

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

Systematic Review Search Strategy A systematic literature search of PubMed and PsycInfo using the search terms emotional conflict task (all fields) OR conflict resolution OR conflict regulation OR conflict adaptation AND NOT review (all fields) AND NOT mice OR mouse OR rat OR rats. Throughout the review process, all preferred reporting items for systematic reviews and meta-analyses guidelines were followed (Moher et al., 2015; Page et al., 2021). In total, $n = 3,923$ articles were identified through the database searches, of which 75 duplicate articles were removed prior to screening. All articles were compiled and screened using Rayyan software (Ouzzani et al., 2016).

RELIABILITY OF EMOTIONAL CONFLICT RESOLUTION

Study Eligibility Articles were included in this review if they (a) were journal articles, book chapters or theses, (b) were published in English, (c) used the emotional conflict task, and (d) included adult human participants in at least one experiment. Articles were excluded in this review if they (a) did not have English full text, (b) were review articles, or (c) reanalyzed data from previous publications. If articles included a conflict resolution analysis, they were deemed eligible for the meta-analysis. In total, $n = 3,848$ articles underwent title/ abstract screening.

Data Extraction Data screening was performed by two independent raters (Jiani Yin and Yi Wang). All articles were collated and screened by both raters for inclusion and exclusion criteria during initial title/ abstract screening, followed by full-text screening for potentially includable articles. A total of $n = 3,486$ articles were excluded based on the title/abstract screening. Full-text articles were either deemed includable for systematic review and meta-analysis (completed a conflict resolution analysis) or only for systematic review (no conflict resolution or adaptation analysis completed). Of the $n = 362$ articles sought for full-text screening, two could not be retrieved, which left 360 articles that underwent full-text screening. After full-text screening, $n = 94$ articles were determined to be eligible for inclusion in the systematic review. Yi Wang extracted effect sizes from studies eligible for meta-analysis. It was possible to include $n = 26$ studies in the meta-analysis of reaction time and $n = 18$ studies in the meta-analysis of accuracy, both of which included results from the current study. Effect sizes were calculated by taking the reported effect size or F or t test statistics from the conflict resolution analysis and converted to Pearson's r using the effect size conversion spreadsheet in Lakens (2013).

Restricted maximum likelihood functions were used, and the analysis was illustrated with forest plots. The heterogeneity of the included effect sizes was calculated using I^2 , where low, medium, and high I^2 values are classified as 25%, 50%, and 75%, respectively. Hedge's Q was also used as a method for testing effect size heterogeneity (Hedges, 1981, 1982a, 1982b). Funnel plot symmetry and publication bias were assessed using Egger's test (Egger et al., 1997). Funnel plots are displayed in the online supplemental materials. Corrections for publication bias were performed using the trim and fill method (Duval & Tweedie, 2000). Further, separate meta-analyses were conducted comparing clinical versus healthy participants using the methods described above. Moderation analyses were also performed, with moderating factors including whether studies reported controlling for repetition priming (Hommel et al., 2004; U. Mayr et al., 2003), contingency learning (Schmidt et al., 2007), and completion of either face or word-responses tasks.

Experimental Study Participants One hundred sixteen participants above the age of 17 years were recruited via the Queensland University of Technology online sign-up portal (SONA system) and word of mouth. The sample size was determined based on GPower calculations using a priori, repeated measures F test, with a power of 0.8, one group, eight measurements, and an assumed medium effect size of 0.4 to 0.5, using the recommended optional setting of Cohen (1988) effect size $F(V)$. This calculation yielded a recommended sample size of 85. Ethical approval was received from the University Human Research Ethics Committee (UHREC: 3625). All participants provided informed consent prior to study commencement. Participants were asked to provide information about their age, gender, and ethnicity in free-text response options. Participants who performed below 70% accuracy on the condition level were excluded.

The Emotional Conflict Task Meta-Analysis Meta-analysis was performed using the metafor in R (Viechtbauer, 2010). Our analysis tested the emotional conflict resolution and adaptation effects, pooled across studies that had

tested for these effects using the emotional conflict task. The emotional conflict resolution effect in this type of study is where faster reaction times are observed on incongruent trials that were preceded by an incongruent trial (il) compared to incongruent trials that were preceded by a congruent trial (cl). Conflict adaptation was determined using the following formula: $(cl - cC) - (il - iC)$. Where means and standard deviations were provided, we computed the relevant effect size using these statistics. Where only statistical tests were reported, effect sizes were calculated using the Lakens (2013) spreadsheet. If neither of these options were available, PlotDigitizer was used to extract the means and standard deviations from illustrations in the manuscript (Huwaldt & Steinhorst, 2015). In cases where the conflict resolution effect was reported as not significant but no other information was included, the effect size was included as 0. This is an approach that we have used in previous meta-analyses that accounts for underreporting of nonsignificant effects and is recommended by Cochrane guidelines as a suitable approach for meta-analyses (Higgins et al., 2021; Ney, O'Donohue, et al., 2022).

The images of faces were sourced from a standard face database containing facial expressions of young, middle-aged, and older Caucasian women and men (Ebner et al., 2010). A total of 32 young Caucasian faces were used: eight happy male faces, eight angry male faces, eight happy female faces, and eight angry female faces. The images were edited using Paintshop Pro such that an emotional word (happiness-related or anger-related) was superimposed on each image. The affective valence of the faces and words can be either congruent or incongruent (see Figure 1). There was a total of 16 emotional words: eight "happy" words (happy, jolly, joyful, pleased, blissful, cheerful, buoyant, and gleeful) and eight "angry" words (angry, irate, fuming, raging, furious, outrage, hostile, and enraged) which were chosen from a combination of two sources: a word database (Warriner et al., 2013) and an online dictionary. The words were presented in the middle of the image in red, size 28-point font. Words of the same affect were grouped into pairs such that the two words matched in letter length and word-use frequency (e.g., fuming and raging). The screen background was set to white. The task consisted of a battery of 64 unique stimuli that differed in face-word congruency (incongruent or congruent) and gender (male and female) (16 stimuli per condition). During the experiment, each participant completed two blocks of stimuli, where they were

YIN, WANG, LIPP, MAYO, AND NEY

Figure 1 Schematic of Trials During the Experiment

Note. The illustrated trials demonstrate a cl trial order. Participants had to respond using S or L keys as quickly and as accurately as possible based on either the facial emotion or the emotion of the word. Trials either consistent of congruent or incongruent face-word pairings. cl = congruent-incongruent. From "FACES—A Database of Facial Expressions in Young, Middle-Aged, and Older Men and Women: Development and Validation," by N.C. Ebner, M. Riediger, and U. Lindenberger, 2010, *Behavior Research Methods*, 42, 351–362 (<https://doi.org/10.3758/BRM.42.1.351>). Copyright 2010 by the Max Planck Institute for Human Development.

required to respond to either the faces or the words. In the face block, participants were instructed to respond according to the emotion expressed by the face while ignoring the word. In the word block, the distractor was the facial expression, and participants were instructed to indicate which emotion was written on the face. Each image was presented twice in both blocks, yielding a total of 256 trials (128 trials per block). These blocks were counterbalanced and corresponding instructions were presented on the screen at the beginning of each block. There were four types of stimulus pairings throughout the experiment that were used to capture the emotional conflict resolution and adaptation effects: trials where the previous trial was congruent and the current trial was also congruent (cC), trials where the previous trial was congruent but the current trial was incongruent (cl), trials where the previous trial was incongruent and the current trial was congruent (iC), and trials where the previous trial was incongruent and the current trial was also incongruent (il). Participants were asked to respond to either the emotion of the words or the face via a keyboard press (S or L) as accurately and quickly as possible. The keys that correspond to either happy or angry (i.e., S or L) were also counterbalanced across participants to account for handedness effects. Before each trial, a fixation cross was shown in the middle of the screen for 500 ms. Each trial was presented until a keyboard response was made or for a maximum of 3 s. If the participants did not provide a response within 3,000 ms, the program presented the next trial. A blank screen was presented for 1,000 ms in between trials (see Figure 1). Once the first

categorization task was completed, the instruction screen for the second categorization task was presented. The structure of the second categorization task was similar to the first.

Procedure The experiment was implemented in Inquisit, which allows for remote testing. The experiment could be accessed and completed on any electronic device (e.g., laptop, phone, and iPads). As the data collection was conducted remotely, viewing distance and other environmental and contextual factors (i.e., quiet room, independent setting, and sitting position) were not controlled for. Participants downloaded Inquisit on their respective device prior to the commencement of the study. At the start of the experiment, participants were presented the participant information sheet containing the project details. After participants had provided consent to take part in the experiment (ticked check box), they were redirected to the experimental task. After completing all the trials, participants completed the demographic questionnaire (age, gender, sex, ethnicity, and type of device used). The experiment was designed to be completed in roughly 30 min.

Statistical Analyses Since this dataset is part of a larger study, full results from this analysis will be published elsewhere. Data were screened for outliers, with responses faster than 150 ms and slower than an individual

RELIABILITY OF EMOTIONAL CONFLICT RESOLUTION

participant's $M + 2$ SDs removed from the analyses. Participants who had extremely low accuracy (<70% accuracy) were excluded from the study analyses. Further, reaction times from trials that followed an error trial were excluded from the analysis of conflict resolution and adaptation effects. For each of the word and face categorization tasks, two 2 (congruency: congruent, incongruent) $\times$ 2 (poser expression: happy, angry) within-subject analyses of variance (ANOVAs) were separately used to determine the overall influence of congruency on performance of the emotional conflict task for reaction times and accuracy. Next, to examine the conflict resolution effect, further four (congruency of previous compared to current trial: cC, cl, iC, il) factor repeated measures ANOVAs were conducted comparing the mean reaction time of cC, cl, iC, and il trials separately for word-response and face-response tasks. The same repeated measures ANOVAs were conducted for accuracy conflict resolution effects. Conflict adaptation and posttrial conflict slowing effects were quantified using the formulae $(cl - cC) - (il - iC)$ and $(il + iC)/2 - (cl + cC)/2$, respectively. All analyses were conducted in SPSS V27.

Transparency and Openness Data from our experimental study are uploaded to the Open Science Framework website (Yin et al., 2023). The study and metaanalysis were not preregistered, though the data and code for the meta-analysis are also available from the Open Science Framework website (<https://osf.io/ndkxp/>).