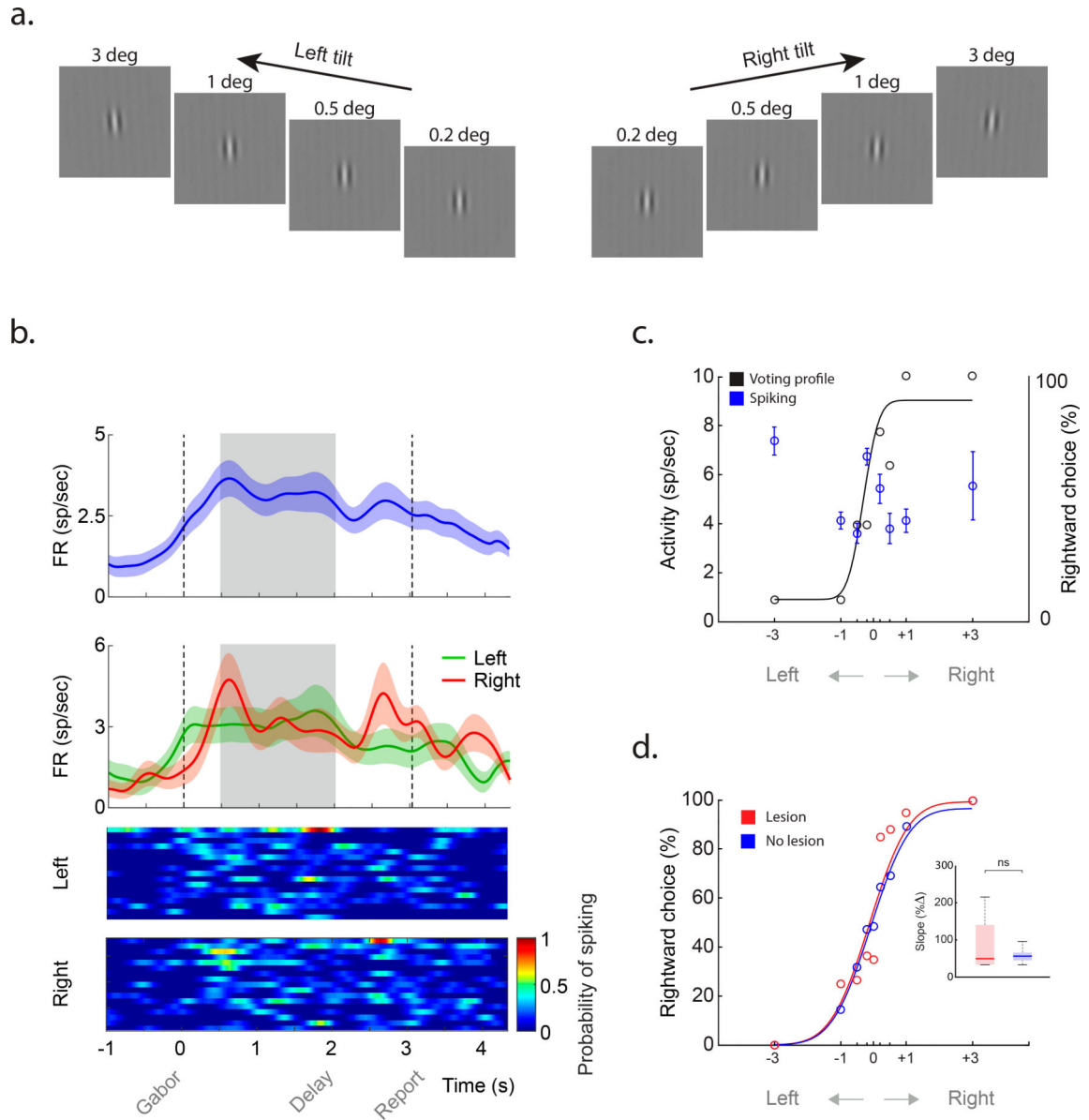
Figure 5. Neuronal responses and effect of dlPFC lesions on sensory perceptual decisions.
(**a**) Gabor stimuli were presented to the participants and varied in orientation from least ambiguous (+3.0 and −3.0 degrees) to most ambiguous (±0.2 degrees). (**b**) Peri-stimulus spike histogram and trial heatmap of an example cell during task performance. The shaded bands represent s.e.m. While the firing activity of this neuron increased during the presentation of Gabors (*top panel*), the neuron was not modulated by variations in the participant's choice of left or right (*middle and bottom panels*). The first vertical dashed line (0 sec) represents Gabor presentation onset and the second vertical dashed line (3 sec) represents the choice option presentation (left/right). The grayed area represents the time period considered for neuronal analysis. (**c**) Average spiking activity ± s.e.m. of the representative cell as a function of Gabor's orientation (29 trials). The corresponding psychometric function of the participant is superimposed (black curve and grey circles). (**d**)

Psychometric profiles with performances for dlPFC lesioned (n = 4) and non-lesioned (n = 14) participants across the different Gabor orientations were similar, displaying no difference in the slope (two-sided Wilcoxon test, p = 0.96). The inset shows the median (red line), quartiles (boxes) and the range (whiskers) of the slopes for lesioned and non-lesioned subjects.
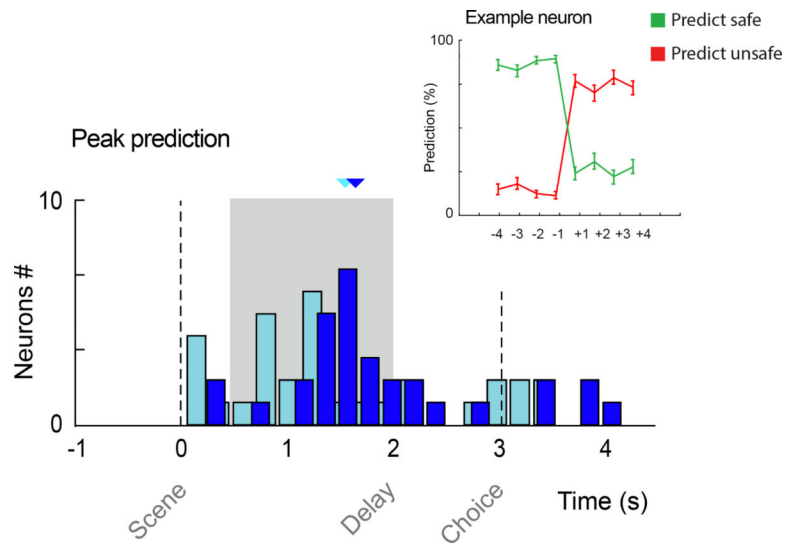
Figure 6. Temporal dynamic of neuronal response.
Peak population prediction timing distribution and means (arrows). The light blue arrow corresponds to trial predictions of choices made near equipoise (ordinal positions −1 and +1) and the dark blue arrow corresponds to those made far from equipoise (ordinal positions −4 and +4). The *inset* illustrates the prediction profile (mean ± s.e.m. for 1000 repetition) obtained from an example cell during image presentation.
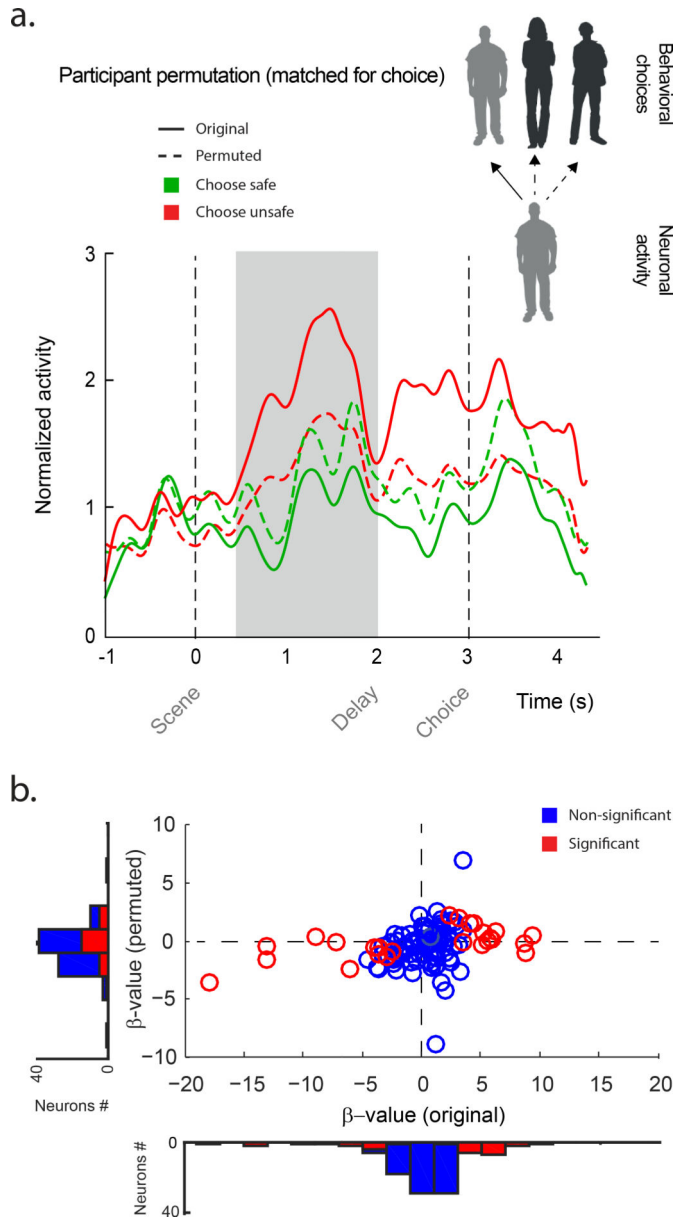
Figure 7. Changes in neuronal activity in the dlPFC correlate with inter-personal variations in opinion.

(**a**) A between-participant permutation procedure was performed in which the same net selections of safe/unsafe were compared, but in which one participant's voting profile was randomly substituted for that of other participants. The solid lines in this peri-stimulus histogram represent the spiking responses of a neuron aligned to the participant's own original voting profile. The dashed lines, in comparison, represent mean spiking responses for the same neuron but aligned to the voting profiles of the other participants (permuted 1000 times). As before, the green and red curves represent neuronal activity during the choice reports of safe and unsafe, respectively. (**b**) Scatter plot depicting the β values describing the correlations between neuronal activity and the participant permuted (y-axis) and non-permuted (x-axis) voting profiles for all cells (n =119). Cells that displayed

significant modulation (two-sided t-test, p<0.05) before the permutation procedure are shown in red (n = 31, permutation).
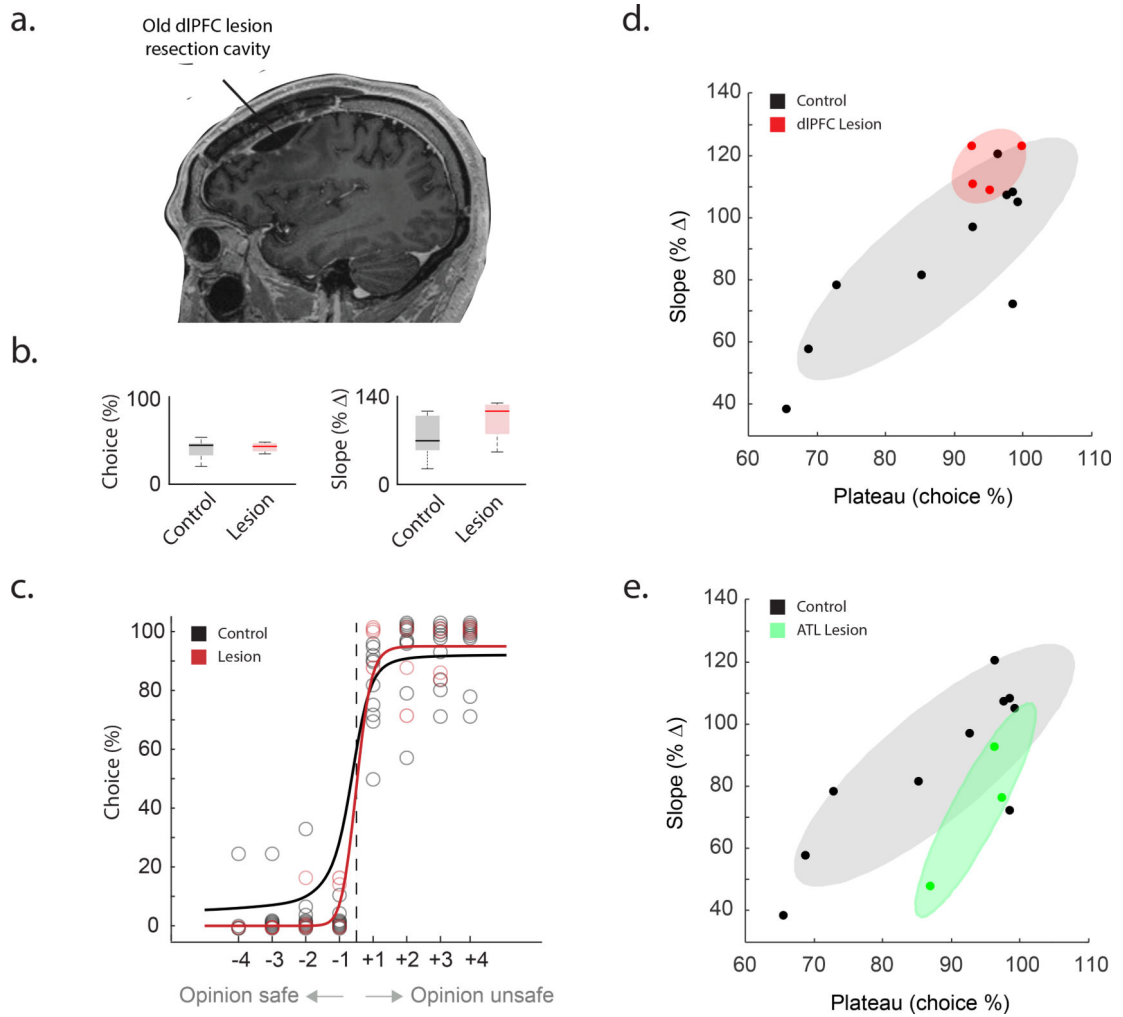
Figure 8. Focal dysfunction of the dlPFC leads to a loss of gradation between opposing decisions but does not influence the participant's motor report.

(**a**) A sagittal *T1* weighted MRI of a participant (total population n = 4) that had previously undergone focal neurosurgical resection of a small lesion localized to the dlPFC. (**b**) On the *left* is the net proportion of trials in which the participants made the choice of safe vs. unsafe ± s.e.m. On the *right* is the slope of their voting profiles (i.e., the rate of change in the participants' choices of safe to unsafe at the inflection of the curve; see Methods). The values shown are the median (red line), quartiles (boxes) and the range (whiskers) for the control (n=11) and lesioned (n=4) subjects. (**c**) Averaged voting profiles of dlPFC lesioned and control participants aligned to equipoise. Supplementary Figure 7 displays the individual performances from which the psychometric curves were constructed separately for clarity. (**d**) Scatter plot depicting the rate of change in choice (slope) vs. average choice at the plateaus of the curves. Individuals with dlPFC lesions (n=4) are displayed in red and those of control (n=11) are displayed in black. The shaded areas represent the centroid and standard-deviation for each respective group. (**e**) Scatter plot for individuals with anterior temporal lobe (ATL) lesions (n=3) and control (n=11). The shaded areas represent the centroid and standard-deviation for each respective group.