Received: 13 April 2021 | Revised: 16 November 2021 | Accepted: 23 November 2021

DOI: 10.1111/psyp.13985

PSYCHOPHYSIOLOGY SPR WILEY

ORIGINAL ARTICLE

Direct feedback and social conformity promote behavioral change via mechanisms indexed by centroparietal positivity: Electrophysiological evidence from a role-swapping ultimatum game

Paul C. Bogdan[1,2] | **Matthew Moore**[1] | **Illia Kuznietsov**[1,3] | **Justin D. Frank**[1] | **Kara D. Federmeier**[1,2,4] | **Sanda Dolcos**[1,2] | **Florin Dolcos**[1,2,4]

[1]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[2]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

[3]Department of Human and Animal Physiology, Lesya Ukrainka Volyn National University, Lutsk, Ukraine

[4]Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Correspondence
Paul C. Bogdan, Sanda Dolcos, and Florin Dolcos, SCoPE Neuroscience Laboratory, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA.
Email: pbogda2@illinois.edu, sdolcos@illinois.edu, and fdolcos@illinois.edu

Funding information
Fulbright Association; Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign; University of Illinois at Urbana-Champaign

Abstract

Our behavior is shaped by multiple factors, including direct feedback (seeing the outcomes of our past actions) and social observation (in part, via a drive to conform to other peoples' behaviors). However, it remains unclear how these two processes are linked in the context of behavioral change. This is important to investigate, as behavioral change is associated with distinct neural correlates that reflect specific aspects of processing, such as information integration and rule updating. To clarify whether these processes characterize both direct learning and conformity, we elicited the two within the same task, using a role-swapping version of the Ultimatum Game—a fairness paradigm where subjects decide how to share a pot of money with other players—while electroencephalography (EEG) data were recorded. Behavioral results showed that subjects decided how to divide the pot based on both direct feedback (seeing whether their past proposals were accepted or rejected) and social observation (copying the splits that others just proposed). Converging EEG evidence revealed that increased centroparietal positivity (P2, P3b, and late positivity) indexed behavioral changes motivated by direct feedback and those motivated by drives to conform. However, exploratory analyses also suggest that these two motivating factors may also be dissociable, and that frontal midline theta oscillations may predict behavioral changes linked to direct feedback but not conformity. Overall, this study provides novel electrophysiological evidence regarding the different forms of behavioral change. These findings are also relevant for understanding the mechanisms of social information processing that underlie successful cooperation.

KEYWORDS
conformity, economic game, ERPs, FMT, LP, P300

https://doi.org/10.1111/psyp.13985

1 | INTRODUCTION

Human behavior is shaped by multiple sources. Most obviously, it is modified by direct learning through feedback (rewards and losses) encountered while interacting with the environment (Niv, 2009). Our behavior is also shaped indirectly by observing the actions of our peers, which oftentimes trigger drives to conform and imitate their behavior (Bandura & Walters, 1977; Shamay-Tsoory et al., 2019; e.g., students naturally picking up the good and bad habits of their colleagues).[1] Conformity and learning through observation are central to adaptive social decision-making, and clarifying their neural mechanisms is a major focus of current research with notable links to clinical conditions, such as autism (Varni et al., 1979) and depression (Thoma et al., 2015). A key question within this area is how the neurocognitive systems associated with processing other people's actions overlap with the machinery dedicated to directly learning from rewards and losses (Bellebaum et al., 2010; Klucharev et al., 2009; Levorsen et al., 2021). In particular, it remains unclear how drives to conform and reward/loss feedback each influence behavioral change.

Behavioral change motivated by direct learning is associated with distinct neural correlates that reflect integrating information, updating rules, and exercising cognitive control (Cavanagh & Frank, 2014; Chase et al., 2011; Eppinger et al., 2017), but it is unclear whether these processes are also common to conformity or instead dissociate the two. The present study fills this gap by using a task that triggers both types of learning to investigate their links to behavioral change, while electroencephalography (EEG) data were recorded. Clarification of these issues is important for understanding the mechanisms behind social information processing.

There are several key challenges in integrating research on direct learning versus learning through observation. The former is generally addressed in settings wherein subjects learn how to make decisions that maximize their rewards (Chase et al., 2011; Donaldson et al., 2016; San Martín et al., 2013). In contrast, conformity studies rarely include a reward dimension (Pierguidi et al., 2019; Wang et al., 2019). Some reward-focused works have notably examined the links between direct learning and learning through observation (Bellebaum et al., 2010, 2012; Peterburs et al., 2021; Rak et al., 2013) but did not specifically examine conformity. These studies typically involved subjects observing the decisions of another person and seeing whether it yielded a reward or loss. Such observations lead to subjects changing their own decision-making. However, it is not clear whether such changes are caused by learning that the action led to some outcome, or are instead the result of a social drive to conform. To better shed light on these different processes, the latter was the sole focus of the present study.

To overcome these challenges, we employed a version of the Ultimatum Game (UG) where participants alternated between the Responder and Proposer roles. In the Proposer role, subjects made offers about how to divide a pool of money, and in the Responder role, subjects evaluated the other players' offers. Throughout the experiment, subjects continuously changed their Proposer behavior to maximize their earnings and achieve a sense of fairness in their interactions. Specifically, their Proposer choices were informed by both direct learning in previous Proposer trials (whether subjects' past offers were accepted or rejected) and conformity based on previous Responder trials (seeing what offers that others proposed). EEG data were recorded to investigate the electrophysiological correlates of behavioral change and to test for overlapping or contrasting patterns associated with each form of behavioral learning. These statistical analyses were made possible by the task manipulations of the present study, which, to our knowledge, is the first that invoked both direct learning and conformity within the same task while circumventing the aforementioned challenges. Below, we elaborate on the advantages of our methodology and provide relevant conceptual details.

1.1 | Event-related potential research on direct learning

The processes underlying direct learning can be captured, at least in part, by examining how past feedback influences a person's future behavior. Pursuing this question experimentally often involves using probabilistic gambling tasks, where subjects make decisions and each choice has some likelihood of yielding a reward or loss. These outcomes serve as feedback and inform subsequent choices. Event-related potential (ERP) studies using gambling tasks have shown that increased centroparietal P2, P3b, and late positivity (LP) elicited by outcome feedback predict that subjects will change their

[1] Conformity is notably distinct from "observational learning"—for example, learning how to start a fire from observing another person do so. While some early literature used "conformity" and "social/observational learning" interchangeably (Bandura & Walters, 1977), the present report uses "conformity" exclusively to refer to behavioral learning based on the social drive to imitate others' behaviors (Bandura & Walters, 1977; Shamay-Tsoory et al., 2019), and uses "observational learning" exclusively to refer to learning that a given action will lead to some outcome based on observing this occur for another person (Bellebaum et al., 2010).

behavior in the next trial (Chase et al., 2011; Donaldson et al., 2016; San Martín et al., 2013). In this context, P2 amplitude is most closely linked to stimulus valence, P3b is more often tied to unsigned prediction error, and LP potentially plays a role in integrating valence and expectation violation (Donaldson et al., 2016; Stewardson & Sambrook, 2020). Each component warrants investigation, and together they are thought to reflect the integration of information and updating of mental models (i.e., model-based learning; Eppinger et al., 2017). Notably, these earlier studies employed tasks where trials included objectively correct/most rewarding options, which is different from our UG design, in which there is no objectively "correct" way to act. Nonetheless, all these designs share a fundamental property—they prompt subjects to continuously change their behavior based on new information. For our design, the lack of objectively correct choices allowed us to also test the effects of conformity, which we will elaborate upon below.

Feedback-related negativity (FRN) and Frontal Midline Theta (FMT; 4–8 Hz) are closely related and have also played key roles in decision-making research. FRN is sensitive to the outcomes of decisions and tracks average gains and losses associated with choices. However, in contrast to centroparietal positivity, FRN often fails to predict changes in subjects' behaviors (Chase et al., 2011; San Martín et al., 2013). On the other hand, increases in FMT have been associated with the recruitment of cognitive control (Cavanagh & Frank, 2014) and the ability to predict subsequent behavioral change in paradigms such as Go-NoGo, and in tasks involving punishment following errors (Cavanagh & Shackman, 2015). FMT has received less attention in tasks wherein subjects make decisions with probabilistic outcomes, but some results suggest that FMT also predicts behavioral change in this area (Mas-Herrero & Marco-Pallarés, 2014). Hence, while overall less is known about the association between FMT effects and behavioral change, this is also a suitable target for the present study. Altogether, this earlier research highlights that learning is composed of multiple processes and that a focus on behavioral change offers a unique perspective on its underlying mechanisms.

1.2 | ERP research on conformity and observational learning

The mechanisms underlying behavioral change through social observation have been the target of much research (reviewed in Olsson et al., 2020; Shamay-Tsoory et al., 2019). Our specific focus here is on conformity (the social drive to copy or mimic another person), which is typically studied by examining the neural correlates that predict subjects will change their behavior to match what they just observed. For example, in one commonly used task, subjects are first asked to make an opinion-based rating about some topic (e.g., how beautiful a face is). Subjects are then shown other people's opinions on the topic, and then later are asked whether they would like to change their initial rating. Typically, subjects tend to change their rating to mirror the group, and increased centroparietal positivity in response to seeing the group's opinion predicts subsequent conformity (Pierguidi et al., 2019; Wang et al., 2019). Paralleling research on direct learning, FRN amplitude increases when seeing that another's opinion is different from one's own but does not predict changes in opinion (Wang et al., 2019). Aside from these similarities with results on direct learning, some differences also emerge. Specifically, in the context of conformity, increased centroparietal positivity has only been identified for the P3b and LP time windows, unlike the results on direct learning which also included P2 effects. Moreover, conformity is also linked to other effects, such as increased centroparietal N2 amplitude (Pierguidi et al., 2019), which creates a further disconnect with direct learning research.

These differences may be linked to deviations in the task design, as the direct learning studies focused on subjects maximizing their rewards. This is not the case for the conformity tasks. The conscious pursuit of goals, such as rewards and fairness, elicits unique neural processes, like early centroparietal positivity (Heydari & Holroyd, 2016). Hence, these branches of research on direct learning and conformity cannot be reliably compared or contrasted from the existing results. Integrating them and clarifying the mechanisms underlying behavioral change can be best done using a singular task in which both direct learning and conformity are tied to goal-related processes such as rewards and fairness. Related works have demonstrated that economic games are robustly suited for this purpose and can effectively evoke both forms of behavioral learning. For example, individuals who are punished by others in these games adapt to meet other players' expectations, which demonstrates direct behavioral learning (Herrmann et al., 2008). Furthermore, when subjects are presented examples of others' economic behavior, they tend to copy what they observed (Chierchia et al., 2020; Sacconi & Faillo, 2010), even when it comes at their own expense (Bicchieri & Xiao, 2009; FeldmanHall et al., 2018). This tendency to mirror the economic behavior of others is thought to reflect conformity and is sensitive to manipulations known to upregulate conformity (e.g., placing subjects face-to-face with other players; Guazzini et al., 2019). Altogether, this earlier research highlights the effectiveness of economic games for broadly integrating different aspects of behavioral learning, but no previous study has done this while recording neural data.

As mentioned above, our focus is on mechanisms of feedback-based/direct learning and conformity, which have not been investigated together in the context of behavioral change. However, there is a growing body of research dedicated to comparing direct learning and observational learning (i.e., learning through observation but *not* via a drive to conform). For example, studies using probabilistic gambling tasks have probed both by having subjects learn a task's structure through making decisions or via exposure to the outcomes of others' decisions (Bellebaum & Colosio, 2014; Bellebaum et al., 2010, 2012). Within these designs, direct and observational learning led to similar levels of accuracy during subsequent trials where learning was tested (Bellebaum & Colosio, 2014; Bellebaum et al., 2010, 2012). Adding to these similarities, both forms of learning are susceptible to some of the same biases (Peterburs et al., 2021). However, this is not always the case (Nicolle et al., 2011), and differences linked to the effect of personality on learning have also been found (Rak et al., 2013). Further similarities and differences have been identified at the EEG level. While P3, FRN, and error-related negativity are involved in both direct and observational learning, their amplitudes and potentially their roles have been shown to differ between these two forms of learning. FRN is relatively dampened during observation (Bellebaum et al., 2010; Huberth et al., 2019; Koban et al., 2012) and shows weaker learning-related modulation (Bellebaum & Colosio, 2014). Centroparietal positivity is also decreased or, in some cases, even absent during observational learning (Huberth et al., 2019; Rak et al., 2013). Functional MRI has uncovered further key overlaps and dissociations. The ventromedial prefrontal cortex is similarly sensitive to both personal rewards and observing others receive rewards. However, the nucleus accumbens shows greater activation for personal rewards, whereas the dorsomedial prefrontal cortex and posterior superior temporal sulcus are more sensitive to others' rewards (Dunne et al., 2016; Morelli et al., 2015). Altogether, these results highlight similarities and differences between direct and observational learning. However, the above studies focused solely on subjects observing other's actions then seeing the outcomes, but social observation also elicits drives to conform regardless of outcomes (Wang et al., 2019).

To achieve a clearer picture of social information processing and decision-making, conformity must also be accounted for, along with its links to feedback-based learning. This topic has been investigated to some degree, but not in the context of behavioral change. In large part, conformity research has targeted error-processing (Wu et al., 2016). For example, typical studies have involved subjects performing a gambling task, where they attempt to maximize their rewards, and a separate social conformity task, using opinion-based ratings (Klucharev et al., 2009; Levorsen et al., 2021). This work has revealed that both direct learning and conformity elicit error signals that are processed by overlapping neural regions (Klucharev et al., 2009), albeit through different neural representations (Levorsen et al., 2021). However, it remains unclear how these results should be interpreted. In the case of direct learning, error signals likely reflect subjects updating their perceived values, but, in the case of the conformity-related information, errors may represent subjects updating their own perceived values or updating their beliefs about others. If the latter is true, comparing feedback processing and drives to conform in terms of error-related signals may not be appropriate. To address these issues, the present research uses a single task in which direct learning and conformity both influence decision-making with regards to a common goal. Additionally, rather than focusing on error-processing, we investigated the neural signatures of behavioral change.

1.3 | The present study

Unlike previous research, the present study employed a novel version of the UG task to motivate behavioral change based on both direct feedback and conformity. UG is a two-player economic game wherein one player (the Proposer) decides how to split a pot of money ($10 in this case) with another player (e.g., they may decide to take $6 and give the other player $4). After receiving the offer, the Responder decides whether to accept or reject it. Acceptance causes the money to be distributed as proposed, whereas rejection causes neither player to receive any money from the pot. Although it may be expected that participants would try to maximize their earnings by accepting every offer, this is typically not the case—instead, participants tend to reject very unequal splits (e.g., $2:$8 or $1:$9; Oosterbeek et al., 2004). For our study, subjects alternated between the Proposer and Responder roles. This allowed their Proposer strategies to be influenced by both direct (accept vs. reject) feedback from Proposer trials and drives to conform from Responder trials.

Based on the available evidence, we expected to identify several effects related to behavioral change based on direct learning and conformity. Regarding the behavioral effects, we expected that subjects would change their Proposer behavior based on direct feedback, shifting towards more generous, safe offers after their offers are rejected, and toward more selfish, risky offers after their offers are accepted. For conformity, we expected that receiving relatively generous offers would encourage subjects to propose more generously in the next trial, and receiving relatively selfish offers would have the opposite effect. In the ERPs, we expected that increased centroparietal positivity when processing feedback would predict that subjects subsequently change their Proposer behavior. Similarly, we expected that the centroparietal positivity elicited by received offers would also predict a subsequent change in Proposer behavior via conformity. For both ERP hypotheses, we expected that centroparietal positivity would predict behavioral changes at the trial level and would also differentiate among subjects who frequently change their behavior versus rarely change their behavior. Finally, we also explored event-related spectral perturbations and tested whether FMT effects are associated with direct behavioral learning and/or conformity (Supporting Information: Method 1.5 and Results 2.5).

2 | METHOD

2.1 | Participants

A total of 40 subjects from the University of Illinois and the surrounding Urbana-Champaign community participated in this study (18 to 39 years old, 50% female, $M_{\text{Age}} = 23.1$, $SD_{\text{Age}} = 5.3$). Concerning the behavioral aims, this sample size is supported by power analyses ($\alpha = .05$, $\beta = .80$, two-tailed) conducted on an independent sample of pilot subjects who completed the same or a similar version of the UG task ($N = 15$; direct feedback effect: $d = 1.82$; conformity: $d = 0.59$), which revealed that 25 subjects would be sufficient to identify significant behavioral effects linked to both direct feedback and conformity. For the EEG aims, power analyses were performed based on the effect sizes of prior related studies. Namely, for the within-subject ERP hypothesis concerning the link between behavioral change and direct feedback, P3 data from Chase et al. (2011; $t[12] = 4.15$, thus $d = 1.15$) suggested that 9 subjects are needed. For the within-subject hypotheses on conformity, P3 data from Wang et al. (2019; $F[1, 24] = 7.25$, thus $d = 0.54$) suggested that 29 subjects are needed. Finally, for the correlations, P3 data from Donaldson et al. (2016; $r = .39$–$.48$) suggested that between 29 and 46 subjects are needed. Our sample size is notably also comparable with other similar studies (San Martín et al., 2013) and is consistent with simulation research (Boudewyn et al., 2018), which has demonstrated that the centroparietal patterns are typically stable and detectable using smaller sample sizes than what is needed for other components, such as error-related negativity, which are not targeted here. All participants were healthy, right-handed, native English speakers, and reported no recent history of psychiatric or neurological conditions.

One participant was excluded due to falling asleep during the experiment, and two participants were excluded due to malfunctions during EEG recording, which resulted in a final set of data from 37 participants. Of these participants, some very rarely changed their Proposer behavior across the experiment, and hence could not be used for comparing trials preceding behavioral Change versus No-Change. Specifically, due to low trial numbers in three/four participants, respectively, 34 subjects could be used for the direct feedback comparison (Change vs. No-Change conditions), and 33 could be used for conformity comparison (Conformity vs. No-Conformity). These rates of attrition are reasonable, as subjects utilize different strategies in social-economic games. However, all 37 subjects could be used for the correlation analyses that link average ERP responses and overall tendencies to change behavior frequently versus rarely. All participants provided written informed consent under a protocol approved for use of deception by the Institutional Review Board and received payment for their participation. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

2.2 | The UG procedure

Participants played a role-swapping UG task for 384 trials, alternating between the Proposer and Responder roles and inputting choices via a keyboard (Figure 1). Participants were told that their UG performance would influence their monetary payment, but, in reality, all participants received equal payment for their participation. Consistent with earlier UG studies (Chang & Sanfey, 2013; Xiang et al., 2013), participants were told (1) that they would interact with a large group of other players, (2) that they would play with different people between trials, and (3) that they should not expect their behavior in one trial to impact the other player's behavior in the following trial. However, subjects were not told that they were playing with a group of 384 players, as this would not have been realistic. This design is suitable to elicit learning, as prior research has demonstrated that even when subjects are told that they are continuously presented with new partners, they still utilize information from past trials to inform future decisions (Xiang et al., 2013).

A role-assignment screen at the start of each trial informed participants of their current roles. For the Proposer role trials, the next screen prompted participants to decide how to split the $10 pot by choosing from five pairs representing the amounts for the Proposer/

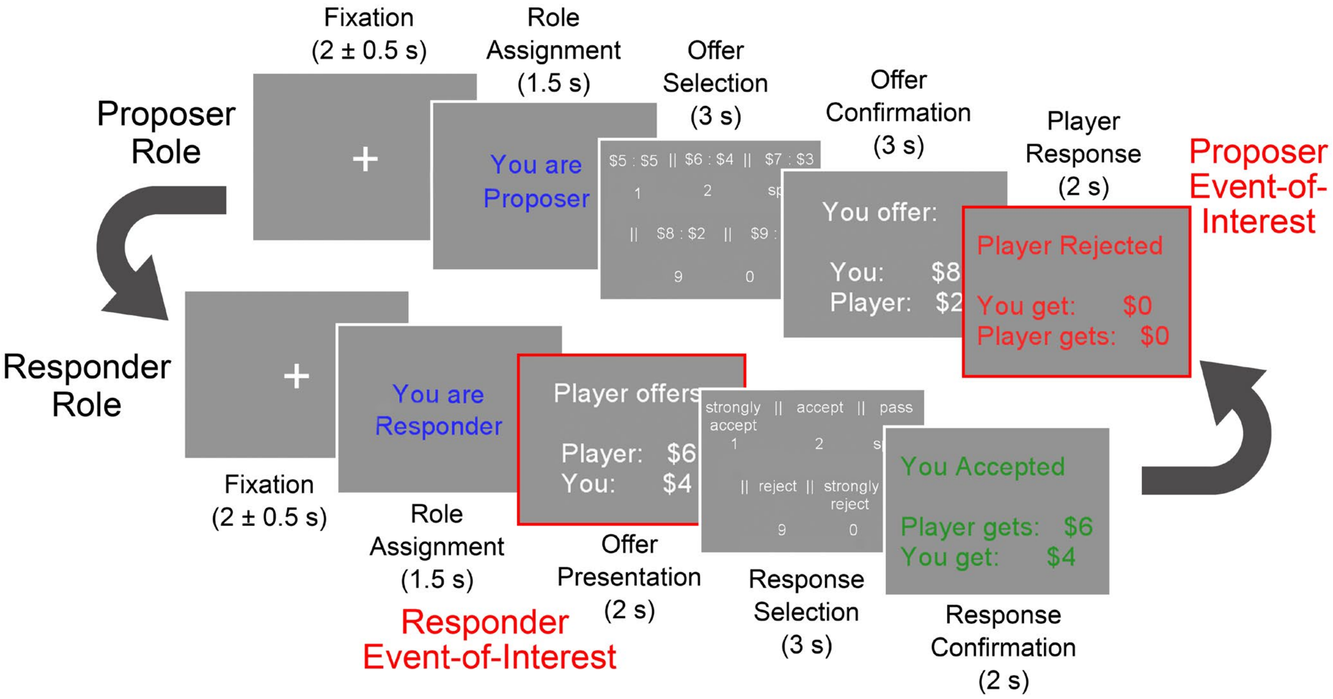

FIGURE 1 Task diagram showing the Ultimatum Game procedure. Subjects switched between playing as Proposers and Responders. The associated trials had similar structures, to ensure that participants believed they were playing with another player in the opposite role. Proposer trial analyses were time-locked to the "Player Response" screen, and Responder trial analyses were time-locked to the "Offer Presentation" screen

Responder, respectively ($5:$5, $6:$4, $7:$3, $8:$2, or $9:$1). Following a delay, subjects were told whether their partner accepted/rejected their offer along with the trial's earnings. For the Responder role trials, the screen after role assignment informed participants of the offer amount that they received for that trial. Subjects had to evaluate this offer and, in the subsequent screen, select from five possible responses (Strongly Reject, Reject, Pass, Accept, or Strongly Accept). In terms of the UG trial outcomes, the Strongly Reject and Reject response options were equivalent to one another, and so were the Strongly Accept and Accept options. Subjects were told to select from among these options based on the degrees of their responses, ranging from "somewhat" sure to "extremely" sure. The use of five options notably differs from most prior UG studies that provide only binary response options: *Accept* and *Reject*. Our protocol used five to ensure that subjects had the same number of options for both the Proposer and Responder roles and to dissuade subjects from moving their hands across trials to make responses. Dissociating between Strongly versus non-Strongly response options is not the focus of the present report, and hence our analyses collapsed Strongly Accept and Accept trials into an *Accept* category, and Strongly Reject and Reject trials into a *Reject* category. Subjects were instructed to only select Pass if they were totally unable to decide. As expected, the Pass response option was rarely chosen (under 2% of trials), and thus did not influence the results.

The 192 Proposer trials and 192 Responder trials were divided into 8 blocks of 48 trials each, separated by short breaks to avoid fatigue. Our focus was on both the effect of a Proposer trial on the subsequent Proposer trial (feedback processing), and the effect of Responder trials on the next Proposer trial (conformity, which is relatively subtle). Hence, in 87.5% of trials, subjects alternated between roles, playing as Proposer after a Responder trial and vice versa. In the other 12.5% of trials, which were interspersed throughout the blocks, subjects played the same role twice in a row. Prior to the experiment, instruction and practice rounds ensured that subjects were familiar with the keyboard controls. Subjects were told that they were playing with other humans, but in reality, they always played with a computer, as is typically done in psychological research using the UG (Chang & Sanfey, 2013; Xiang et al., 2013). When the subject played as a Proposer, the computer was more likely to reject lopsided offers. Specifically, the computer rejected offers of $5:$5 in 1% of trials, offers of $4:$6 in 12% of trials, offers of $3:$7 in 45% of trials, offers of $2:$8 in 67% of trials, and offers of $1:$9 in 75% of trials. The computers' accept and reject responses were not influenced by past trials and did not follow any pseudo-random sequence. These rejection rates emulated human behavior identified in a pilot experiment and reflect patterns of behavior reported in the literature (Oosterbeek et al., 2004). When the subject played as a Responder, they received offers of \$5:\$5 in 32% of trials, offers of \$4:\$6 in 18% of trials, offers of \$3:\$7 in 17% of trials, offers of \$2:\$8 in 17% of trials, and offers of \$1:\$9 in 17% of trials. This distribution was pseudo-randomly determined prior to the experiment so that in each block participants received similar rates of equal, moderately unequal, and highly unequal offer amounts. The average of these received offers (\$3.3:\$6.7) was consistent with typical human behavior (Oosterbeek et al., 2004).

2.3 | Analytic strategy

2.3.1 | EEG preprocessing

EEG data were continuously recorded for each run/block, at a sampling rate of 2,048 Hz, with a 64-channel electrode cap and three EOG electrodes, using a BioSemi ActiveTwo System and the ActiView software (BioSemi BV, Amsterdam, the Netherlands). EOG channels were located at the outer canthi of the left and right eyes and below the right eye. Data were processed using the MNE Python package (Gramfort et al., 2013, 2014). Data were first referenced to Fz (subsequently average referenced before data analyses—see below), down-sampled to 256 Hz, subjected to low-pass finite impulse response filtering at 30 Hz and high-pass finite impulse response filtering at 0.1 Hz. Second, artifact rejection and correction were performed using the MNE implementation of Autoreject (Jas et al., 2017), a fully automated algorithm that identifies outlier electrodes on a trial-by-trial basis by measuring the peak-to-peak differences, interpolates electrodes whose differences surpass a peak-to-peak amplitude threshold, and rejects trials that have an excessive number of interpolated electrodes. AutoReject defines a unique peak-to-peak for each subject, using a cross-validation-based method (median threshold = 150 μV, range = 92–340 μV). This procedure excluded 15.4% of trials (see below for the final trial counts associated with each condition), and yielded a dataset with an average of 4.2 interpolated electrodes per included trial. Visual inspection of the trials' data confirmed that the AutoReject consistently identified trials containing artifacts (e.g., muscle movements and blinks; see below for details on the number of trials remaining following trial rejection). Third, the data were re-referenced to an average reference, and the reference used during importing (Fz) was added back to the data, to be able to better compare results with prior work in social cognition that have also used an average reference (Bailey & Kelly, 2017; Kröger et al., 2013; Schmitz et al., 2012). For the analyses below, independent component analysis (ICA) cleaning was not performed, given the stability and posterior location of the targeted centroparietal effects. However, to ensure the robustness of our results, confirmatory analyses were performed using a data set that was cleaned via ICA (Supporting Information: Method 1.2), and all results were replicated (Supporting Information: Results 2.2).

The ERP analysis focused on the P2, P3b, and LP components. Amplitudes were measured using the medial centroparietal electrodes (C1, Cz, C2, CP1, CPz, CP2), consistent with earlier research on direct learning and conformity (Chase et al., 2011; Donaldson et al., 2016; Wang et al., 2019). P2 was defined as the mean amplitude of the 200–300 ms window, P3b as the mean of the 300–500 ms window, and LP as the mean of the 500–800 ms window. Mean amplitudes were used as this achieves a more stable measure than peak-based approaches (Clayson et al., 2013; Keil et al., 2014). Additionally, as we describe below, our different conditions contained different numbers of trials, which can bias peak measurements but not mean amplitude measurements (Clayson et al., 2013; Keil et al., 2014). These windows are consistent with the timings of the components shown by the waveforms. Related literature contains disagreements with regard to timing cutoffs (Chase et al., 2011; Donaldson et al., 2016; San Martín et al., 2013; Wang et al., 2019), but our definitions broadly align with these prior studies. Each of these amplitude measures was submitted to the same series of analyses described below. Time-frequency decomposition patterns were also analyzed to clarify the role of FMT in learning (Supporting Information: Method 1.5).

2.3.2 | Behavioral analysis

Concerning the direct-learning focus, we first tested whether acceptances prompt subjects to propose more selfishly in the next trial and whether rejections prompt subjects to propose more generously in the next trial. We calculated subjects' average changes in behavior following each form of feedback and submitted these values to one-sample *t*-tests. We next compared the magnitudes of each change effect by converting both types of change as positive values (i.e., multiplying changes towards selfishness by −1). These values were submitted to paired *t*-tests evaluating whether subjects changed their behavior more due to rejection or acceptance. Concerning the conformity focus, we compared the effects of receiving a "high" offer (\$4 or more) versus receiving a "low" offer (\$3 or less) on changes in subjects' Proposer behavior. Dividing the offers into groups of \$5 and \$4 ("high") versus groups of \$3, \$2, and \$1 ("low") results in the data being split at the median, as subjects received \$5 in 33% of trials and received each other offer in 16.7% of trials (similar to Wei et al., 2013). Splitting at the median maximizes statistical power relative to other splits. Psychologically, offers of $5 and $4 are also similar in that they are both overwhelmingly accepted (Vavra et al., 2018), and both reflect what subjects themselves tend to propose most frequently (Oosterbeek et al., 2004).

2.3.3 | EEG analysis

As our focus was on the neural correlates of feedback processing that predict changes in subjects' Proposer behavior, our analysis began by identifying trials preceding such a change. Specifically, our first within-subject analysis concerned the ERP correlates of direct behavioral learning. The analysis targeted the "Player Response" screen (Figure 1), where the subject is informed of whether their proposed offer was accepted or rejected. To measure the effects of this feedback, we identified the Proposer trials which preceded subjects changing their behavior accordingly—that is, trials where subjects proposed more generously after rejection or more selfishly after acceptance. For example, a Proposer trial[n] where the subject proposed $6 and it was accepted would be considered a "Change" trial if the subject subsequently proposed $7, $8, or $9 in trial[$n+1$] or trial[$n+2$]. On the other hand, it would be considered a "No-Change" trial if the subject had subsequently proposed $5 or $6. Note that changes in the opposite direction (e.g., proposing more generously after acceptances) count as "No-Change," as they do not reflect behavioral learning. This approach utilized all proposer trials except for the last one of each block. All of the included subjects ($N = 34$) had high numbers of trials in each of these conditions. The "Change" condition was associated with an average of 52 artifact-free trials per subject, and the "No-Change" condition with 102 clean trials. Multilevel logistic regression was performed to assess whether ERP amplitude predicts subsequent Change versus No-Change. To ensure that these results were not due to other variables which may influence centroparietal positivity and behavioral change (e.g., the offer amount or the response), the logistic regression models controlled for these factors:

$$\begin{gathered}\text{Change}\,[n+1]\ \sim 1+\ \text{ERP}\,[n]\ +\text{proposed}\,[n]\\ +\ \text{response}\,[n]\ +\ \text{proposed}\,[n]*\text{response}\,[n]\\ +(1+\ \text{ERP}\,[n]\ +\text{proposed}\,[n]\\ +\ \text{response}\,[n]\ +\ \text{proposed}\,[n]*\text{response}\,[n]\ |\,\text{Subject})\end{gathered}$$

Next, to measure whether the effect of ERP amplitude on change was specific to either acceptance or rejection feedback, follow-up analyses were conducted that included an ERP × Response interaction. A positive interaction would reveal that the effect of ERP amplitude on behavioral change is strongest following acceptances, while a negative interaction would reveal that the effect of ERP amplitude is strongest following rejections. This was done using the following logistic regression model (the random effect covariance structure was slightly simplified to avoid convergence errors):

$$\begin{gathered}\text{Change}\,[n+1]\ \sim 1+\ \text{ERP}\,[n]\ +\text{response}\,[n]\ +\text{ERP}\,[n]*\text{response}\,[n]\\ +\text{proposed}\,[n]\ +\ \text{proposed}\,[n]*\text{response}\,[n]\\ +(1+\ \text{ERP}\,[n]\ +\ \text{response}\,[n]\ +\ \text{ERP}\,[n]*\text{response}\,[n]\ |\,\text{Subject})\\ +(1+\ \text{proposed}\,[n]\ +\ \text{response}\,[n]\ +\ \text{proposed}\,[n]*\text{response}\,[n]\ |\,\text{Subject})\end{gathered}$$

Paralleling these within-subject tests, we performed analogous across-subject correlations. Specifically, we followed a similar procedure as Donaldson et al. (2016) and measured the correlation between subjects' average likelihood of changing their proposer behavior and their average ERP responses across both Change and No-Change conditions. These analyses could be conducted on all 37 subjects. As with the within-subject analyses, a follow-up test was done analyzing only rejected trials, which was limited to the set of 34 subjects, as those three who rarely changed their behavior also rarely had their offers rejected. Finally, again paralleling the within-subject analysis, a follow-up multiple regression was performed measuring the link between ERP amplitudes and behavioral changes, using only accepted trials or only rejected trials.

Our second analysis concerned the ERP correlates of conformity. We specifically examined Responder trials, and the "Offer Presentation" screen (Figure 1), where subjects are informed of what offer they received. We organized the trials into "Conformity" versus "No-Conformity" conditions based on the offer amount subjects received and their subsequent change in behavior. A "Conformity" Responder trial was defined as one which preceded a conforming change in Proposer behavior. That is, if the subject received a low offer ($3 or less) in the current Responder trial, and in the next trial changed their Proposer behavior to be more selfish, this is defined as a "Conformity" trial. Alternatively, if the subject received a low offer and their subsequent Proposer behavior remained unchanged or became more generous, this would be defined as a "No-Conformity" trial. For example, if the subject proposed $6 in Proposer trial[$n$−1] and received an offer of $2 in Responder trial[$n$], then trial[$n$] would be defined as a "Conformity" trial if subjects proposed $7, $8, or $9 upon returning to the Proposer role. On the other hand, trial[n] would be defined as a "No-Conformity" trial if the subject proposed $5 or $6 when they returned to the Proposer role. An analogous definition was used for instances where subjects received high offers ($4 or more). These trials were defined as "Conformity" trials if they prompted subjects to be more generous, and No-Conformity trials if subjects did not change or subsequently acted more selfishly. All of the included subjects ($N = 33$) had high numbers of trials in each of these conditions. The No-Conformity condition was associated with an average of 117 trials, and the Conformity condition with an average of 47 trials. Similar to the analysis of direct behavioral learning, to assess whether ERP amplitude predicts conformity, multilevel logistic regressions were used:

$$\text{Conformity}\,[n] \sim 1 + \text{ERP}\,[n] \\ + (1 + \text{ERP}\,[n] \mid \text{Subject})$$

Controlling for other variables (e.g., the offer amount subjects received) was not necessary, as this covariate did not influence ERP amplitudes. Also paralleling the analysis of direct behavioral learning, we performed similar across-subject correlations that linked the likelihood of conforming to a given offer with subjects' average ERP amplitudes across the Conformity and No-Conformity conditions.

Finally, to confirm that the centroparietal positivity associated with the across-subject correlations is indeed a signature of learning, and not of other phenomena such as task engagement, we performed an analysis measuring the link between change frequency and centroparietal amplitude following the "Offer Selection" screen (i.e., a screen stimulus which would not be relevant to learning; Figure 1). This was expected to yield null results, which would provide evidence against the idea that significant correlations are linked to individual differences in subjects' levels of task engagement.

2.3.4 | Alternative EEG analyses

To confirm the robustness of the present findings, additional analyses were carried out using alternative statistical procedures, which are reported in the Supporting Information. These analyses, in large part, aim to increase consistency with earlier research on the neural correlates of behavioral change and conformity (Chase et al., 2011; San Martín et al., 2013; Wang et al., 2019). These include: (a) analyses which rely on t-tests and linear regression rather than logistic regression, (b) analyses using data cleaned via ICA, (c) conformity analyses that treat offers as continuums rather than collapsing them into high ($5, $4) versus low ($3, $2, $1) groups, (d) conformity analyses that use alternative grouping definitions (e.g., analyses which defined high and low relative to subjects' own previously proposed offers), and (e) conformity analyses that control for direct feedback effects. In every instance, these additional analyses replicated the findings reported in the main text.

2.4 | Software and additional multilevel model details

All statistics were performed using R and R Studio (R Core Team, 2013). The multilevel regressions were fit using the *lme4* package (Bates et al., 2014). Models were fit using restricted maximum likelihood (REML) and a full variance-covariance structure for all of the random effects (other than where noted above). Convergence errors did not occur. Fixed effect significance was calculated using the *lmerTest* package, which relies on Satterthwaite's degrees of freedom method (Kuznetsova et al., 2017). The use of REML and Satterthwaite approximation minimizes the type I error rate (Luke, 2017) and adheres to contemporary best practices (Meteyard & Davies, 2020). The multilevel regressions reported in the main text were all maximal models. That is, they included random slopes for each predictor, as this minimizes the type I error rate (Barr et al., 2013; Schielzeth & Forstmeier, 2008) and also adheres to contemporary best practices (Meteyard & Davies, 2020).

3 | RESULTS

3.1 | Behavioral results

The behavioral findings showed the expected direct learning and conformity patterns. In general, subjects showed a high degree of flexibility in their Proposer behavior, and no single offer amount was selected in more than half of the trials (Figure 2a). Subjects tended to propose more generously after their previously proposed offers were rejected ($t[33] = 10.4$, $p < .001$) and more selfishly after their previously proposed offers were accepted ($t[33] = -7.5$, $p < .001$; Figure 2b). Direct comparison of these effects revealed that the effect of rejection was significantly larger than that of acceptance ($t[33] = 8.0$, $p < .001$). We also found that subjects tended to shift their behavior to conform with the offer they just received. They proposed more generously after receiving a high offer ($t[32] = 2.70$, $p = .011$) and more selfishly after receiving a low offer ($t[32] = 2.88$, $p = .007$; Figure 2c).

3.2 | ERP results

3.2.1 | Behavioral change due to direct feedback

Confirming our first ERP hypothesis, we found that increased centroparietal positivity when processing the accept versus reject feedback predicted behavioral

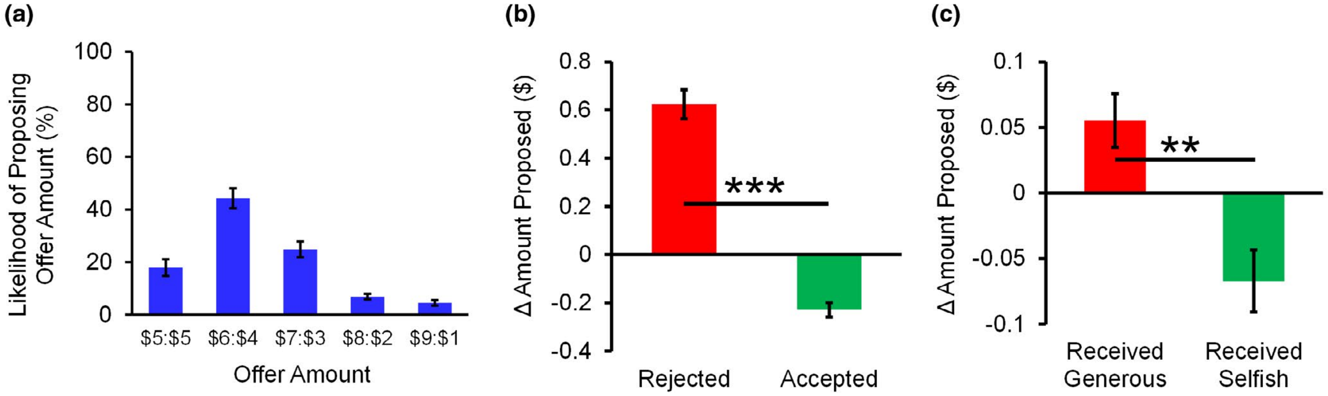

FIGURE 2 Behavioral (a) change due to both (b) direct feedback and (c) conformity. (a) Histogram shows the likelihood that subjects proposed any given amount. (b) After subjects' proposed offers are rejected, they tend to shift toward proposing more generous offers, which are less likely to be rejected. After their offers are accepted, they tend to shift toward proposing more selfish offers, which have the potential for larger payouts but carry an increased risk of rejection. These shifts are changes in response to feedback and are instances of direct behavioral learning. (c) After subjects received offers that were more generous than the median offer ($5 or $4; "high" offers), they tended to shift toward proposing more generously. After subjects received offers that were more selfish than the median ($3, $2, or $1; "low" offers) they tended to shift toward proposing more selfishly. In both cases, subjects mirror the offer they had just received, and hence this is a form of conformity. Error bars represent one standard error above and below the mean. $^{**}p < .01$; $^{***}p < .001$

change. Specifically, multilevel logistic regressions that controlled for other influencing variables—for example, the offer amount that subjects proposed and the other player's response—revealed that P3b and LP but not P2 amplitude predicted behavioral change. That is, high P3b and LP predicted that subjects would propose more selfishly after acceptance and more generously after rejection (P2: Odds Ratio [OR] = .004, $p = .24$; P3b: OR = .011, $p = .049$; LP: OR = .017, $p = .009$; Figure 3; Table 1). Next, given that the behavioral results showed that rejection had a greater impact on behavioral change than acceptance, we tested whether similar patterns exist in the ERP data by incorporating an ERP × Response interaction as a predictor within the regression. This revealed that increased P2 specifically predicted behavioral change after rejection (interaction: OR = −.023, $p = .005$) but not in general (main effect: OR = .006, $p = .504$). This suggests that the role of P2 was tied to negative feedback. On the other hand, while the effects of P3b and LP were slightly larger for rejections (interactions: P3b: OR = −.016, $p = .011$; LP: OR = −.010, $p = .076$ [marginal]), main effects remained significant (main effects: P3b: OR = .014, $p = .074$ [marginal]; LP: OR = .010, $p = .010$), meaning that P3b and LP play a general role for both positive and negative feedback. These patterns replicate using analyses that treat centroparietal positivity as the dependent variable rather than as a predictor (i.e., linear regressions; Supporting Information: Results 2.1). Altogether, these findings demonstrate that centroparietal mechanisms play a key role in behavioral change, in a manner that is consistent with the behavioral findings above.

Additionally, we found analogous across-subject evidence linking centroparietal positivity and behavioral change. Namely, subjects who showed increased centroparietal positivity also tended to change their behavior in response to feedback more often (P2: $r = .45$, $p = .005$; P3b: $r = .42$, $p = .010$; LP: $r = .38$, $p = .021$; Figure 4a). This effect remains if only considering trials wherein the proposed offer was rejected (P2: $r = .45$, $p = .007$; P3b: $r = .41$, $p = .015$; LP: $r = .36$, $p = .035$; Figure 4b). The effect numerically dampens if examining only accepted offers (P2: $r = .37$, $p = .026$; P3b: $r = .31$, $p = .07$; LP: $r = .25$, $p = .13$), but this is to be expected given the earlier results on rejection sensitivity. To confirm that these rejection-based results were not due to individual differences in the offer amounts that subjects proposed, a multiple regression that controlled for this was performed, and the same patterns emerged (P2: β [standardized] = .46, $p = .009$; P3b: $\beta = .42$, $p = .015$; LP: $\beta = .38$, $p = .026$). Taken together, these converging pieces of both within- and across-subject evidence show that established centroparietal P2, P3b, and LP patterns linked to direct learning also appear within this UG context, where fairness norms are relevant. Additionally, these findings create a foundation that can be compared to correlates of conformity.

3.2.2 | Conformity results

Confirming our second ERP hypothesis, we found that increased centroparietal positivity, when Responder subjects processed received offers, predicted conformity-related changes in their subsequent behavior as Proposer.

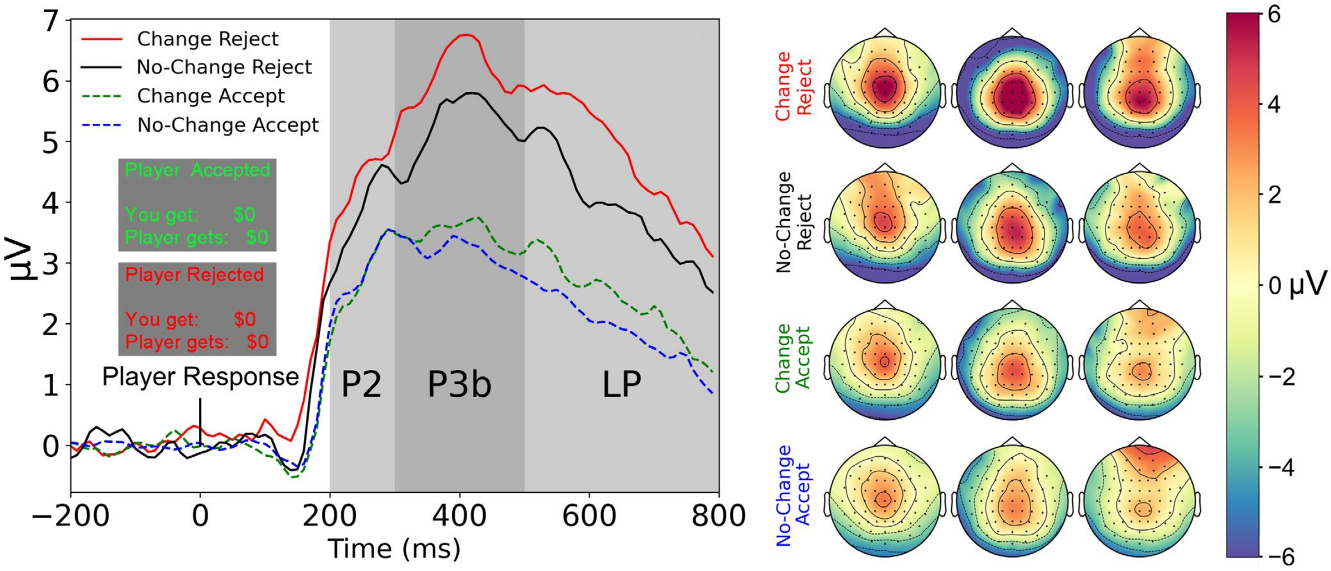

FIGURE 3 ERP correlates of direct behavioral learning following proposed offer acceptance versus rejection. Waveforms dissociate between Proposer trials where subjects subsequently changed their behavior based on feedback (proposed more selfishly after acceptance or more generously after rejection) versus trials where this did not occur (subjects either did not change their behavior or changed it in the opposite direction). P2 exclusively predicted behavioral learning after rejections but not after acceptance. P3b and LP showed more general effects common to both types of feedback. As our analyses relied on controlling for other variables that may influence the ERP and behavioral data—e.g., the offer amount proposed and the offer x response interaction—these covariate effects were removed via subject-by-subject regressions before plotting. In the Supporting Information, Figure S1 shows the same waveforms based on just the Cz electrode, which was representative

TABLE 1 Increased centroparietal positivity underlies behavioral change prompted by both direct feedback and conformity

	P2		P3b		LP	
	Change	No-change	Change	No-change	Change	No-change
Feedback	3.17	2.89	4.34*	3.53	3.32**	2.33
Only accept	2.63	2.76	3.50	3.26	2.60	2.06
Only reject	4.09	3.56	5.93	5.09	4.88	3.88
Conformity	1.89*	1.61	2.43*	2.06	2.74*	2.21
Only accept	1.85	1.62	2.25	2.15	2.36	2.21
Only reject	2.18	1.91	2.99	2.41	3.46	2.76

Note: The data reflect the waveforms shown in Figures 3 and 5. Significance markings for the bolded rows reflect the primary Change versus No-Change and Conformity versus No-Conformity logistic regression results.

For completeness, "only accept" and "only reject" results are included for the conformity analysis, although some conditions (e.g., Conformity-Change + only reject) included no trials for one or two participants. These participants were excluded.

* $p < .05$; ** $p < .01$.

Specifically, logistic regression revealed that P2 (OR = 0.02, p = .013), P3b (OR = 0.02, p = .018) and LP (OR = 0.02, p = .022) amplitudes predicted subsequent conformity (Figure 5; Table 1). Furthermore, these findings replicate using analyses that treat centroparietal positivity as the dependent variable (t-tests; Supporting Information: Results 2.1) or analyses that use alternative approaches to defining conformity (Supporting Information: Results 2.3 and 2.4).

Again, paralleling our approach for direct learning, we examined across-subject correlations and found that subjects who showed increased centroparietal positivity upon receiving an offer were also more likely to conform to others' behavior (P2: r = .37, p = .028; P3b: r = .42, p = .007; LP: r = .48, p = .004; Figure 6). This link is also identified if we exclude the subjects omitted from the within-subject analysis (P2: r = .32, p = .066; P3b: r = .41, p = .017; LP: r = .38, p = .028). Next, we confirmed that these across-subject positivity effects indeed reflect an overlap between direct behavioral learning and conformity, rather than some other aspect of individual differences or subjects' overall levels of task engagement. This was done by testing whether subjects' frequencies of changes in their

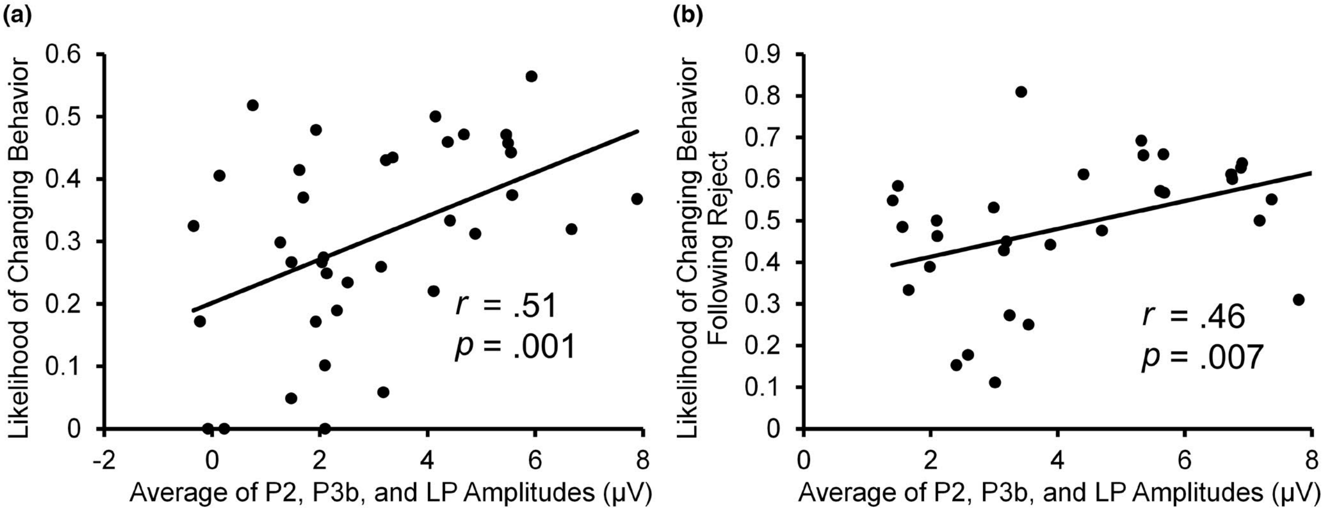

FIGURE 4 Across-subject patterns linking centroparietal positivity to direct learning. ERP amplitudes are shown as the averages of the P2, P3b, and LP components, as similar patterns were generally identified for each one individually. (a) Subjects who show higher centroparietal amplitudes when observing whether proposed offers are accepted or rejected tend to change their Proposer behaviors more frequently. (b) This direct learning effect is also identified if examining only trials where subjects' proposed offer was rejected

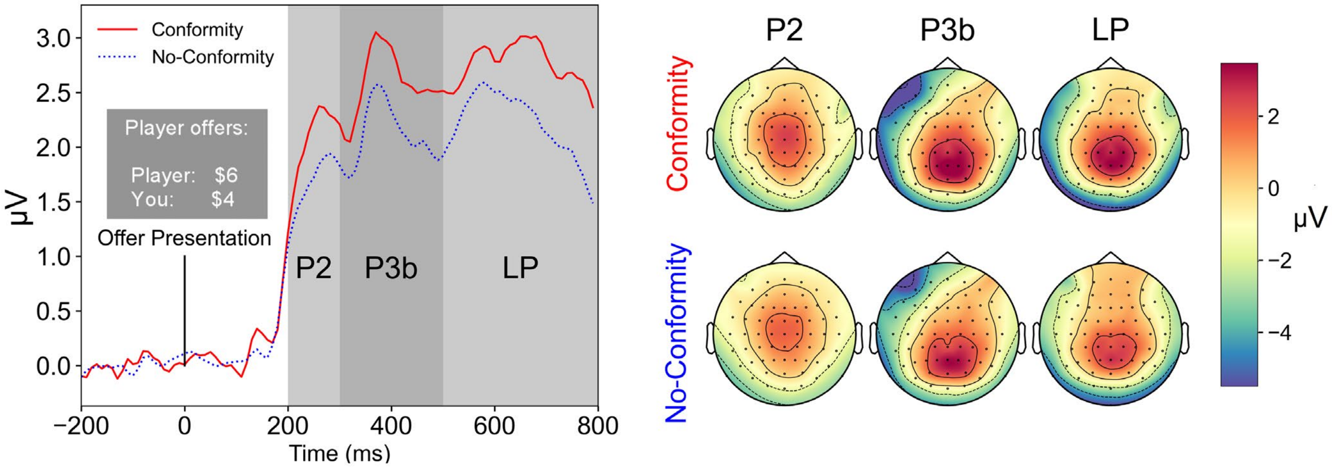

FIGURE 5 ERP correlates predicting subsequent conformity. Increased centroparietal responses elicited by received offers predicted that subjects would subsequently conform their Proposer behavior to this offer. This pattern was present within the P2, P3b, and LP time ranges. The waveforms are plotted at a 10 ms resolution

Proposer behavior was correlated with their average centroparietal response to an unrelated Proposer trial task screen, which generated the expected null results when examining either all trials or just rejected trials (*p*s > .35). In summary, we found that conformity involves centroparietal P2, P3b, and LP, which overlaps with the ERP correlates of direct behavioral learning and suggests common neural mechanisms.

4 | DISCUSSION

The goal of the present study was to investigate the links between the neural correlates of direct learning and conformity in terms of predicting behavioral change. Our novel use of the Ultimatum Game motivated subjects to continuously change their behavior based on both direct feedback and observations, which prompted conformity. This design permitted analyses that were not possible in studies separately investigating these types of learning. Regarding the behavioral results, subjects changed their Proposer behavior in response to feedback and based on what they saw their partners propose. Regarding the EEG data, first, we found that increased centroparietal positivity (P2, P3b, and LP) to direct feedback predicted subsequent changes in Proposer behavior. Additionally, subjects who showed increased centroparietal positivity to feedback, on average, tended to change their behavior more frequently.

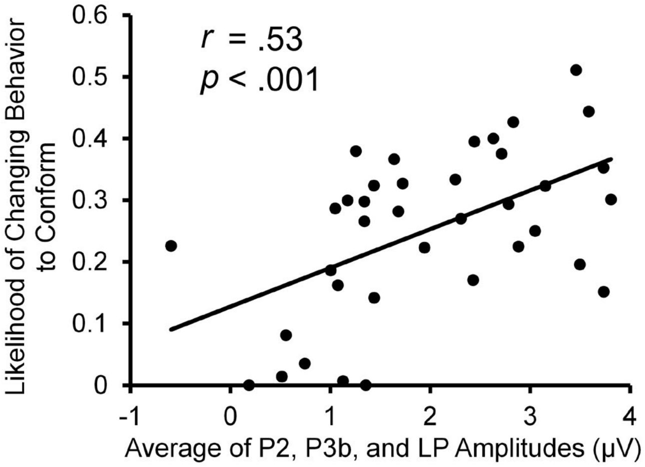

FIGURE 6 Across-subject patterns linking centroparietal positivity to conformity. Subjects who show increased centroparietal positivity are more likely to conform to the other player's behavior. This result on conformity parallels our earlier findings on the neural correlates which predict behavioral change due to direct feedback (Figure 4)

Second, we found similar patterns linked to conformity. Centroparietal positivity when processing others' offers predicted that subjects would subsequently conform to the other person, and also differentiated among subjects who frequently versus rarely conformed to others. In sum, centroparietal positivity plays a key role in both direct behavioral learning and conformity. This conclusion is supported by converging evidence from both within-subject analyses and across-subject correlations—statistical tests which were, notably, independent of one another. We describe the implications of these findings and how they can be integrated with the relevant literature.

The present study adds to the research pointing to overlapping neural systems involved in multiple types of learning, and it demonstrates that direct feedback and conformity drive behavioral change via common mechanisms. This focus on behavioral change notably contrasts prior decision-making research investigating the links between direct learning, observational learning, and conformity. This earlier work predominantly focused on similarity with regard to error processing, updating decision values, and learning the task structure (Bellebaum et al., 2010; Burke et al., 2010; Klucharev et al., 2009; Wu et al., 2016). By focusing instead on centroparietal positivity and behavioral change, our results add to the depth of this close similarity. We specifically found increased P2, P3b, and LP amplitudes when processing direct feedback or when observing others' behavior both predict that subjects will subsequently change their behavior. These findings were highly robust and replicated across a wide variety of analytic approaches, including regressions that account for both direct learning and conformity in unison or regressions that predict behavior change in terms of ERP responses modulating the impact of past observations (Supporting Information: Results 2.1–2.4).

Our findings on direct behavioral learning extend earlier research showing that centroparietal positivity predicts learning with probabilistic gambling tasks (Chase et al., 2011; Donaldson et al., 2016; San Martín et al., 2013). However, unlike this previous work, our study focused on subjects changing their behavior within a social context, where there notably is no objectively correct manner in which to act. Hence, our findings demonstrate that centroparietal positivity indexes a general behavioral change system that is active across a variety of task structures and contexts. To a degree, this is to be expected, given that social reward processing involves similar neural pathways as non-social processing (Behrens et al., 2008; Izuma et al., 2008). Nonetheless, confirming this point is useful.

Our findings on the mechanisms of conformity promoting behavioral change extend earlier ERP literature on conformity, which used opinion-based tasks and similarly reported P3b and LP effects (Pierguidi et al., 2019; Wang et al., 2020; Yuan et al., 2019). Beyond the ERP literature, most fMRI studies on conformity have focused on error processing—for example, the effects of one's choice being different from the group—or to a lesser extent, on the neural correlates predicting opinion changes (reviewed by Wu et al., 2016). To the best of our knowledge, just one prior fMRI study examined behavioral conformity, and it notably found that when subjects processed others' behaviors, increased temporoparietal junction (TPJ) activity predicted conformity (Wei et al., 2013). As the TPJ is thought to be a neural generator of P3b and LP (Linden, 2005), these earlier results are consistent with our findings. Additionally, because our study also examined behavioral change motivated by direct feedback and demonstrated overlapping effects between the two, this speaks to the interpretation of the neural patterns and suggests that the theoretical viewpoints used to understand behavioral change due to feedback processing are also relevant to understanding conformity. For example, identifying overlapping centroparietal positivity effects for both feedback processing and conformity may suggest that social conformity involves "context-updating," a theoretical mechanism thought to be indexed by P3b (Donchin, 1981; Polich, 2007) and potentially also LP (Hajcak & Foti, 2020). As P3b is additionally involved in model-based decision-making (Eppinger et al., 2017), this may suggest that social conformity operates on model-based processes. More broadly, this P3b and LP overlap suggests that the literature on direct learning may be useful for understanding conformity and vice versa.

Along with the effects identified within those later windows, P2 was notably found to predict changes in

Proposer behavior due to direct feedback processing, but only for rejected trials. Identifying this direct behavioral learning role of P2 is consistent with previous work (Donaldson et al., 2016; San Martín et al., 2013). However, its specific association with rejected trials is somewhat surprising, given that positivity within this time range is thought to be associated with rewarding outcomes ("Reward positivity"; Donaldson et al., 2016; Heydari & Holroyd, 2016). This "reward positivity" interpretation is not universally supported, as some authors have identified similar P2 sensitivity with gains versus losses (San Martín et al., 2013) or increased P2 sensitivity with losses (Martinez-Selva et al., 2019; Schuermann et al., 2012), although these latter findings are exceptional cases.

The link between offer rejection and P2 amplitude may be explained by our task's social-fairness aspect: when subjects' proposed offers are rejected, they may believe that the other player is being unfair. Unfairness can elicit anger (Pillutla & Murnighan, 1996), which enhances early centroparietal positivity (Angus et al., 2015; Tsypes et al., 2019) and may encourage behavioral change (Lench et al., 2016). Anger could also explain why a P2 effect was identified in our experiment, but not in the earlier studies of conformity that did not involve sensitive topics such as fairness (Pierguidi et al., 2019; Wang et al., 2019)—that is, receiving a selfish offer would elicit anger, which in turn prompts subjects to propose selfishly themselves. Alternatively, if P2 reflects reward positivity in this case, that may also explain why we found P2 involvement in conformity. Subjects are more likely to change their behavior to match others if they feel a sense of gratitude (Valk et al., 2017), and thus increased P2 amplitude may reflect positive emotional responses promoting conformity.

4.1 | Caveats

First, further clarification is needed as to whether the centroparietal effects should be interpreted as a single large component or three separate ones (P2, P3b, and LP). Although the present findings point to possible specificity regarding P2 differences between learning in response to negative feedback (reject) versus positive feedback (accept), further research is needed to clarify this dissociation. It is known that subjects perceive avoiding losses as being more important than acquiring gains (Tversky & Kahneman, 1992), and some previous research has corrected for this by having losses be in smaller quantities (e.g., gains as +1,250 points and losses as −625 points; Tunison et al., 2019), but this was not possible within the current task. Hence, it is unclear whether perceived value or gaining versus losing itself is responsible for the P2 dissociation between accepted versus rejected trials. Concerning the correlation findings, our interpretation is that they reflect subjects' overall learning tendencies. However, levels of task engagement could potentially confound these results, as engagement could simply upregulate both centroparietal amplitude and behavioral change. Although our correlation analyses of the non-learning screens provide evidence against this alternative explanation, future research would benefit from designs that similarly rule out other alternative possible explanations.

5 | CONCLUSION

In summary, we investigated the neural processes associated with behavioral change motivated by direct feedback and conformity using a novel role-swapping UG task. This allowed subjects to change their behavior in response to both types of learning events. We found overlapping centroparietal positivity effects associated with behavioral changes in response to both direct feedback and from a drive to conform to others' behaviors. These overlapping robust effects spanned the P2, P3b, and LP time windows. Moreover, exploratory analyses identified FMT effects, which exclusively predicted behavioral change in response to direct feedback but not conformity (see Supporting Information: Discussion). Taken together, our results shed light on how computational systems involved in direct behavioral learning may be repurposed for learning via social information. Additionally, our results suggest that past findings on direct learning are likely to be relevant for understanding conformity and vice versa. Overall, these findings suggest that future studies on learning would benefit from describing how both forms can emerge out of a single overarching framework rather than addressing the two in isolation.

ACKNOWLEDGMENTS

This research was carried out in part at the University of Illinois' Beckman Institute for Advanced Science & Technology. During the preparation of this manuscript, P.C.B. was supported by a Predoctoral Fellowship provided by the Beckman Foundation, M.M. was supported by Pre- and Postdoctoral Fellowships provided by the Beckman Foundation, I.K. was supported by a Fulbright Scholarship, and F.D. was supported by a Helen Corley Petit Scholarship in Liberal Arts and Sciences and an Emanuel Donchin Professorial Scholarship in Psychology from the University of Illinois. The authors wish to thank members of the Dolcos Lab for assisting with data collection and preprocessing.

AUTHOR CONTRIBUTIONS

Paul C. Bogdan: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft.

Matthew Moore: Conceptualization; Investigation; Methodology; Validation; Writing – review & editing. **Illya Kuznietsov:** Conceptualization; Funding acquisition; Investigation. **Justin D. Frank:** Formal analysis; Software. **Kara D. Federmeier:** Methodology; Writing – review & editing. **Sanda Dolcos:** Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing – review & editing. **Florin Dolcos:** Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Visualization; Writing – original draft; Writing – review & editing.

DATA AVAILABILITY STATEMENT

The data sets generated and analyzed during the current study, along with the analysis code, are available to qualified researchers from the corresponding authors, upon manuscript publication and Institutional Review Board (IRB) approval.

ORCID

Paul C. Bogdan https://orcid.org/0000-0002-4362-6084
Matthew Moore https://orcid.org/0000-0002-8499-0464
Illia Kuznietsov https://orcid.org/0000-0002-6113-0322
Kara D. Federmeier https://orcid.org/0000-0002-7815-1808
Sanda Dolcos https://orcid.org/0000-0002-7646-8246
Florin Dolcos https://orcid.org/0000-0003-2230-4139

REFERENCES

Angus, D. J., Kemkes, K., Schutter, D. J., & Harmon-Jones, E. (2015). Anger is associated with reward-related electrocortical activity: Evidence from the reward positivity. *Psychophysiology*, *52*(10), 1271–1280. https://doi.org/10.1111/psyp.12460

Bailey, A. H., & Kelly, S. D. (2017). Body posture and gender impact neural processing of power-related words. *The Journal of Social Psychology*, *157*(4), 474–484. https://doi.org/10.1080/00224545.2016.1242469

Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. *1*). Prentice-Hall.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. arXiv preprint. https://doi.org/10.18637/jss.v067.i01

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Bellebaum, C., & Colosio, M. (2014). From feedback-to response-based performance monitoring in active and observational learning. *Journal of Cognitive Neuroscience*, *26*(9), 2111–2127. https://doi.org/10.1162/jocn_a_00612

Bellebaum, C., Jokisch, D., Gizewski, E., Forsting, M., & Daum, I. (2012). The neural coding of expected and unexpected monetary performance outcomes: Dissociations between active and observational learning. *Behavioural Brain Research*, *227*(1), 241–251. https://doi.org/10.1016/j.bbr.2011.10.042

Bellebaum, C., Kobza, S., Thiele, S., & Daum, I. (2010). It was not MY fault: Event-related brain potentials in active and observational learning from feedback. *Cerebral Cortex*, *20*(12), 2874–2883. https://doi.org/10.1093/cercor/bhq038

Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208. https://doi.org/10.1002/bdm.621

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, *107*(32), 14431–14436. https://doi.org/10.1073/pnas.1003111107

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology - Paris*, *109*(1–3), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience*, *8*(3), 277–284. https://doi.org/10.1093/scan/nsr094

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, *23*(4), 936–946. https://doi.org/10.1162/jocn.2010.21456

Chierchia, G., Piera Pi-Sunyer, B., & Blakemore, S.-J. (2020). Prosocial influence and opportunistic conformity in adolescents and young adults. *Psychological Science*, *31*(12), 1585–1601. https://doi.org/10.1177/0956797620957625

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*(2), 174–186. https://doi.org/10.1111/psyp.12001

Donaldson, K. R., Oumeziane, B. A., Hélie, S., & Foti, D. (2016). The temporal dynamics of reversal learning: P3 amplitude predicts valence-specific behavioral adjustment. *Physiology & Behavior*, *161*, 24–32. https://doi.org/10.1016/j.physbeh.2016.03.034

Donchin, E. (1981). Surprise!... surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Dunne, S., D'Souza, A., & O'Doherty, J. P. (2016). The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. *Journal of Neurophysiology*, *115*(6), 3195–3203. https://doi.org/10.1152/jn.00046.2016

Eppinger, B., Walter, M., & Li, S.-C. (2017). Electrophysiological correlates reflect the integration of model-based and model-free decision information. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(2), 406–421. https://doi.org/10.3758/s13415-016-0487-3

FeldmanHall, O., Otto, A. R., & Phelps, E. A. (2018). Learning moral values: Another's desire to punish enhances one's own punitive behavior. *Journal of Experimental Psychology: General*, *147*(8), 1211–1224. https://doi.org/10.1037/xge0000405

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., & Parkkonen, L. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Guazzini, A., Panerati, S., Filindassi, V., Collodi, S., & Levnajic, Z. (2019). *Social norm spreading in real and virtual environments: Pro-social versus pro-self norm*. Paper presented at the International Conference on Internet Science

Hajcak, G., & Foti, D. (2020). Significance?... significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: An integrative review. *Psychophysiology*, *57*(7), e13570.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808

Heydari, S., & Holroyd, C. B. (2016). Reward positivity: Reward prediction error or salience prediction error? *Psychophysiology*, *53*(8), 1185–1192. https://doi.org/10.1111/psyp.12759

Huberth, M., Dauer, T., Nanou, C., Román, I., Gang, N., Reid, W., Wright, M., & Fujioka, T. (2019). Performance monitoring of self and other in a turn-taking piano duet: A dual-EEG study. *Social Neuroscience*, *14*(4), 449–461. https://doi.org/10.1080/17470919.2018.1492968

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*(2), 284–294. https://doi.org/10.1016/j.neuron.2008.03.020

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*(1), 1–21. https://doi.org/10.1111/psyp.12147

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

Koban, L., Pourtois, G., Bediou, B., & Vuilleumier, P. (2012). Effects of social context and predictive relevance on action outcome monitoring. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(3), 460–478. https://doi.org/10.3758/s13415-012-0091-0

Kröger, A., Bletsch, A., Krick, C., Siniatchkin, M., Jarczok, T. A., Freitag, C. M., & Bender, S. (2013). Visual event-related potentials to biological motion stimuli in autism spectrum disorders. *Social Cognitive and Affective Neuroscience*, *9*(8), 1214–1222. https://doi.org/10.1093/scan/nst103

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1494–1502. https://doi.org/10.18637/jss.v082.i13

Lench, H. C., Tibbett, T. P., & Bench, S. W. (2016). Exploring the toolkit of emotion: What do sadness and anger do for us? *Social and Personality Psychology Compass*, *10*(1), 11–25. https://doi.org/10.1111/spc3.12229

Levorsen, M., Ito, A., Suzuki, S. & Izuma, K (2021). Testing the reinforcement learning hypothesis of social conformity. *Human Brain Mapping*, *42*(5), 1328–1342.

Linden, D. E. (2005). The P300: Where in the brain is it produced and what does it tell us? *The Neuroscientist*, *11*(6), 563–576. https://doi.org/10.1177/1073858405280524

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Martinez-Selva, J. M., Muñoz, M. Á., Sanchez-Navarro, J. P., Walteros, C., & Montoya, P. (2019). Time course of the neural activity related to behavioral decision-making as revealed by event-related potentials. *Frontiers in Behavioral Neuroscience*, *13*, 191. https://doi.org/10.3389/fnbeh.2019.00191

Mas-Herrero, E., & Marco-Pallarés, J. (2014). Frontal theta oscillatory activity is a common mechanism for the computation of unexpected outcomes and learning rate. *Journal of Cognitive Neuroscience*, *26*(3), 447–458. https://doi.org/10.1162/jocn_a_00516

Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Morelli, S. A., Sacchet, M. D., & Zaki, J. (2015). Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. *NeuroImage*, *112*, 244–253. https://doi.org/10.1016/j.neuroimage.2014.12.056

Nicolle, A., Symmonds, M., & Dolan, R. J. (2011). Optimistic biases in observational learning of value. *Cognition*, *119*(3), 394–402. https://doi.org/10.1016/j.cognition.2011.02.004

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. https://doi.org/10.1016/j.jmp.2008.12.005

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, *21*(4), 197–212. https://doi.org/10.1038/s41583-020-0276-4

Oosterbeek, H., Sloof, R., & Van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, *7*(2), 171–188. https://doi.org/10.1023/B:EXEC.0000026978.14316.74

Peterburs, J., Frieling, A., & Bellebaum, C. (2021). Asymmetric coupling of action and outcome valence in active and observational feedback learning. *Psychological Research Psychologische Forschung*, *85*(4), 1553–1566. https://doi.org/10.1007/s00426-020-01340-1

Pierguidi, L., Guazzini, A., Imbimbo, E., Righi, S., Sorelli, M., & Bocchi, L. (2019). Validation of a low-cost EEG device in detecting neural correlates of social conformity. Paper presented at the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, *68*(3), 208–224. https://doi.org/10.1006/obhd.1996.0100

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rak, N., Bellebaum, C., & Thoma, P. (2013). Empathy and feedback processing in active and observational learning. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 869–884. https://doi.org/10.3758/s13415-013-0187-1

Sacconi, L., & Faillo, M. (2010). Conformity, reciprocity and the sense of justice. How social contract-based preferences and beliefs explain norm compliance: The experimental evidence. *Constitutional Political Economy*, *21*(2), 171–201. https://doi.org/10.1007/s10602-009-9080-x

San Martín, R., Appelbaum, L. G., Pearson, J. M., Huettel, S. A., & Woldorff, M. G. (2013). Rapid brain responses independently predict gain maximization and loss minimization during economic decision making. *Journal of Neuroscience*, *33*(16), 7011–7019. https://doi.org/10.1523/jneurosci.4242-12.2013

Schielzeth, H., & Forstmeier, W. (2008). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schmitz, J., Scheel, C. N., Rigon, A., Gross, J. J., & Blechert, J. (2012). You don't like me, do you? Enhanced ERP responses to averted eye gaze in social anxiety. *Biological Psychology*, *91*(2), 263–269. https://doi.org/10.1016/j.biopsycho.2012.07.004

Schuermann, B., Endrass, T., & Kathmann, N. (2012). Neural correlates of feedback processing in decision-making under risk. *Frontiers in Human Neuroscience*, *6*, 204. https://doi.org/10.3389/fnhum.2012.00204

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019). Herding brains: A core neural mechanism for social alignment. *Trends in Cognitive Sciences*, *23*(3), 174–186. https://doi.org/10.1016/j.tics.2019.01.002

Stewardson, H. J., & Sambrook, T. D. (2020). Evidence for parietal reward prediction errors using great grand average meta-analysis. *International Journal of Psychophysiology*, *152*, 81–86. https://doi.org/10.1016/j.ijpsycho.2020.03.002

Thoma, P., Norra, C., Juckel, G., Suchan, B., & Bellebaum, C. (2015). Performance monitoring and empathy during active and observational learning in patients with major depression. *Biological Psychology*, *109*, 222–231. https://doi.org/10.1016/j.biopsycho.2015.06.002

Tsypes, A., Angus, D. J., Martin, S., Kemkes, K., & Harmon-Jones, E. (2019). Trait anger and the reward positivity. *Personality and Individual Differences*, *144*, 24–30. https://doi.org/10.1016/j.paid.2019.02.030

Tunison, E., Sylvain, R., Sterr, J., Hiley, V., & Carlson, J. M. (2019). No money, no problem: Enhanced reward positivity in the absence of monetary reward. *Frontiers in Human Neuroscience*, *13*, 41. https://doi.org/10.3389/fnhum.2019.00041

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Valk, S. L., Bernhardt, B. C., Trautwein, F.-M., Böckler, A., Kanske, P., Guizard, N., Collins, D. L., & Singer, T. (2017). Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. *Science Advances*, *3*(10), e1700489. https://doi.org/10.1126/sciadv.1700489

Varni, J. W., Lovaas, O. I., Koegel, R. L., & Everett, N. L. (1979). An analysis of observational learning in autistic and normal children. *Journal of Abnormal Child Psychology*, *7*(1), 31–43. https://doi.org/10.1007/BF00924508

Vavra, P., Chang, L. J., & Sanfey, A. G. (2018). Expectations in the ultimatum game: Distinct effects of mean and variance of expected offers. *Frontiers in Psychology*, *9*, 992. https://doi.org/10.3389/fpsyg.2018.00992

Wang, L., Li, L., Shen, Q., Zheng, J., & Ebstein, R. P. (2019). To run with the herd or not: Electrophysiological dynamics are associated with preference change in crowdfunding. *Neuropsychologia*, *134*, 107232. https://doi.org/10.1016/j.neuropsychologia.2019.107232

Wang, Y., Cheung, H., Yee, L. T. S., & Tse, C.-Y. (2020). Feedback-related negativity (FRN) and theta oscillations: Different feedback signals for non-conform and conform decisions. *Biological Psychology*, *153*, 107880. https://doi.org/10.1016/j.biopsycho.2020.107880

Wei, Z., Zhao, Z., & Zheng, Y. (2013). Neural mechanisms underlying social conformity in an ultimatum game. *Frontiers in Human Neuroscience*, *7*, 896. https://doi.org/10.3389/fnhum.2013.00896

Wu, H., Luo, Y., & Feng, C. (2016). Neural signatures of social conformity: A coordinate-based activation likelihood estimation meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *71*, 101–111. https://doi.org/10.1016/j.neubiorev.2016.08.038

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *Journal of Neuroscience*, *33*(3), 1099–1108. https://doi.org/10.1523/jneurosci.1642-12.2013

Yuan, B., Wang, Y., Roberts, K., Valadez, E., Yin, J., & Li, W. (2019). An electrophysiological index of outcome evaluation that may influence subsequent cooperation and aggression strategies. *Social Neuroscience*, *14*(4), 420–433. https://doi.org/10.1080/17470919.2018.1488766

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

FIGURE S1 ERP correlates of Change versus No-Change based on just Cz. Waveforms dissociate between Proposer trials where subjects subsequently changed their behavior in accordance with feedback versus trials where this did not occur. The data reflects the exact same conditions and procedure as Figure 3, in the main text, but the waveforms are now based on just the Cz electrode data, rather than on the average of a centroparietal cluster

FIGURE S2 Topographic plots showing that frontal midline theta is linked to direct learning but not conformity. All topographic plots reflect the average of the 4–8 Hz theta range and the 200–300 ms time window. The topographic plots are shown for: (a) Change Proposer trials, (b) Conformity Responder trials, (c) No-Change Proposer trials, and (d) Non-Conformity Responder trials. (e) When seeing whether a proposed offer was accepted or rejected (feedback), increased FMT predicted that subjects would subsequently change their behavior. Red circles

indicate frontal electrodes showing significant differences ($p < .05$, two-tailed) between the change versus no-change conditions and also showing significant Change × Type interactions. Interestingly, significant differences were also identified in posterior locations, corresponding to electrodes O2 and PO8. (f) Finally, no significant differences linked to conformity were found for any electrode

FIGURE S3 Spectrograms showing that frontal midline theta is linked to direct learning but not conformity. (a) When seeing whether a proposed offer was accepted or rejected (feedback), increased FMT predicted that subjects would subsequently change their behavior. (b) On the other hand, when subjects received an offer, no meaningful FMT patterns were found, linked to conformity. Power was averaged across the F1, Fz, and F2 electrodes. Dark red areas indicate time-frequency bins associated with significant learning effects ($p < .05$, two-tailed). The dashed white rectangles show the window used for the *t*-tests reported in the text

How to cite this article: Bogdan, P. C., Moore, M., Kuznietsov, I., Frank, J. D., Federmeier, K. D., Dolcos, S., & Dolcos, F. (2021). Direct feedback and social conformity promote behavioral change via mechanisms indexed by centroparietal positivity: Electrophysiological evidence from a role-swapping ultimatum game. *Psychophysiology*, 00, e13985. https://doi.org/10.1111/psyp.13985