Neurophysiological correlates of collective perceptual decision-making

Kristina G. Baumgart1,2 | Petr Byvshev1,3 | Alexa-Nicole Sliby1,4 | Andreas Strube1,5 | Peter König1,2 | Basil Wahn1,6

1Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
2Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
3Department of Communications and Networking, Aalto University, Espoo, Finland
4Institute for Molecular and Cellular Cognition, Center for Molecular Neurobiology ZMNH, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
5Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
6Department of Psychology, University of British Columbia, Vancouver, BC, Canada

Correspondence
Kristina G. Baumgart, Institute of Cognitive Science, University of Osnabrück, Wachsbleiche 27, Osnabrück 49069, Germany. Email: baumgartkg@gmail.com

Funding information
We acknowledge the support of a research fellowship of the German Research Foundation (DFG) awarded to BW. Furthermore, we acknowledge the support by H2020-H2020-FETPROACT-2014 641321—socSMCs for PK and the support from the Deutsche Forschungsgemeinschaft (DFG).

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14545

[Correction added on 13 March 2020, after first online publication: URL for peer review history has been corrected.]

Abstract
Humans frequently perform tasks collaboratively in daily life. Collaborating with others may or may not result in higher task performance than if one were to complete the task alone (i.e., a collective benefit). A recent study on collective benefits in perceptual decision-making showed that dyad members with similar individual performances attain collective benefit. However, little is known about the physiological basis of these results. Here, we replicate this earlier work and also investigate the neurophysiological correlates of decision-making using EEG. In a two-interval forced-choice task, co-actors individually indicated presence of a target stimulus with a higher contrast and then indicated their confidence on a rating scale. Viewing the individual ratings, dyads made a joint decision. Replicating earlier work, we found a positive correlation between the similarity of individual performances and collective benefit. We analyzed event-related potentials (ERPs) in three phases (i.e., stimulus onset, response and feedback) using explorative cluster mass permutation tests. At stimulus onset, ERPs were significantly linearly related to our manipulation of contrast differences.

© 2019 The Authors. European Journal of Neuroscience published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
validating our manipulation of task difficulty. For individual and joint responses, we found a significant centro-parietal error-related positivity for correct versus incorrect responses, which suggests that accuracy is already evaluated at the response level. At feedback presentation, we found a significant late positive fronto-central potential elicited by incorrect joint responses. In sum, these results demonstrate that response- and feedback-related components elicited by an error-monitoring system differentially integrate conflicting information exchanged during the joint decision-making process.

**KEYWORDS**
collective benefit, confidence sharing, error positivity, error-monitoring, error-related negativity, joint action, social conflict

## 1 | INTRODUCTION

In the words of Lyn Collins popularized by Rob Base and DJ EZ Rock, “It takes two to make a thing go right.” In a world inundated by such popular culture references, it is evident that there is a common socio-cultural belief in the benefits of collaboration. We collaborate and perform joint tasks with each other often, whether it is friends moving a couch, parents deciding how to raise their child, or sports referees examining foul play. Accordingly, the dynamics of collaborative behavior are commonly studied to examine how groups working together have the potential to achieve higher performance than the best individual member in a group (i.e., collective benefit; Bahrami et al., 2010, 2012a, 2012b; Fusaroli et al., 2012; Pescetelli, Rees, & Bahrami, 2016). Research investigating group collaboration and its benefits encompasses many domains such as problem solving (Laughlin, Bonner, & Miner, 2002; Laughlin, Hatch, Silver, & Boh, 2006; Trouche, Sander, & Mercier, 2014), motor performance (Ganesh et al., 2014; Masumoto & Inui, 2013; Wahn, Karlinsky, Schmitz, & König, 2018; Wahn, Schmitz, König, & Knoblich, 2016) and perceptual tasks (Bang et al., 2014; Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Mahmoudi et al., 2015; Wahn, Czeszumski, & König, 2018; Wahn, Kingstone, & König, 2017, 2018; Wahn, Schwandt, et al., 2016). These studies suggest that information exchange and group dynamics are factors contributing to collective benefit.

In recent years, collective benefits have been frequently investigated in joint perceptual decision-making paradigms (Bahrami et al., 2010, 2012a, 2012b; Fusaroli et al., 2012; Pescetelli et al., 2016). Generally, a perceptual decision is one that consults sensory information and results in choosing one of many, or few, options (Heerken, Marrett, & Ungerleider, 2008). As perceptual decision-making paradigms have been extensively investigated for individuals performing the task alone, a solid basis (e.g., in terms of task procedure and analyses) for extending earlier individual decision tasks to joint tasks is provided (Bauer, Cheadle, Parton, Müller, & Usher, 2008; Heerken et al., 2008; Li, Han, Lei, Holroyd, & Li, 2011). Moreover, these paradigms allow researchers to control and manipulate the flow of information between co-actors, making them highly suitable to investigate collective benefits in joint perceptual decision-making tasks.

Using joint perceptual decision-making paradigms, researchers investigated several factors affecting collective benefit, such as the type and quality of information exchanged and performance similarities (Bahrami et al., 2010, 2012a, 2012b). In particular, over a series of studies, Bahrami and colleagues discuss and examine the role of joint decision-making and collective benefit within a contrast discrimination task (Bahrami et al., 2010, 2012a, 2012b). Dyads were tasked with individually identifying a contrast difference in a target stimulus which is present in one of two intervals, each displaying circularly arranged Gabor patches. In conditions of dyadic disagreement, dyads either reached a consensus verbally or by means of confidence sharing, via a graphic scale, when communicating non-verbally. Bahrami et al. (2010) found that dyads attained a collective benefit when co-actors were allowed to speak freely about the task, and a trend toward this benefit (Bahrami et al., 2012b) when co-actors communicated non-verbally through a confidence rating scale. Additionally, Bahrami et al. (2010) found that dyads with similar sensitivities were more likely to achieve a collective benefit than less similar dyads and that the constraints of these sensitivity similarities were softened in non-verbal conditions of the paradigm (Bahrami et al., 2012b). Taken together, these studies highlight how factors such as individual performance differences and the type of information exchanged between co-actors (i.e., verbally vs. non-verbally via confidence scales) influence collective benefits in perceptual decision-making.

Further studies have also investigated other factors affecting collective benefit in perceptual decision-making paradigms. As noted above, Bahrami et al. (2012b) found verbal communication to be the most successful mode in acquiring collective benefit, which led Fusaroli et al. (2012) to analyze linguistic links and dyadic verbal adaptation. They found...
that pairs who aligned themselves linguistically (i.e., communicated using similar language) achieved greater benefit. Pescetelli et al. (2016) examined the effect of social information on confidence decisions, finding those who were able to communicate reliably (i.e., more accurately communicate uncertainty) were able to achieve higher task performance. Furthermore, Bang et al. (2017) solidified that dyads who communicated their uncertainty similarly when using confidence rating scales obtained higher collective benefit only if they were also similarly sensitive to the task. This speaks to the importance of sensitivity similarity in the achievement of collective benefit. Overall, there are many important factors that affect a dyad’s ability to attain collective benefit.

In addition to investigating achievement of collective benefit in perceptual decision-making tasks and related factors, several models have been proposed to gain a better understanding of cooperative behavior in these tasks and its relation to collective benefits. In particular, using a signal-detection approach, Sorkin, Hays, and West (2001) proposed an optimal model, in which all information (i.e., accurate assessment of one’s own perceptual capabilities and uncertainties) is assumed to be available to the co-actors and perfectly exchanged between them. Thus, this optimal model sets the upper boundary of the collective benefit that can be attained. Yet, Bahrami et al. (2010) argued that in a real-world setting, joint decision-making bears different constraints and limitations, whereby collective benefit can be reduced or disappear entirely under certain conditions. For example, when there are large differences in the perceptual sensitivities of co-actors within a dyad, suboptimal weighting of information occurs as co-actors may still similarly weight their co-actors’ perceptual judgments, resulting in a reduction of collective benefit (Bahrami et al., 2010). In other words, people often assume that others’ have similar perceptual capabilities (and consequently weigh judgments similarly) but this is not always the case. Hence, Bahrami et al. (2010) proposed the weighted confidence sharing (WCS) model as a realistic alternative, accommodating for possible shortcomings in the exchange of information.

Concerning the physiological mechanisms of perceptual decision-making with investigation by EEG, certain cognitive processes have been found to be reflected in differences of electrical activity in the brain. Generally, neural correlates of decision-making can be investigated in different phases such as stimulus presentation, response and feedback. In the following, we introduce the correlates associated with these phases. For example, perceptual decision-making paradigms in which individuals performed decision-making tasks alone have revealed that attentional processes and visual awareness can be evaluated during stimulus presentation (Koivisto, Salminen-Vaparanta, Grassini, & Revonsuo, 2016; Philiastrides & Sajda, 2006; VanRullen & Thorpe, 2006). The source of top-down biasing signals in visual perception derives from a network of areas in the frontal and parietal cortices (Kastner & Ungerleider, 2000). Components found at electrode sites in parietal locations are reported to be primarily concerned with the spatial component of perception, while frontal components are related to target detection, alerting and motor representation (Kanwisher & Wojciek, 2000). Early components (N200 and P300) are typically linked to processes of visual awareness, global attentional processes and change detection (Eimer & Mazza, 2005; Koivisto & Revonsuo, 2010).

The decision-making process has been further analyzed on a physiological level during the response phase, where response and response conflict can be evaluated before and after a response is given. For example, cognitive processes occurring before a decision indicated with a motor command can lead to differences in the readiness potential (Bereitschaftspotential; Kornhuber & Deecke, 1965). This potential has been shown to reflect cognitive processes even before a motor response is given and is connected to motor preparation. Furthermore, response conflict has been associated with modulations of the readiness potential (Hackley & Valle-Inclán, 1999), which has implications for joint tasks in which contrary partner information could lead to a response conflict. In a review by Bartholow (2010), the readiness potential has been specifically discussed as being of interest in social cognitive tasks involving conflict or the need for cognitive control.

Several other components have been proposed to be associated with response error and performance monitoring after a response has been given. For example, the term rrror-related negativity (ERN) is used for a frontal negativity directly following response errors (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990). The ERN has been initially linked to error detection mechanisms, but ERN-like activity has also been found in correct trials (Bartholow et al., 2005). With this new evidence of correct-related negativity, the frontal response negativity stemming from the anterior cingulate cortex has instead been proposed as a response to fulfill a conflict-monitoring function (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Additionally, in response error monitoring, an error positivity (Pe) occurring after the ERN has been consistently found (Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). Some research suggests that the Pe may reflect the emotional significance of an error and is generally larger when participants are consciously aware that an error has been made (Orr & Carrasco, 2011). Moreover, Error-Related Feedback Negativity (fERN), which is similar to the response ERN and elicited after incorrect feedback (Mitnner, Braun, & Coles, 1997), is useful in evaluating error monitoring. In summary, differences in electrical brain activity during response and feedback phases, such as the ERN, ERN and Pe, are consistently related to error- and conflict-monitoring processes.

Some headway has also been made in understanding joint tasks on a neurophysiological level, and insights have been
used to establish a framework for understanding neurophysiological correlates of group tasks. It has been suggested that the portrayal of personal and observed actions occurs in a largely equivalent manner, where observation of an action elicits similar activation as performing the action itself (Sebanz, Knoblich, & Prinz, 2003). Sebanz, Knoblich, Prinz, and Wascher (2006) provide further evidence that shared representations of actions are formed between individuals when these actions are performed in a social context. Moreover, the neurophysiological representations of joint action have been investigated using a Go-NoGo task, where comparisons of task responses in various social contexts established that individuals display response inhibition when accommodating for the actions of a partner (Tsai, Kuo, Jing, Hung, & Tzeng, 2006). While these studies bring some insight to neurophysiological aspects of joint action and social processes, little is known specifically concerning joint decision-making.

Many researchers have investigated the neurophysiological correlates of joint action (Kourtis, Sebanz, & Knoblich, 2013; Picton, Saunders, & Jentzsch, 2012; Sebanz et al., 2006; Tsai et al., 2006). Many others (Bang et al., 2014; Fusaroli et al., 2012; Mahmoodi et al., 2015; Pescetelli et al., 2016) have extended the paradigm by Bahrami et al. (2010). However, none have yet quantified this particular behavior on a neurophysiological level. Therefore, the current study aims to bridge the gap by adapting the non-verbal condition of the Bahrami et al. (2012b) perceptual decision-making paradigm with the addition of EEG measurements. The Bahrami et al. (2012b) paradigm in particular allows us to break down joint decision-making and observe neurophysiological behavior at critical points during the decision-making process.

We first sought to adapt and reproduce the behavioral results of the paradigm. In particular, Bahrami et al. (2012b) found that as the sensitivity similarity between dyad members increases, the ratio of sensitivity between dyad and best member increases. As a second step, we assessed the metacognitive sensitivity of co-actors, which is a measure of the relationship between confidence and accuracy (Fleming & Dolan, 2012; Fleming & Lau, 2014; Pescetelli et al., 2016). Earlier studies found that there is a significant positive relationship between performance and confidence scale rating; the more confident the rating, the greater the performance. In the current study, we tested whether this is also the case for the present design. Bahrami et al. (2012b) compared results against model predictions of the proposed WCS and optimal models, finding neither model as a suitable predictor of joint performance under the parameters set for the study paradigm. In line with these analyses conducted by Bahrami et al. (2012b), and to further evaluate the possibility of achieving collective benefit, we compared the dyad performance with predictions of the WCS and optimal models.

Related to the neurophysiological correlates of collective decision-making, the paradigm by Bahrami et al. (2012b) offers a suitable design to investigate the neurophysiological underpinnings of collective perceptual decision-making. The exchange of information via confidence scales is especially suitable for EEG as this method avoids large motor artifacts that accompany speech. As a consequence, the design of the paradigm allows for the comparison of individual and joint decisions revealing neurophysiological and behavioral insights of a joint decision from start to completion. Despite its suitable design, other aspects of the original paradigm were adapted to accommodate the needs of the neurophysiological measurements. We exploratively examined electrophysiological potentials elicited by the fronto-parietal attention network and response- and error-related systems in the following phases: stimulus presentation, response and feedback. In particular, we first assessed EEG signals at the presentation of Gabor patches in the stimulus presentation phase, where cognitive, sensory and attentional processes could be encoded. Secondly, we exploratively examined EEG activity at individual and joint response phases—here, conflict- and error-monitoring processes might be at play. Finally, we investigated potentials at the presentation of feedback, where conflict- and error-monitoring processes may also be relevant.

2 | METHODS

2.1 | Participants

Participants were recruited from the student body of the University of Osnabrück, Germany. A total of 52 participants (33 female, age range 19–31 years) participated in the experiment after giving informed consent. All participants were healthy with normal or corrected-to-normal vision. Six pairs were excluded from analysis due to performing below chance level or failure to complete all experimental trials. Thus, results are reported from 40 participants, comprised of 20 dyads, where 62.5% of participants reported previously knowing their partner. Each dyad consisted of one co-actor for whom we measured EEG (“EEG participant”) and one co-actor for whom only behavioral data were collected (“non-EEG participant”). Participants were recruited by departmental email and independently recruited for each role of the dyad. Participants could request to be paired with a friend, and requests were accommodated when availability of both parties overlapped. Participants received either monetary compensation or course credit for their participation. The study was approved by the ethics committee of the University of Osnabrück.

2.2 | Experimental setup

The current study adapted a non-verbal signal-detection paradigm from Bahrami et al. (2012b), involving a two-interval forced-choice (2IFC) design (i.e., participants could
not remain undecided and were forced to choose from two options). Dyads were seated in a dark experimental room at the same table approximately 60 cm from individual display screens (BenQ 24”, resolution = 1,920 × 1,080, 60 Hz refresh rate). Displays were controlled via PsychoPy 1.85.1 (Peirce, 2008) for Python 3.5 (Python Software Foundations). A divider was placed between participants such that they were unable to view their co-actor’s display. As both displays were connected to the same graphics card, an additional black cardboard cover was used to mask one half of the screen. For technical reasons, the EEG participant used the left half of the left display while the non-EEG participant used the right half of the right display.

Participants were each assigned a color to represent their corresponding tasks and responses: the EEG participant was blue and the non-EEG participant was yellow. Responses were made using the right hand on individual response boxes. Participants were provided with earplugs to reduce external noise interference and were instructed not to talk. The experimental phase contained 16 blocks of 16 trials after completion of one training block of 16 trials, which allowed participants to become familiar with the task. Experimenters remained present during training to ensure participants understood experimental instructions. Completion of both the training and experimental phases required approximately one and a half to two hours. A fifteen-minute break was given after the eighth block and additional breaks could be taken when necessary.

2.3 | Stimuli

Stimuli were set to match those previously presented in the Bahrami et al. (2012b) study. The stimuli consisted of two sets of six vertically oriented Gabor patches (SD of Gaussian envelope = 0.38°, spatial frequency: 1.8 cycles/°, contrast: 10%), which were presented in circular formation (radius = 6°) equidistant from each other. Target stimuli were generated by increasing baseline contrast of a single Gabor patch in one of the following increments: 1.5%, 3.5%, 7% and 15%, with an increase of 1.5% representing more difficult trials (as the target is most similar to the distractor contrasts) and an increase of 15% representing trials which are typically easier. Location and interval in which the target stimulus appeared was randomized throughout the experiment. Contrast of the target stimulus was pseudorandomized such that all four contrasts occurred equally within an experimental session. Therefore, the difficulty of the task was randomized within a dyad such that all dyads experienced an equal number of trials per difficulty within an experimental session.

2.4 | Procedure

In the experiment, dyads performed a joint contrast discrimination task. In this task, participants were shown two intervals of stimuli in which one interval contained a target stimulus of different contrast. Dyads were first asked to individually indicate in which interval the contrast difference occurred and were then asked to indicate how confident they were in said decision. Following this, dyads were shown their individual decisions as well as that of their co-actor. One co-actor was then chosen to make a final joint decision for the dyad, indicating first in which interval the difference appeared and subsequently how confident the pair was. The choice was then displayed to both co-actors, and individual and joint decision feedback was given.

In detail, a white fixation cross marked the start of each trial. The non-EEG participant initiated each trial with a button press. The white fixation cross then turned black and was presented alone for a randomly jittered amount of time. After which, the first stimulus interval was presented for 85 ms, followed by a single black fixation cross for 1,000 ms, and the second stimulus interval for 85 ms (Figure 1). Participants then made a private individual “Interval Choice” indicating where the target stimulus occurred using a single button press, left for the first interval and right for the second.

Following this button press, a bidirectional confidence rating scale was presented requiring participants to indicate confidence in their interval choice. A marker in the respective color of the participant was included in the center of the scale. Participants were forced to choose an interval as defined by the 2IFC design and were only allowed to indicate confidence for their chosen interval. Private confidence ratings were made by moving the marker left for the “First” interval and right for the “Second” interval. The marker could be moved five increments in either direction; the farther from the central line, the more confident the decision. The experiment imposed no time limitations for any decisions made and therefore the experiment could not continue unless both participants made a decision, ergo interval forced-choice task.

After confidence rating submissions, individual interval choices were shared in the participants’ respective colors at the top or bottom of the screen. To avoid spatial bias (i.e., attention is allocated to a specific section of the display), the position of individual choices on the screen was randomized. Following this, individual confidence ratings were displayed and one dyad member, indicated by instruction color, was randomly chosen to make a joint decision for the dyad. As enacted in the individual decision, the joint decision consisted of an interval choice followed by a confidence rating. This confidence rating for the joint decision was chosen by the same participant who performed the joint decision. After the joint interval choice was made, the chosen interval was revealed to both participants. Both the interval choice and confidence rating were displayed after the respective choice. A black fixation cross was presented for 500 ms, followed by presentation of individual and joint performance feedback for 2,000 ms. Joint feedback was always presented in white
text in the center of the screen while individual feedback was presented in the participants' respective colors in randomized positions (top or bottom). Feedback informed participants whether their joint and individual decisions were correct or not (i.e., “CORRECT” or “WRONG” was displayed on the screen). After feedback offset, a white fixation cross signaled the end of the trial and the start of a new one.

2.5 Behavioral analysis

Behavioral analysis investigated factors contributing to collective benefit and compared the fit of WCS and optimal models, adapted as closely as possible from Bahrami et al. (2010, 2012b) and Sorkin et al. (2001). Contrast difference of the target stimulus, which varied by trial, was used as a measure of difficulty per trial to calculate a psychometric curve separately for joint and individual performance. The slope of the curve was then used to measure individual and dyad sensitivities. As in Bahrami et al. (2012b), performance was quantified as the proportion of second interval choices per reported locational contrast difference, which was positive if the target appeared in the second interval and negative if it appeared in the first. Collective benefit, which determined whether dyads achieved group benefit or not, was defined as the ratio of the dyad slope ($S_{dyad}$) to the more sensitive participant (i.e., the participant with the higher slope, $S_{max}$). Values above one indicate a collective benefit.

To further investigate factors contributing to collective benefit, participants’ use of the confidence rating scales was analyzed. We assessed in two ways how the similarity of confidence scale usage between co-actors of a dyad might contribute to collective benefit. The first approach compared the individual confidence value distributions of a dyad, defining a Distribution Difference Index (DDI), which quantifies the difference between the two distributions. This measure assessed scale-usage similarity as the difference in the number of times each respective confidence level was used between dyad members. The DDI was
plotted against collective benefit on a per dyad basis. A DDI value of zero would indicate perfectly identical usage of all confidence levels. The DDI is expressed as

$$\text{DDI}_l = \frac{\sum_{k=1}^{5}(a^k_i - b^k_i)^2}{\text{max}_l[\sum_{k=1}^{5}(a^k_i - b^k_i)]}$$ (1)

where $a^k_i$ and $b^k_i$ represent the proportion of answers given by each subject in the $i$-th pair at the $k$-th confidence level.

The second approach quantified how similarly dyads performed at any confidence level. To do this, we proposed the Confidence Performance Difference Index (CPDI), which captures the difference in using confidence levels by the summed squared deviations. The individual performance of each participant was plotted against expressed confidence rating per trial. Then, the difference vector for a dyad was calculated to evaluate the CPDI. A CPDI value of zero would indicate perfectly identical performance across all confidence levels. The CPDI, a variation of the DDI, is expressed as

$$\text{CPDI}_l = \frac{\sum_{k=1}^{5}(c^k_i - d^k_i)^2}{\text{max}_l[\sum_{k=1}^{5}(c^k_i - d^k_i)]}$$ (2)

where the CPDI$_l$ is a Confidence Performance Difference Index of $i$-th pair of subjects, $c^k_i$ and $d^k_i$ represent accuracy of subjects in the $i$-th pair at the $k$-th confidence level. Compared to more sophisticated measures like the Kullback–Leibler divergence, it avoids the complexities of handling division by zero, as is the case with the Kullback–Leibler divergence when one category is not used by a participant at all. Furthermore, the CPDI has the advantage of simplicity making it transparent to non-experts as well compared to the Kullback–Leibler divergence, with which many are unfamiliar.

Additionally, we sought to compare dyad performance in the present task to the performance predictions of the WCS and optimal models. The optimal model (for exact derivations of formulas, see Sorkin et al., 2001; Bahrami et al., 2012b) predicts dyad sensitivity as

$$S_{\text{dyad}}^{\text{opt}} = \sqrt{S_1^2 + S_2^2}$$ (3)

where $S_1$ and $S_2$ represent the sensitivities of each individual (Bahrami et al., 2012b; Sorkin et al., 2001). In this model, dyad sensitivity is expected to always be better than that of the best individual as it assumes that all necessary information is available and communicated to both co-actors. This then sets a maximal boundary of collective achievement. To acknowledge differences in dyad information exchange, the WCS model is weighted by shared confidence, expressing dyad sensitivity as

$$S_{\text{dyad}}^{\text{WCS}} = \frac{S_1 + S_2}{\sqrt(2)}$$ (4)

Given the equational variances, the models diverge further in instances of high dyad sensitivity difference (Bahrami et al., 2012b).

### 2.6 EEG data acquisition

Data were recorded using asalab™ with two 72 Refa amplifiers (ANT Neuro) and gel-based Ag/AgCl 128 electrode cap (waveguard™ caps), which were set based on the $10^{-5}$ international positioning (Oostenveld & Praamstra, 2001). For technical reasons, only 64 electrodes were used according to a 64 electrode cap layout. All recordings were referenced to the Cz electrode, with the ground positioned below the left clavicle utilizing a disposable cardiac electrode. Data were acquired at a 1,024 Hz sampling rate, and scalp impedances were kept below 10 kΩ. Eye movement was controlled through bipolar recordings of horizontal and vertical electrooculogram (EOG), made possible by placing disposable cardiac electrodes above and below the left eye, centrally aligned with the pupil.

### 2.7 EEG preprocessing

Analysis and processing of EEG data were performed using the FieldTrip toolbox for MATLAB (Oostenveld, Fries, Maris, & Schoffelen, 2011). Mastoidal electrodes M1 and M2 were excluded from analysis due to excessive noise. The raw data were resampled to 256 Hz and Butterworth filtered at the third order at 0.1 Hz (high-pass) and 20 Hz (low-pass). To later analyze the different phases in the decision process, the data set was epoched according to three phases, namely at the onset of Gabor patch presentation (stimulus presentation phase), at individual and joint responses of the EEG participant (response phase) and at the onset of feedback presentation (feedback phase).

Epochs ranged from $-2$ to $2 \text{ s}$ relative to the trial-defining trigger. These triggers were set at the onset of stimulus and feedback presentation as well as at the button press indicating individual and joint interval decisions. These epochs were then centered, that is the mean of each epoch was removed from the respective epoch as a first baseline correction in order to prepare the data for independent component analysis (ICA; Jung et al., 2000). Clear outliers representing technical or movement artifacts were rejected manually after visual inspection to further clean the data for the following ICA procedure. Subsequently, the EEG signal was separated into independent additive subcomponents using ICA. This allows for the identification of components clearly based on eye movement artifacts. Components containing prominent eye movement artifacts were then visually rejected.

Back-projection of the remaining non-artifactual components resulted in the corrected EEG data. All remaining trials with samples exceeding an absolute value of 80 μV in any EEG channel in a time frame from $-1$ to $1.5 \text{ s}$ were deemed artifactual and were excluded from further analysis. For each
subject and channel, the data were z-transformed, that is, the mean of all samples of the respective subject and channel was subtracted from each sample and this difference was divided by the standard deviation. Please note that large inter-individual differences exist in EEG signals (Melnik et al., 2017). However, as a consequence of this z-transformation all subjects contribute equally to the grand average. The data were split into four different bins: (a) the onset of Gabor patches with a baseline from −100 to 0 ms from stimulus onset, (b) individual decisions with a baseline from 0 to 100 ms from the button press indicating the individual decision, (c) joint decisions with a baseline from 0 to 100 ms from the button press and (d) the onset of feedback with a baseline ranging from −100 to 0 ms where all combinations of feedback were considered for analysis (Figure 2).

2.8 | EEG statistical analysis

Grand averages (GAs) were calculated by the unweighted mean of z-transformed data of each participant for each trial, phase and feature of interest separately. For each phase, we defined conditions that we aim to compare. Please note that the term “condition” here is used only to separate experimental trials on the basis of certain criteria, not that these are literally under experimental control. In particular, at stimulus presentation and at the individual decision, a feature of interest is represented in the correctness of the response, that is, conditions that will be compared were defined as trials with a correct or incorrect joint response. At feedback presentation and the joint decision button press, features of interests were the correctness of the individual and joint response and whether responses of participants conflicted. Hence, conditions are correct or incorrect individual and joint responses, respectively, and conflicting or non-conflicting responses.

All statistical tests were conducted in electrode space and were corrected for multiple comparisons using non-parametrical permutation tests of clusters of neuronal activity (Maris & Oostenveld, 2007). Please note that this procedure allows a data-driven investigation of the physiological substrate of joint decision processes without a bias to pre-selected electrodes or time frames. In this procedure, a statistical test is performed at every electrode and time point. Adjacent data points exceeding a pre-defined threshold are then considered as a cluster. Subsequently, a random permutation test is performed to evaluate the probability of all existing clusters exceeding the pre-defined threshold. Importantly, in contrast to other exploratory procedures it handles multiple comparison corrections intrinsically and efficiently. For the present purpose—analyzing EEG signals of a novel paradigm—it combines two important advantages that it is not dependent on a specific theoretical framework, but allows an unbiased investigation of the physiological decision process. Clusters were formed based on neighboring time-channel bins exceeding a parametric statistical test value corresponding to a $p$-value of .05. Clustering was restricted to samples having at least two

**FIGURE 2** Information about EEG data processing (baseline and event lock) as well as statistical tests performed are listed for each respective phase. For each test, cluster analysis intervals are specified in bold [Colour figure can be viewed at wileyonlinelibrary.com]
neighboring channels also exceeding the threshold, where neighbors of channels were defined by a triangulation algorithm implemented in the FieldTrip toolbox. Using Monte Carlo resampling, 5,000 matrices with randomly permuted condition labels within each subject were generated, as were corresponding clusters for each of these matrices.

For each cluster, cluster-level statistics were calculated using the cluster mass (Bullmore et al., 1999), which is the sum of statistical values of all samples included in a cluster. Clusters obtained from the original data matrix were compared with the permutation distribution of the largest cluster masses resulting from each random permutation of data labels. A $p$-value represents the proportion of clusters in random permutation exceeding the cluster mass value of those obtained from the original data matrix. This test was applied two-sided for negative and positive clusters separately, and resulting $p$-values were Bonferroni corrected by the factor 2 for the two-sided test accordingly.

Using this procedure in the stimulus presentation phase, a linear trend of the four levels of contrasts was evaluated for data locked to the onset of the presentation of Gabor patches using a dependent samples linear regression test implemented in the FieldTrip toolbox for MATLAB (Oostenveld et al., 2011). Prior to this test, conditions were averaged according to the four levels of contrasts. The test was applied to samples in a time frame ranging from 0 to 1,000 ms. Additionally, dependent sample $t$ tests were applied to compare conditions in different time frames. In particular, at the presentation of Gabor patches, conditions were averaged separately according to correct and incorrect individual responses for each participant, and the cluster permutation test was applied in a time frame from 0 to 1,000 ms.

In the response phase, correct and incorrect responses with EEG data locked to the button press indicating the individual decision were tested in a time frame from −1,000 to 1,500 ms. Data locked to the button press of the EEG participant indicating a joint decision were tested in four different ways. First, conditions were averaged according to conflicting and non-conflicting individual responses, where in conflicting trials individual responses differed while in non-conflicting trials they were the same. Secondly, conditions were averaged according to correct and incorrect joint decisions. Correct versus incorrect and conflicting versus non-conflicting contrasts were tested in a time frame ranging from −1,000 to 1,500 ms. Thirdly, the absolute difference of correct and incorrect responses versus the absolute difference of conflicting versus non-conflicting responses was assessed in a time frame ranging from 0 to 1,500 ms. Finally, an interaction of conflict and correctness was assessed using the same procedure, that is, the absolute difference between conflicting correct and incorrect trials was tested versus the absolute difference between non-conflicting correct and incorrect trials in a time frame ranging from 0 to 1,500 ms.

In the feedback phase, conditions were averaged either according to feedback representing the correctness of individual response or to feedback representing the correctness of joint response. Correct and incorrect feedback conditions were tested for significant differences for individual and joint feedback separately. Furthermore, individual conflicting feedback was tested against individual non-conflicting feedback, where individual responses disagree or agree, respectively. These tests of feedback were done using the cluster permutation procedure described above. Additionally, an interaction of individually conflicting conditions and correctness of the joint interval choice was tested using the same cluster permutation procedure and was further quantified using paired-samples $t$ tests. All cluster permutation tests at feedback onset were applied in a time frame ranging from 0 to 1,500 ms. ERP plots were obtained by averaging trials of a certain condition at first within a participant. Then, a grand average (GA) was produced by the unweighted average of the obtained averaged ERPs of each participant. These grand averages were plotted for visualization.

3 | RESULTS

3.1 | Behavioral results

Although this work focuses on the neurophysiological mechanisms supporting joint decisions, the experiments speak to questions on the behavioral level as well. Therefore, in the following section we report the behavioral results in close analogy to previous studies.

3.1.1 | Comparison of dyad with the better participant

Dyad sensitivity was compared to the sensitivity of the better participant to assess whether a dyad achieved collective benefit. On average, dyad sensitivity was not significantly different from the sensitivity of the better participant. Figure 3b left bar, effect size $-0.1458$, $t(19) = -1.517, p = .146$, one-sample $t$ test. That is, pairs did not significantly attain a collective benefit. In spite of this, a significant relationship between the ratio of individual sensitivity similarity ($S_{\text{min}}/S_{\text{max}}$) and the ratio of sensitivity between dyad and best member ($S_{dyad}/S_{\text{max}}$) was found. As the ratio of similarity increases, the ratio between dyad and best member increases; therefore, the possibility of achieving collective benefit also increases (Pearson $r = .877, p < .001, n = 20$).

3.1.2 | Confidence rating scale usage and performance vs. collective benefit

The relationship between performance and expressed confidence level was used as a measure indicating metacognitive sensitivity on an individual level (Fleming & Dolan, 2012; Fleming & Lau, 2014; Pescecelli et al., 2016). A positive relationship was found between performance and expressed
level of confidence, where accuracy increased as expressed confidence level increased, thus indicating higher metacognitive sensitivity (Figure 3c). Following this, the Distribution Difference Index (DDI) was used to assess how confidence scale usage might affect collective benefit. The DDI quantified this distribution by taking the difference in the number of times each confidence level between a dyad was used. The data do not suggest a significant relationship between the DDI and collective benefit (Figure 3d, Pearson $r = -0.3$, $p = 0.19$, $n = 20$). That is, how differently a pair distributed their confidences across the rating scale did not predict collective benefit.

Further investigation of confidence scale usage was assessed by the Confidence Performance Difference Index (CPDI). The CPDI quantified the difference in performance between a dyad for each expressed confidence level. Confidence level performance accuracy was used to calculate CPDI per dyad. The observed relationship between CPDI and collective benefit suggests that CPDI is a predictive measure of collective benefit in that, as CPDI increases the likelihood of achieving collective benefit decreases significantly (Figure 3e, Pearson $r = -0.62$, $p = 0.003$, $n = 20$). Other measures, such as the KL-divergence, have been commonly used to assess differences in distributions, similarly to our DDI and CPDI measures. We therefore used the KL-divergence to test whether we would get similar results as for the DDI and CPDI. Indeed, we found that the KL-divergence also predicted the collective benefit similarly to the CPDI (Figure S1, Pearson $r = -0.64$, $p = 0.003$, $n = 20$). Thus, supporting similar previous findings, in order to achieve a collective benefit, it is not decisive to use different confidence levels with similar frequency, but to use different confidence levels to indicate comparable levels of individual performance.

### 3.1.3 Comparison of dyad with weighted confidence sharing and optimal models

Dyad performance was compared to the predictions of the WCS and optimal models to further investigate the possibility of achieving collective benefit. The WCS model predicted significantly higher dyad performance than was observed,
Figure 3a,b right bar, effect size 0.2456, $t(19) = -4.82, p < .001$, one-sample $t$ test, although it seems to predict a similar linear dependency on performance similarity. Considering the optimal model predicts higher overall collective benefit in terms of relative sensitivity difference compared to the WCS, the optimal model also predicted significantly higher dyad performance than was observed, Figure 3a, $t(19) = -5.3, p < .001$, one-sample $t$ test. Thus, in spite of the significant deviations, of the models tested, the WCS model performs best and specifically describes a similar linear dependency on performance similarity.

3.1.4 | Correctness vs. familiarity

Dyad familiarity was evaluated to exclude it as a confounding variable affecting achievement of collective benefit. In a post-experiment survey, participants were asked to what degree, if at all, they had previously known their partner. No significant difference in dyad performance was found based on reported familiarity ($r = -.12, p = .61, n = 20$).

3.2 | EEG results

3.2.1 | Stimulus presentation

We first sought to investigate cortical mechanisms at stimulus presentation. At the presentation of Gabor patches, no significant cluster representing a significant difference in EEG signals between correct and incorrect individual decisions was found (all $p$-values > .085). The linear regression statistics for the cluster analysis of dependent samples of different contrast levels at the presentation of Gabor patches revealed two significant clusters, one positive and one negative, corresponding to a dipole. The positive cluster included samples from electrodes in fronto-central to parieto-central positions, 297–859 ms after stimulus onset (Cluster Mass (CM) = 4,530; $p = .002$; confidence interval of $p$ (CI) = [0.0008; 0.0032]). The negative cluster included samples from left-temporal areas 109–196 ms after stimulus onset (CM = -3,975; $p < .001$; CI = [0.0002; 0.0016]). The highest parametric positive statistical $t$-value was found at channel AF4 at 488 ms, Figure 4 top, $t(19) = 5.88, p < .001, d = 1.31$. The highest parametric negative statistical $t$-value was found at channel FT7 at 353 ms, Figure 4 bottom, $t(19) = -6.75, p < .001, d = -1.51$. These results indicate a modulation of the EEG signals by the experimental manipulation of contrast differences at a relatively late latency of topographically widespread event-related potentials.

3.2.2 | Individual decision

At the button press indicating the individual interval decision, we investigated whether EEG signals were modulated by the correctness of the response. The permutation test of clusters composed of $t$-values obtained from dependent samples $t$ tests revealed one significant negative cluster, thus indicating lower values for correct individual decisions, as compared to incorrect individual decisions, at central to fronto-central right hemispheric electrodes from 578 to 1,097 ms after response (CM = -3,296.6; $p = .009$; CI = [0.0062; 0.0114]). The highest parametric negative statistical $t$-value was found at channel C4 at 1,018 ms, Figure 5, $t(19) = -5.01, p < .001, d = -1.23$. We can conclude that event-related potentials elicited after the individual response were modulated by the correctness of the response. No differences have been found before the response despite a time frame ranging well before the response trigger was tested, indicating that response preparation was not affected by factors of correctness or conflict.

3.2.3 | Joint decision

At the joint decision, we conducted tests to find manifestations of conflicting individual decisions as well as of the correctness of the joint decision within event-related potentials. Similarly to the reported difference after the individual decision, a significant negative cluster was found representing a significant difference between correct and incorrect joint interval decisions after the response of the EEG participant. This suggests lower values for correct joint interval decisions as compared to incorrect joint interval decisions at centro-parietal electrodes, 699–1,223 ms after response (CM = -3,426; $p = .0156$; CI = [0.0112; 0.019]). The highest parametric negative statistical $t$-value was found at channel CPz at 781 ms, Figure 6, $t(19) = -6.57, p < .001, d = -1.47$. The permutation test of clusters representing differences between conflicting and non-conflicting conditions revealed two significant clusters, one positive and one negative. The positive cluster, with a broad frontal topography, indicated a higher value for non-conflicting as compared to conflicting conditions, 106–391 ms after response (CM = 3,561.7; $p = .0488$; CI = [0.0428; 0.0548]). The highest parametric positive statistical $t$-value was found at channel AF7 at 195 ms, $t(19) = 6.438, p < .001, d = 1.44$. The negative cluster, corresponding to a higher value for conflicting conditions, was found at central to parietal electrodes, 82–1,500 ms after response (CM = -21,201; $p < .001$; CI = [0; 0.001]). The highest negative parametric statistical value was found at channel CP1 at 203 ms, $t(19) = -7.54, p < .001, d = -1.69$. Thus, both main effects of conflict, at frontal and parietal electrode sites, and correctness at parietal electrode sites, were found in the statistical test of event-related potentials. Again, only a time frame after the button press was found to be significantly affected by tested factors. We then tested whether those conditions
differed in their magnitudes. In order to do so, the absolute differences of these conditions were tested for significant differences using the previously introduced cluster permutation procedure.

This test yielded one negative cluster representing a significantly lower absolute difference between correct versus incorrect as compared to conflicting versus non-conflicting joint interval decisions, thus indicating a higher absolute difference between values of conflicting versus non-conflicting conditions as compared to correct versus incorrect conditions. This cluster included electrodes in frontal and parietal positions connected by a narrow bridging electrode in a time frame 328–406 ms after response (Figure 6, CM = -771.7, p = .013, CI = [0.01; 0.016]). For this effect, the highest negative parametric statistical value was found at channel F2 at 340 ms, t(19) = -4.57, p < .001, d = -1.02. This test indicated that the conflict condition elicited a higher difference in event-related potentials at both frontal and parietal electrode sites.

Finally, we investigated whether the different correctness and conflict trials interacted. There was a significant interaction 1,172–1,445 ms after response, indicating higher differences between correct and incorrect interval decisions in non-conflicting conditions as compared to conflicting conditions (CM = 946.5, p = .005, CI = [0.003; 0.007]). This interaction was found at centro-parietal electrode sites. The highest parametric statistical value was found at channel C2 at 1,293 ms, Figure 6, t(19) = 4.93, p < .001, d = 0.66. In conclusion, our results indicate differential representations of correctness and conflict in frontal and parietal event-related slow evoked potentials after the button press indicating the joint decision.

FIGURE 4 The averaged ERPs over each participant are shown for AF4 (top), and FT7 (bottom). Red and blue shaded boxes show significant time frames of the significant positive and negative clusters representing a linear regression of contrast differences, respectively. On the left, topographies of these clusters with averaged t-values are accordingly color-coded and significant electrodes are highlighted. Warm colors (yellow-red) represent positive t-values, cool colors (blue) negative t-values [Colour figure can be viewed at wileyonlinelibrary.com]
3.2.4 | Feedback

We evaluated three different conditions represented in the feedback display: (a) correctness of the individual decision, (b) correctness of the joint decision and (c) conflict of individual decisions. (a) No significant cluster was found representing differences between a correct and incorrect individual interval decision (all $p$-values > .07).

**FIGURE 5** The averaged ERPs of correct and incorrect responses over each participant at the button press indicating individual interval choice are shown. The red shaded area represents the time frame of the cluster representing a significant difference between correct and incorrect trials. For this cluster, a topography representing averaged $t$-values over the significant time frame with highlighted electrodes is shown. Warm colors (yellow-red) represent positive $t$-values, cool colors (blue) negative $t$-values [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 6** Average ERPs corresponding to conditions with different conflicting and different correctness situations of the joint interval choice are shown for Fz (middle) and CPz (bottom). Colored shaded areas show significant time frames of corresponding significant clusters. Red corresponds to the negative cluster representing differences between correct and incorrect trials. Yellow and blue correspond to the significant positive and negative clusters of conflicting situations. Green refers to the cluster testing the difference of the correctness contrast versus the conflict contrast and black represents the respective interaction. Topographies (top) are accordingly color-coded and represent averaged $t$-values over the cluster. Warm colors (yellow-red) represent positive $t$-values, cool colors (blue) negative $t$-values [Colour figure can be viewed at wileyonlinelibrary.com]
(b) Furthermore, no significant cluster representing differences between feedback about conflicting and non-conflicting joint decisions was found (all p-values > .143). (c) A significant negative cluster was found for differences between feedback conditions, signaling a correct versus incorrect joint decision in trials where the EEG participant made the joint decision (CM = −5.060; p = .004; CI = [0.002; 0.006]). This frontal-to-parietal cluster included samples in a time frame 449–883 ms after feedback onset. The highest negative parametric statistical value was found at channel FC1 at 645 ms Figure 7, t(19) = −6.52, p < .001, d = −1.46. This series of tests indicated that only the joint correctness condition elicited consistent differences in event-related potentials after feedback presentation.

Subsequently, we further investigated if this difference was modified by a conflict. We found a significant interaction 434–824 ms after feedback onset at frontal-to-central electrode sites, indicating higher differences between correct and incorrect joint interval decisions in non-conflicting conditions as compared to conflicting conditions (CM = 2,799.2; p < .001; CI = [0; 0.0001]). The highest parametric statistical value was found at channel FCz at 793 ms, Figure 7, t(19) = 6.55, p < .001, d = 1.46. In the same time frame, 434–824 ms, both conflicting correct responses, M = 0.67, SD = 0.58; t(19) = 2.293, p = .0167 d = 0.71, one-sided and conflicting incorrect responses, M = 0.66, SD = 0.53; t(19) = 2.73, p = .007, d = 0.73, one-sided elicited significantly larger ERPs than non-conflicting correct responses (M = 0.48, SD = 0.33) at FCz. Additionally, conflicting correct responses, t(19) = −2.68, p = .007, d = −0.21, one-sided and conflicting incorrect responses, t(19) = −3.64, p < .001, d = −0.26, one-sided elicited significantly smaller ERPs than non-conflicting incorrect responses (M = 0.933, SD = 0.57) at the same channel and time. These results indicate that conflicting correct and conflicting incorrect responses evoked a significantly lower potential than non-conflicting incorrect decisions as well as a significantly higher potential than non-conflicting correct decisions. Finally, we investigated whether differences in event-related potentials, resulting from the initially tested conditions, differed from each other. The permutation test of clusters representing significant differences between differences of joint versus individual correctness, joint correctness versus conflict conditions, and of individual correctness versus conflict conditions yielded no significance (all p values > .213).

In summary, these results show that feedback-related differences in EEG signals were elicited by incorrect versus correct responses.

**FIGURE 7** Average ERPs corresponding to conditions with different conflicting situations and the correctness of the joint interval choice are shown for FCz. The yellow shaded area represents the time frame of the significant cluster of the correctness contrast. The red shaded area represents the significant interaction. Topographies on the left are accordingly color-coded and represent positive (yellow/red) and negative (blue) t-values. Significant electrodes are highlighted [Colour figure can be viewed at wileyonlinelibrary.com]
In the present study, we investigated factors contributing to collective decision-making and neurophysiological correlates of collective decision-making by adapting a perceptual decision-making task. Similar to the non-verbal condition in Bahrami et al. (2012b), we found a trend toward collective benefit for dyads communicating via confidence rating scales. Moreover, we found that as the sensitivity similarity between co-actors increased so does the possibility for attaining collective benefit. Furthermore, performance improved as expressed confidence level increased, indicating higher metacognitive sensitivity. Evaluation of confidence rating scale usage revealed no significant relationship between the DDI, the difference in the frequency with which each confidence level was used between a dyad, and collective benefit. However, the CPDI, the difference in performance at each confidence level between a dyad, was found to be a predictive measure of collective benefit. Neither the optimal nor the WCS model predictions significantly fit the data, yet the WCS model produced a similar linear tendency on performance similarity. Dyad familiarity had no effect on dyad performance.

In addition to evaluating factors contributing to decision-making and collective benefit, we aimed to understand the neurophysiological underpinnings of such phenomena using EEG. In order to achieve a wide coverage of potential mechanisms in this paradigm, a pure explorative cluster analysis method was chosen. We have previously introduced several components related to such processes, but due to the novelty of EEG in this paradigm an analysis detached from specific hypotheses regarding the timing and waveforms of potential effects was implemented.

We examined electrophysiological potentials in three phases: stimulus presentation, response and feedback. The stimulus presentation phase contains information regarding the initial evaluation of contrast differences. Investigation of neurophysiological potentials here revealed modulation of EEG signals, thus indicating successful contrast manipulations, as potentials were linearly related to the range of contrasts presented in the paradigm. Secondly, we investigated individual and joint response phases where trials were evaluated based on conflicting information and correctness. We found that ERPs evoked after individual response were modulated by correctness. At the joint decision, event-related potentials were affected by both correctness and conflict, where conflict over correctness produced a larger significant difference at respective electrode sites. Furthermore, an interaction between correctness and conflict indicated differential cortical representations of correctness and conflict. Lastly, we investigated neurophysiological potentials for the feedback phase, where the correctness of both individual responses and the final joint response were presented. After feedback presentation, joint correctness evoked significant differences in event-related potentials, which was further modified by conflict. Moreover, existing social conflict dissipated feedback-related differences evoked by correctness. In the following, each of these results will be discussed in turn.

4.1 The dyad vs. the better participant

No significant difference was found when comparing the dyad to the sensitivity of better participants; therefore, pairs did not significantly achieve collective benefit. This is not surprising, as Bahrami et al. (2012b) did not find that dyads achieved a collective benefit in the non-verbal condition, which the present study was modeled after. Perhaps the current exchange and quality of information is not adequate for proper integration or is insufficient information to achieve benefit. It may be of interest to explore alternative means of non-verbal communication, such as providing cumulative performance scores at the end of each block as an added means to assess a partner's reliability.

As Bahrami et al. (2012a) and others (Wahn et al., 2017) have found, a combination of types of information exchanged may enhance collective benefit in a variety of tasks. Wahn et al. (2017) explored the effect of information type exchanged on collective benefit in a visuospatial task. They found that collective benefit could be achieved by providing either performance scores or information about a co-actor, but when provided with the combination of both information types, dyads achieved a higher performance earlier than in either condition alone. In the current paradigm, it is possible that the number of trials was an insufficient amount of time to observe collective benefit. With this in mind, the current exchange of information may not be enough to overcome individual differences and/or the limitations of non-verbal communication to achieve benefit. Perhaps providing a combination of additional information types to the non-verbal confidence rating scale would allow for dyads to achieve collective benefit within the allotted number of trials. Further experimentation would be needed in order to address either of these alternatives.

One could also potentially improve upon the confidence scale used in the current paradigm by implementing a more rigid structure. For example, each level of confidence could be assigned a more concrete classification, numerical or descriptive, therefore further aligning the dyad's use of the scale. Bahrami et al. (2012b) found that dyads communicating verbally, as compared to non-verbally or not at all, achieved the greatest benefit and fulfilled the predictions of the WCS model. This suggests that verbal communication is most advantageous when it comes to achieving collective benefit.
benefit. Under further investigation, Fusaroli et al. (2012) analyzed how dyads verbally adapted to each other and found that pairs who aligned linguistically to the task at hand locally, but more importantly, globally achieved greater benefit. As Borg (1982) proposed, the use of a scale with the addition of verbal anchors allows better for the possibility of inter-individual comparisons by creating a fixed foundation for all individuals to utilize. Therefore, adding descriptive factors to the scale may help dyads align linguistically on a more local and global level adding a fixed “verbal” concept to a non-verbal task.

Although dyads did not significantly achieve collective benefit, we found a significant correlation between the similarities of the perceptual sensitivities and the collective benefit. This finding suggests that dyads with similar perceptual sensitivities were more likely to achieve a collective benefit. Bahrami et al. (2010, 2012b) also observed this relationship between sensitivity and collective benefit suggesting perceptual sensitivity similarity as an important predictor of collective benefit. Furthermore, Wahn, Schmitz et al. (2016) and Wahn, Czeszumski, and König (2018) found this pattern to hold true for joint visuospatial and dynamic motor tasks in which no communication was allowed between dyads, suggesting this predictor as important in tasks other than perceptual decision-making. Additionally, Wahn, Czeszumski, and König (2018) found this phenomenon to also hold true for triads illustrating the global application that sensitivity differences may play in predicting collective benefits.

### 4.2 The confidence rating scale vs. performance and collective benefit

Overall, participants were well calibrated, expressing a positive metacognitive sensitivity, on an individual level, which was demonstrated by the positive relationship between performance and expressed confidence level. However, dyads did not significantly accrue collective benefit to an advantage. To understand why this happens, the dynamics of the confidence rating scale were further investigated. We first looked at the difference between a pair’s confidence distributions and found no significant relationship between confidence distribution difference and dyad sensitivity, further suggesting possible dissimilar expression of information between a dyad. In other words, knowing how often each confidence rating was used and how different this distribution was between co-actors was not sufficient information to predict collective benefit. This is reasonable as the distribution reveals nothing about performance at a particular confidence level.

Furthermore, if a pair expresses their answers similarly, this does not imply that the information was expressed accurately or interpreted similarly. While the most extreme values along a scale leave little room for inter-individual differences in interpretation, lack of concrete descriptors allows for the possibility of diverging personal interpretations of middle values. This may make expressing one’s degree of certainty or uncertainty more difficult and the distribution of these values will only tell us how often they are used. Therefore, it is appropriate to assess the performance differences between a pair at each confidence level, which we evaluated through the CPDI measure.

The CPDI compares the alignment of a dyad based on the performance of each dyad member at each confidence level. Here, a relationship was found between CPDI and collective benefit. Thus, the more similar the usage of the confidence rating scale between a dyad, the greater the ratio of sensitivity between dyad and best member, thereby increasing the possibility of achieving collective benefit. Pescetelli et al. (2016) similarly reported a positive correlation between the average metacognitive sensitivity of individuals in a pair and dyadic performance when reporting confidence through wagers. Our results may therefore solidify the effectiveness of the rating scale as an indicative measure of performance between dyads of high sensitivity differences. This coincides with what Fusaroli et al. (2012) found in that dyads more linguistically aligned to the task at hand achieved higher performance than those who were not. However in the current study, as was observed in the Bahrami et al. (2012b) study, with the given experimental conditions no collective benefit was achieved. This suggests that the internal conceptualization of the scale between a pair is not the same, where the concept of each confidence value is not similar enough to be advantageous. More specifically, how one co-actor internalizes the value of confidence level “2” may not be similar enough to how the other co-actor internalizes that same value.

### 4.3 WCS model results

To further assess the possibility of achieving collective benefit to an advantage, we compared our data to the predictions of previously established models. Our data significantly did not fulfill the WCS model as proposed by Bahrami et al. (2010). Instead, the model overestimated dyad performance, similarly to what was found in Bahrami et al. (2012b). Despite significant deviations, the model produced a similar linear trend as our data. As the WCS model predicts dyad performance given appropriate expression of information, it would appear that information in the current study was not expressed adequately in order to achieve the model’s predictions. This may be as a result of the way in which the confidence scale was understood between dyads.

### 4.4 Participant familiarity

Bahrami et al. (2010, 2012a, 2012b) employed participants who were familiar with each other. This parameter was not
restrictive in our study, and no relationship was found between familiarity and correctness. In other words, familiar dyads were no more accurate or successful than unfamiliar dyads. This lends to a real-world application of such communication measures, in that relative strangers can conceivably communicate effectively to achieve collective success. For example, such a measure could be applied across a company to achieve success among colleagues who are unfamiliar with each other or are from different departments, but need to collaborate on a task. On the other hand, Brennan and Enns (2015) found a relationship between familiarity and visibility of a partner when performing a joint visual search task. Familiar pairs were reported to outperform unfamiliar pairs when visible to each other, but when separated by a partition, familiar and unfamiliar pairs performed similarly. It is possible this effect was present in the current paradigm as dyads were separated by a partition. Further experimentation is required to investigate whether familiar dyads would have outperformed unfamiliar dyads had they been visible to each other also for the present task.

4.5 | Stimulus presentation

We first sought to understand basic sensory, attentional and cognitive processes at hand during stimulus presentation when investigating neurophysiological aspects of joint decision-making. Detection of the odd Gabor patch builds the fundamental information needed for this joint task through accumulation of evidence, ultimately reaching a decision, which is expressed via button press. The analysis of EEG signals at stimulus presentation revealed a significant linear relationship for contrast differences in the target stimulus starting with a widespread central positive and a left-temporal negative deflection from 350 ms after stimulus onset. This coincides well with the architecture of the attentional system, which can be conceptualized with control regions in frontoparietal areas that generate modulatory signals and sensory processing areas in the occipitotemporal cortex where these modulatory signals influence ongoing visual processing (Scolari, Seidl-Rathkopf, & Kastner, 2015). At such temporal and spatial locations, diverse potentials related to typical P300 components are found. These are related to processes of stimulus discrimination. Hence, our finding could be related to processes typically connected with P300 components, which are generated when the neuronal representation of the stimulus environment is changed or updated (Polich, 2007).

Midline frontal-to-parietal P300 components are typical for this kind of oddball task, where these components are generated when perceptual stimulus discrimination occurs (Polich, 2007). P300 components are often subdivided into the P3a and P3b components, with the classical P3b component typically found after a target stimulus in oddball paradigms. It is described as having a parietal distribution on the scalp and has been linked to context updating, context closure and event categorization (Bledowski et al., 2004; Donchin & Coles, 1988; Kok, 2001; Polich, 2007; Verleger, 1988). In contrast, the predominantly frontal P3a component has been linked to an earlier stage of information gathering rather than memory and context updating (Berti, 2016). Yet, we cannot directly term our finding as a classical P300 component as the positive deflection is rather small with an overall negative polarity based on a comparison with the baseline. At the same time, temporal and spatial properties are consistent with an interpretation of our findings as a P300 component. Overall, we provide evidence for processes related to stimulus discrimination or attentional processes. However, our design was not suited to disentangle attentional processes from other potential explanations of the effect, such as basic sensory or cognitive processes that also could be at play in this evaluation. We found a clear signal that is correlated with stimulus discriminability, and future research could directly manipulate this phase to further evaluate the nature of this signal.

4.6 | Individual and joint responses

In the explorative analysis of the response phases, we addressed both the time before and after response for potential EEG responses based on correctness and conflict. The cluster analysis of differences at both individual and joint responses revealed no significant differences before a response was made, indicating no differences in action preparation based on conflict situations or correctness of the respective interval choice. However, significant differences occurring directly after response indicate differential processing of conflict and correctness of responses, suggesting a social influence based on information exchange. Both conflicting situations and incorrectness elicited a late slow positive potential at posterior electrode sites. This slow potential resembles the error positivity (Pe) typically found after incorrect responses.

A differential encoding is suggested by a significant positive cluster at frontal electrode sites for non-conflicting conditions. This is further supported by a significantly higher difference between conflicting conditions as compared to correctness conditions. The interaction indicates more pronounced differences for the error positivity, typically found at parietal electrode sites. A conflict situation does not strictly represent an error, but in contrast to the correctness of the response, one can assume that participants were aware of a mismatch. Considering our results to be related to the typical error positivity, these results comply with previous findings (Orr & Carrasco, 2011; Steinhauser & Yeung, 2010) that suggest a relationship between the slow evoked Pe potential and conscious awareness of an error, while further suggesting Pe as a function of evidence accumulation that an error has been
made. The presence of a conflict is prominent information which should be taken into account as evidence in an error-evaluation process.

Early components of the cluster indicating differences based on correctness and conflict after response may be associated with typical ERN components. Yet, ERN components are typically found early after a participant’s response has been given, and thus, a baseline correction shortly after the response could diminish effects. Using a baseline before the participant’s response, however, would entangle potentials based on motor action with potential cognitive processes of error monitoring. Hence, we used the 0–100 ms post-response baseline for our analysis.

With regard to our results of this analysis, the found early deflections could be associated with error-monitoring processes to a degree, as we found differences at early components only for conflicting versus non-conflicting responses. These findings support the notion of a conflict-monitoring function of this early evaluation phase, which is typically represented within the ERN component (Botvinick et al., 2001). Nevertheless, Boldt and Yeung (2015) found only the Pe was predictive of graded changes in confidence, while the ERN failed to predict variations of confidence in single trials, suggesting an all-or-none signal. Evidence of correct-related negativity proposes the ERN component to include response elicited emotion and comparison (Vidal, Hasbroucq, Grapperon, & Bonnet, 2000). This may implicate social context as an important factor in decision-making. The finding of an earlier deflection for response conflict, but not for correctness conditions, could be due to the fact that information that an error was made was not readily available, in contrast to the information that a conflict is present. In other studies, participants are immediately aware of an error, such as in flanker tasks where the correct response is readily available. However, in the current study the response is made after stimulus offset. Hence, we comply with theories reporting that the ERN relates to response conflict and not with theories which support it as pure error representation.

In fact, we found no error representation (correctness contrast), and this might be due to the fact that there is no clear representation of an error or feedback about an error. Di Gregorio, Maier, and Steinhauser (2018) suggest that representation of a correct response is necessary to elicit the ERN component. While we show presence of this ERN, a clear representation of this response may not have been formed. This is supported by the division of confidences reported, in that participants on average reported their confidence in the lowest confidence level, 53.28% of the time. Overall, this suggests participants were generally uncertain in their reported decision and therefore this may be associated with an unclear representation of a correct response. Nevertheless, we see a Pe for both the correctness contrast as well as the conflict contrast, even though the information indicating the correctness of response was not available. While error-monitoring functions of the ERN and Pe are well established in literature, we provide evidence for an additional differential processing of social conflict encoded within EEG signals in this paradigm.

4.7 Feedback presentation

Our analysis of the feedback phase revealed a significant negative cluster, which was observed in the time window of 450–900 ms following feedback presentation. This component may represent the late positive potential (LPP). Typically, this potential is distributed over frontal-to-parietal central electrode sites (Cunningham, Espinet, DeYoung, & Zelazo, 2005; Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Naumann, Bartussek, Diedrich, & Lauffer, 1992; Wiens & Syrjänen, 2013), which aligns well with the location of the negative cluster found in the current study. Bradley (2009) reported that the LPP is elicited by emotional stimuli, which is interpreted to be related to enhanced attention and visual processing. Our results indicate a stronger response for conditions where the joint interval decision made by the EEG participant was incorrect, which can accordingly be interpreted as more emotionally arousing given the increased personal culpability in this decision. A significant interaction indicating higher differences for correct versus incorrect conditions when there was no conflict is particularly interesting, indicating a differential error encoding for different conflicting situations. However, these results could also be accounted for by interpreting conflicting situations as situations with higher perceptual uncertainty. The current design is not suitable to disentangle those two interpretations, meaning that both mechanisms could be at play, thus indicating a need for further investigation concerning decision conflict situations.

5 Conclusion

The current study sought to investigate the neurophysiological correlates of joint action and collective decision-making by combining an already-established perceptual decision-making paradigm with EEG. Similar to Bahrami et al. (2012b), we found that dyads were overall unable to achieve collective benefit to an advantage. While neither the WCS nor the optimal models provided a significant fit for observed behavior, the WCS model predicted the same dependence on individual performance similarity with higher performance offset. Through investigating the relationship between Confidence Performance Difference Index (CPDI) and collective benefit, we sought to better understand the confidence rating scale as a means of communication. Our results indeed suggest that the CPDI could be a predictive measure of collective benefit.
In addition, our results for the neurophysiological data suggest the presence of attentional, error-evaluation or social conflict-monitoring processes in each investigated phase in the current paradigm. In particular, in the stimulus presentation phase, a contextual working memory update is possibly represented by the linear relationship of contrast differences. When a joint interval decision is indicated by a button press, a social conflict due to divergent individual decisions may be represented by a frontal response negative potential in the response phase. Subsequently, incorrect decisions and social conflict may be represented in a late positive potential, which could be associated with the late error positivity. Similarly, in the feedback phase a late positive potential elicited after the presentation of feedback by incorrect and conflicting situations is indicative of emotional arousal and an allocation of attentional resources.

In summary, our results suggest that response- and feedback-related components and potentials elicited by an error-monitoring system may differentially integrate socially conflicting information exchanged during the joint decision-making process. The current study serves as an initial exploratory step toward understanding the underlying mechanisms involved in dyadic interaction and exchange of information. Further research could facilitate understanding of the suggested underlying mechanisms and solidify effective communication techniques for dyads of diverging sensitivities.

ACKNOWLEDGEMENTS

We would like to thank Paola Suarez-Ramirez and Raul Sulaimanov for their contributions to data collection.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

DATA AVAILABILITY STATEMENT

Data for this study are available on https://osf.io/bsw4t/

AUTHOR CONTRIBUTIONS

Study Design: PK, BW, KGB, PB, A-NS, and AS. Data Acquisition: KGB, PB, A-NS, and AS. Experimental Task Programming: KGB, PB, and A-NS. Behavioral Data Analysis: KGB, PB, and A-NS. EEG Data Analysis: AS. Literature Research: KGB, PB, A-NS, and AS. Manuscript Writing: KGB, PB, A-NS, and AS. Manuscript Revisions: KGB, PB, A-NS, AS, PK, and BW.

ORCID

Kristina G. Baumgart https://orcid.org/0000-0002-3174-581X

REFERENCES

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012a). Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making. Journal of Experimental Psychology: Human Perception and Performance, 38(1), 3–8. https://doi.org/10.1037/a0025708
Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012b). What failure in collective decision-making tells us about metacognition. Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 1350–1365. https://doi.org/10.1098/rstb.2011.0420
Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. Science, 329, 1081–1085. https://doi.org/10.1126/science.1185718
Bang, D., Aitchison, L., Moran, R., Castanon, S. H., Rafiee, B., Mahmoodi, A., … Summerfield, C. (2017). Confidence matching in group decision-making. Nature Human Behaviour, 1, 0117. https://doi.org/10.1038/s41562-017-0117
Bang, D., Fussaroli, R., Tylen, K., Olsen, K., Latham, P. E., Lau, J. Y. F., … Bahrami, B. (2014). Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision-making. Consciousness and Cognition, 26, 13–23. https://doi.org/10.1016/j.concog.2014.02.002
Bartholow, B. D. (2010). On the role of conflict and control in social cognition: Event-related brain potential investigations. Psychophysiology, 47(2), 201–212.
Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. Psychophysiology, 42, 33–42. https://doi.org/10.1111/j.1469-8986.2005.00258.x
Bauer, F., Cheadle, S. W., Parton, A., Müller, H. J., & Usher, M. (2008). Gamma flicker triggers attentional selection without awareness. Proceedings of the National Academy of Sciences, 105(5), 1666–1671.
Berti, S. (2016). Switching attention within working memory is reflected in the P3a component of the human event-related brain potential. Frontiers in Human Neuroscience, 9, 701. https://doi.org/10.3389/fnhum.2015.00701
Bledowski, C., Prvulovic, D., Hochtetter, K., Scherg, M., Wirbal, M., Goebel, R., & Linden, D. E. J. (2004). Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. Journal of Neuroscience, 24(42), 9353–9360. https://doi.org/10.1523/JNEUROSCI.1897-04.2004
Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. Journal of Neuroscience, 35(8), 3478–3484.
Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. Medicine and Science in Sports and Exercise, 14(5), 377–381.
Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624
Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology, 46*(1), 1–11. https://doi.org/10.1111/j.1469-8986.2008.00702.x

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*, 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Brennan, A. A., & Enns, J. T. (2015). What’s in a friendship? Partner visibility supports cognitive collaboration between friends. *PLoS ONE, 10*(11), e0143469. https://doi.org/10.1371/journal.pone.0143469

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging, 18*(1), 32–42. https://doi.org/10.1109/42.750253

Cunningham, W. A., Espinet, S. D., DeYoung, C. G., & Zelazo, P. D. (2005). Attitudes to the right-and-left: Frontal ERP asymmetries associated with stimulus valence and processing goals. *NeuroImage, 28*(4), 827–834. https://doi.org/10.1016/j.neuroimage.2005.04.044

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52*(2), 95–111. https://doi.org/10.1016/S0301-0511(99)00044-7

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *NeuroImage, 172*, 427–436. https://doi.org/10.1016/j.neuroimage.2018.01.081

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

Eimer, M., & Mazza, V. (2005). Electrophysiological correlates of change detection. *Psychophysiology, 42*(3), 328–342. https://doi.org/10.1111/j.1469-8986.2005.00285.x

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. *Psychophysiological Brain Research, 1*, 192–195.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science, 23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Ganesh, T., Takagi, A., Osu, R., Yoshioka, T., Kawato, M., & Burdet, E. (2014). Two is better than one: Physical interactions improve motor performance in humans. *Scientific Reports, 4*, 3824. https://doi.org/10.1038/srep03824

Hackley, S. A., & Valle-Inclán, F. (1999). Accessory stimulus effects on response selection: Does arousal speed decision making? *Journal of Cognitive Neuroscience, 11*(3), 321–329. https://doi.org/10.1162/089892999563427

Heerken, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*, 467–479. https://doi.org/10.1038/nrn2374

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163–178.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91–100.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience, 23*, 315–341. doi.org/10.1146/annurev.neuro.23.1.315

Koivisto, M., & Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience and Biobehavioral Reviews, 34*, 922–934. https://doi.org/10.1016/j.neubiorev.2009.12.002

Koivisto, M., Salminen-Vaparanta, N., Grassini, S., & Revonsuo, A. (2016). Subjective visual awareness emerges prior to P3. *European Journal of Neuroscience, 43*, 1601–1611. https://doi.org/10.1111/ejn.13264

Kok, A. (2001). On the utility of the P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557–577. https://doi.org/10.1093/soej/soj048

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv. European Journal of Physiology, 284*(1), 1–17.

Kourtis, D., Sebanc, N., & Knoblich, G. (2013). Predictive representation of other people’s actions in joint action: An EEG study. *Social Neuroscience, 8*(1), 31–42. https://doi.org/10.1080/17479919.2012.694823

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*(2), 605–620. https://doi.org/10.1006/0749-5978(02)00003-1

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology, 90*(4), 644–651. https://doi.org/10.1037/0022-3514.90.4.644

Li, P., Han, C., Lei, Y., Holroyd, C. B., & Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing: An ERP study. *Psychophysiology, 48*, 1129–1133.

Mahmoodi, A., Bang, D., Olsen, K., Zhao, Y. A., Shi, Z., Brobek, K., … Bahrami, B. (2015). Equality bias impairs collaborative decision-making across cultures. *Proceedings of the National Academy of Sciences of the United States of America, 112*(12), 3835–3840. https://doi.org/10.1073/pnas.1421692112

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jnemeth.2007.03.024

Masamoto, J., & Inui, N. (2013). Two heads are better than one: Both complementary and synchronous strategies facilitate joint action. *Journal of Neurophysiology, 109*, 1307–1314. https://doi.org/10.1152/jn.00776.2012

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S., Hairston, W. D., Ferris, D. P., & König, P. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data? *Journal of Neuroscience Methods, 284*(1), 1–17.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task. *European Journal of Neuroscience, 9*, 150. https://doi.org/10.1111/j.1469-8986.1997.00010.x
Baumgart KG, Byvshev P, Ulrich R, Schiltz K, Driesen J, Byvshev P. (2020). Multisensory teamwork: Using a tactile or an auditory display to augment joint attention. *European Journal of Neurology*, 27(8), 1564–1572. https://doi.org/10.1177/1351027820911753

Baumgart KG, Byvshev P, Ulrich R, Schiltz K, Driesen J, Byvshev P. (2020). Multisensory teamwork: Using a tactile or an auditory display to augment joint attention. *European Journal of Neurology*, 27(8), 1564–1572. https://doi.org/10.1177/1351027820911753

Baumgart KG, Byvshev P, Ulrich R, Schiltz K, Driesen J, Byvshev P. (2020). Multisensory teamwork: Using a tactile or an auditory display to augment joint attention. *European Journal of Neurology*, 27(8), 1564–1572. https://doi.org/10.1177/1351027820911753

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.