COGNITION AND EMOTION, 2017
https://doi.org/10.1080/02699931.2017.1386622

Mixed matters: fluency impacts trust ratings when faces range on valence but not on motivational implications

Michal Olszanowski[a], Olga Katarzyna Kaminska[a] and Piotr Winkielman[a,b,c]

[a]SWPS University of Social Sciences and Humanities, Warsaw, Poland; [b]Department of Psychology, University of California, San Diego, USA; [c]Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK

ABSTRACT
Facial features that resemble emotional expressions influence key social evaluations, including trust. Here, we present four experiments testing how the impact of such expressive features is qualified by their processing difficulty. We show that faces with mixed expressive features are relatively devalued, and faces with pure expressive features are relatively valued. This is especially true when participants first engage in a categorisation task that makes processing of mixed expressions difficult and pure expressions easy. Critically, we also demonstrate that the impact of categorisation fluency depends on the specific nature of the expressive features. When faces vary on valence (i.e. sad to happy), trust judgments increase with their positivity, but also depend on fluency. When faces vary on social motivation (i.e. angry to sad), trust judgments increase with their approachability, but remain impervious to disfluency. This suggests that people intelligently use fluency to make judgments on valence-relevant judgment dimensions – but not when faces can be judged using other relevant criteria, such as motivation. Overall, the findings highlight that key social impressions (like trust) are flexibly constructed from inputs related to stimulus features and processing experience.

ARTICLE HISTORY
Received 6 May 2016
Revised 22 September 2017
Accepted 26 September 2017

KEYWORDS
Fluency; trust; face evaluation; facial expressions; social judgment

Social judgments are influenced by facial features that resemble emotional expressions, but also facial structure and gender (Ambady & Weisbuch, 2011; Brownlow & Zebrowitz, 1990; Dotsch, Hassin, & Todorov, 2016; Knutson, 1996; Montepare & Dobish, 2003; Neth & Martinez, 2009; Stirrat & Perrett, 2010; Zebrowitz, Kikuchi, & Fellous, 2010). Among such judgments are attractiveness, dominance, and importantly, trust – a key element in relationships, trade, politics, and governance (Rezlescu, Duchaine, Olivola, & Chater, 2012; Wojciszke, Bazinska, & Jaworski, 1998). Data-driven analyses show that faces high on a dimension termed "valence" (which resemble smiles) are judged trustworthy and approachable, and this dimension accounts for over 60% of the variance in first impressions (Oosterhof & Todorov, 2008). Another important dimension of facial features relates to social dominance and explains about 18% of the variance. Faces high on this dimension resemble anger and are judged as untrustworthy and threatening – basically inducing avoidance motivation (Oosterhof & Todorov, 2008, p. 11090). This combination of features (related to both valence and social motivation) makes impressions of trust from faces not only important, but also informative for investigating basic mechanisms that shape social judgments in general.

Research on the mechanisms of social judgments established that the impact of target features on evaluations is qualified by their processing fluency, defined as the ease of perceptual and conceptual mental operations (Jacoby, Kelley, & Dywan, 1989; for a recent review, Schwarz, 2015). Fluency typically enhances evaluation – an observation initially formulated in the "hedonic fluency hypothesis" (Winkielman, Schwarz, Fazendeiro, & Reber, 2003) and subsequently confirmed in a variety of studies using ratings, actions, psychophysics, and physiology (e.g. Alter & Oppenheimer, 2006; Carr, Brady, & Winkielman, 2017; Carr, Rotteveel, & Winkielman, 2016; Lee & Labroo, 2004; Topolinski,

CONTACT Michal Olszanowski molszanowski@swps.edu.pl

Likowski, Weyers, & Strack, 2009; Winkielman & Cacioppo, 2001; Winkielman, Halberstadt, Fazendeiro, & Catty, 2006). In the domain of faces, several studies found that increasing perceptual and conceptual processing fluency of neutral faces boosts ratings of attractiveness and trust (Halberstadt & Winkielman, 2014; Owen, Halberstadt, Carr, & Winkielman, 2016; Winkielman, Olszanowski, & Gola, 2015, Study 4).

The role of fluency becomes especially important when considering expressive features of faces. In the real world, faces often contain features of multiple emotions (Aviezer et al., 2008; Russell, 1997; Sebe et al., 2007). This may lead perceivers to experience perceptual and cognitive disfluency in their attempt to process the mixed face, especially when perceivers need to identify its emotion. Consistent with this idea, we found that mixed happy-angry expressions elicit lower trust ratings, even when compared to purely angry faces (Winkielman et al., 2015). This devaluation of mixed expressions was enhanced in a condition that made such mixed faces categorically disfluent, as when the participants first needed to assign a face to its emotional category (i.e. happy vs. angry). This finding is important as it shows that evaluations depend not only on objective features associated with the stimulus (including its objective mix), but also on the perceiver's current goals or mindset. As such, fluency and its evaluative consequences derive from the interaction between stimulus features and the perceiver's task. Correspondingly, in the emotion categorisation condition, the disfluency of categorisation statistically mediated trust judgments (Winkielman et al., 2015).

The present research aims to better understand the specific mechanisms by which mixed facial expressions influence social judgments, such as trust. We examine how disfluency in categorising facial expressions interacts with specific features of facial expressions related to dimensions of valence and motivation. This is theoretically important, as it addresses what kind of features influence key social judgments, what types of conflicts between features matter, and how this process is modified by the perceiver's goals. Before we detail our hypotheses, it is useful to speculate on the potential mechanisms linking expressive features and fluency with the valence and motivational aspects of facial expressions.

Faces, features, and fluency: three possibilities

As discussed earlier, facial features underlying social judgments (such as trust) vary not only on a dimension related to valence but also on a dimension related to social motivation. Also, note that specific emotional facial expressions can have different valence and motivational implications. For example, expressions of happiness typically signal positive valence and motivational approach. Consistently, seeing smiles promotes cooperation (van Kleef, de Dreu, & Manstead, 2010) and approach behaviour (Winkielman, Berridge, & Wilbarger, 2005). Expressions of sadness indicate negative valence but also low dominance, which elicit approach-like reactions of sympathy and compassion (Clark & Taraban, 1991; Eisenberg, 2000; Van Kleef et al., 2008). Expressions of anger indicate negative valence, but also, as mentioned, high dominance. In many contexts, such faces elicit avoidance reactions and fear in observers (Adams, Ambady, Macrae, & Kleck, 2006; Dimberg & Ohman, 1996; Horstmann, 2003; Marsh, Ambady, & Kleck, 2005; but see Wilkowski & Meier, 2010). Thus, the above analysis suggests that (i) a mix of happiness and anger varies on both valence and motivation, (ii) a mix of happiness and sadness varies more on valence, and (iii) a mix of sadness and anger varies more on motivation. Examining such mixes gives insight into the interaction between valence and motivational dimensions in key social judgments, expanding our understanding of how expression-related facial features contribute to trust and other evaluative judgments.

Our main question pertains to how valence and motivational aspects of facial expressions interact with processing fluency in determining trust judgments. There are three distinct theoretical possibilities. The first possibility is that disfluency leads to a broad devaluation effect for mixed expressions. Basically, disfluency (of any kind) elicits negative affect, which is then used to inform any evaluation of any associated stimuli. This possibility predicts that any mixture of faces (e.g. happy-angry, happy-sad, or angry-sad) will reduce trust evaluation, assuming the mixture causes disfluency (see Figure 1, left panel).

The second possibility is that mixed expressions only elicit disfluency and consequent devaluation when they are mixed on motivational implications. Theoretically, this is consistent with classic (Schneirla, 1959) and recent (Bach et al., 2014) models that highlight the importance of motivational conflicts related to approach-avoidance. This prediction fits with earlier findings that disfluency in categorising happy-angry mixes lowers trust ratings (Winkielman et al., 2015). If so, we should not see fluency effects on happy-sad blends (which vary on valence), but we

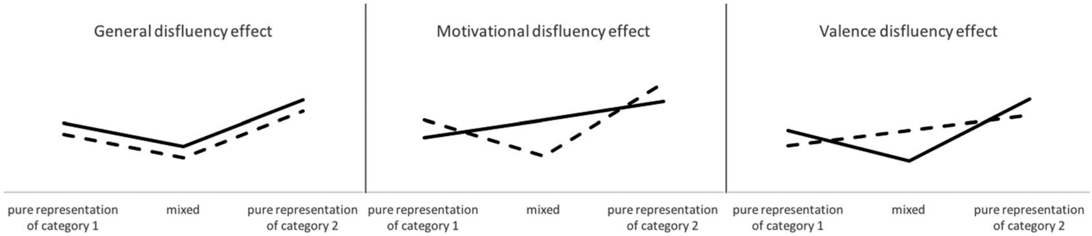

Figure 1. Hypothetical patterns of evaluations representing different theoretical predictions tested by current experiments: 1 (left panel) – any mixed expressions are devalued; 2 (middle panel) – only expressions mixed on valence are devalued; 3 (right panel) – only expressions mixed on motivation are devalued. Solid lines represent stimuli that vary on valence dimension (e.g. angry and happy or sad and happy faces as "pure representations", and their blends as "mixed"). Dashed lines represent stimuli that vary on motivational dimension (e.g. sad and angry faces as "pure representations", and their blends as "mixed").

should observe fluency effects on angry-sad blends, which vary on motivation (see Figure 1, middle panel). It is also worth noting that the same prediction about angry-sad blends can be derived from models that highlight their differences in arousal (Calder, Burton, Miller, Young, & Akamatsu, 2001).

However, most theoretical and empirical considerations support the third possibility, which assumes that valence matters most. Three separate lines of research suggest that valence conflicts are particularly effective in reducing evaluation. First, attitude research shows that mixed valence states are particularly aversive (Nohlen, van Harreveld, Rotteveel, Lelieveld, & Crone, 2014). Second, the literature on "valence-focus" highlights the primacy of this dimension in evaluation (Barrett, 2006). Third, literature on subjective experiences shows that participants' use of any experience (including fluency) reflects its applicability for the judgment at-hand, along with availability of other relevant information (for theory, see Schwarz, 2010; for empirical examples see Cho & Schwarz, 2008; and Study 3 in Carr, Hofree, Sheldon, Saygin, & Winkielman, 2017). Recall that fluency, in-and-of-itself, generates changes in valence. Such changes should matter most when judging other stimuli that vary on valence (e.g. happy-sad blends). By the same token, fluency-related hedonic changes should be less relevant when targets vary primarily on motivation (e.g. angry-sad blends). This is because in the latter case, participants see that there are no informative valence differences present in the stimuli (both emotions have negative valence) and instead should derive their trust judgments from visible differences in motivational implications of the faces. If so, we should observe no fluency effects on angry-sad mixes, but robust fluency effects on happy-sad mixes, for which the fluency-induced valence differences are relevant (see Figure 1; right panel). We will address alternative theoretical perspectives in the general discussion.

Current studies

To assess the relative role of mixed features on valence-related and motivation-related dimensions, the current studies used facial expressions varying in valence or in motivation. In Experiments 1 and 2, we used blends of happiness and sadness, which vary more on valence than motivation. In Experiment 3, we used blends of anger and sadness, which vary more on motivation than valence. In Experiment 4, we combine happy-sad and angry-sad blends. To summarise the predictions, trust judgments should be influenced by expressive features related to valence (Experiments 1, 2 and 4) as well as by features related to motivation (Experiment 3 and 4), replicating Oosterhof and Todorov (2008). Specifically, happiness should elicit more trust than sadness (i.e. valence effect) and sadness should elicit more trust than anger (i.e. motivation effect). In terms of fluency, the predictions depend on the particular account. If disfluency has a generic negative effect, then decreases in trust should occur regardless of the type of expressions that are mixed (e.g. happy-sad or angry-sad), assuming the task generates disfluency. In other words, participants should decrease trust for any "mixed" (disfluent) expressions and increase trust for "pure" (fluent) expressions, especially when participants are first asked to assign expressions to respective categories. If the effect requires motivational conflict, it should *only* occur for angry-sad blends. However, if the effect requires valence conflict, it should *only* occur for happy-sad blends. As discussed, most theoretical and empirical evidence lines up with the latter prediction.

Experiment 1

This study examined participants' trust judgments from male and female expressions, varying from happiness to sadness (thus primarily differing in valence). Participants first categorised a face and then evaluated the face's trustworthiness. As in previous studies, categorical fluency was manipulated by the nature of a classification task. In the control condition, participants classified the face's gender. This should ensure equally fluent processing for all faces, including those expressing blends of anger and happiness. In the experimental condition, participants classified the face's emotion. This should make processing of mixed-emotion faces selectively disfluent. As discussed, if such faces vary on valence, this should lead to their devaluation on subsequent trust judgments. Note here that in both conditions, participants performed a classification task, thus controlling for possible non-specific effects due to the very act of categorising the face.

We also wanted to assess whether our categorisation manipulation changes the perceived intelligibility of mixed-expressions, since inscrutable faces might be less trusted. To assess this, we also collected judgments of readability of targets' intentions. Specifically, we asked, "How clear (readable) are the person's intentions?" Note that readability judgments should be higher for happiness. Happiness (compared to anger or sadness) is less confused with other emotions (Elfenbein & Ambady, 2003), better recognised under many visual conditions (Smith & Schyns, 2009), and more socially expected (Ekman & Oster, 1979). In short, the more smiling a person exhibits, the more readable he/she should be judged. To ensure that faces were appropriately classified, we also collected error data during the categorisation task (Beale & Keil, 1995).

Method

Participants

Seventy-eight undergraduates from the University of California, San Diego (UCSD) participated for course credit. We did not track each participant's gender and age but the mean age of the participant population is about 21 years, SD = 5 years, and is 70% female.

Stimuli

Stimuli were Caucasian faces from the WSEFEP set containing 30 individuals (Olszanowski et al., 2015). From this set, we selected 18 individuals (nine females and nine males) who provided two expressions each (sadness and happiness). Faces were selected on expression quality (i.e. not because of specific trust or attractiveness level, and were not standardised to limit natural variations on those traits). Mixed expression faces were constructed using FantaMorph 5 software by combining (in different proportions) two source images of "pure" expressions. Transitions within 13 steps of each expressions pair were delineated using over 100 facial landmarks. The 50/50 midpoint was discarded to avoid using expressions that could not be objectively categorised as predominantly happy or sad, resulting in 14 pictures from each poser (see Figure 2 for examples).

Procedure

Participants were randomly assigned to an experimental group. On each of the 54 trials, participants saw a fixation cross for one second, followed by a face for three seconds. The face picture was then followed by a categorisation question which remained on the screen until the participant's response. There was no time limit for response and categorisation time was measured from the relative onset of the question. The task was manipulated between-subjects, and thus in the expression categorisation condition, participants classified each face (using the "z" and "/" keys) as either happy or sad. In the gender categorisation condition, participants classified each face as either male or female. Participants needed to categorise quickly, as the trial automatically advanced to the judgment phase. Next, for each face, participants responded to two questions in sequence on a 9-point scale: (1) "How trustworthy is that person?" (1 = *not at all*, 9 = *extremely*), and (2) "How clear (readable) are the person's intentions?" (1 = *not clear*, 9 = *clear*).

Data preprocessing and analysis strategy

Fluency was measured by categorisation speed. To reduce the impact of outliers and individual differences, we first computed each participant's log-transformed reaction times (RTs) on the categorisation task, and then averaged those log-RTs for each stimulus type. Data were analyzed with a 2 (categorisation: expression vs. gender) × 14 (emotion: sadness to happiness) repeated-measures ANOVA. On our hypothesis, in the expression categorisation condition, the fluency (RT) curve should assume an inverse U-shape, while the evaluative (trust) curve should follow a U-shape. Accordingly, we tested for the presence of quadratic contrasts

Figure 2. Examples of stimuli used in current experiments. Top and middle rows show sad-happy stimuli used in Experiments 1, 2 & 4, representing pure expressions of sadness on the far left side (frame no1) and happiness on the far right side (frame no15), and mixed expressions with the dominance of sadness (frame no 6) and dominance of happiness (frame no9). The bottom row showing analogical frames of angry-sad stimuli (Experiments 3 & 4), with anger on the far left and sadness on the far right. Images courtesy of Olszanowski et al. (2015) by permission.

and their interaction with classification condition. To directly test our hypothesis that processing fluency contributes to differences in evaluations for targets with "pure" and "mixed" expression, we explored fluency as a mediator. Our mediation tests employed the multilevel approach, which provided unbiased estimates of indirect and total effects (Bauer, Preacher, & Gil, 2006). To verify that the results reported below were robust, we also analyzed the critical RT and trust ratings using multilevel modelling (MLM), including random intercepts for participants and stimuli properties (e.g. to control for different trust levels due to facial features of portrayed individuals). The results of the MLM analyses, which replicate the results reported below, appear as an appendix (see Appendix A).

Results

Fluency

Figure 3a shows that categorisation of "mixed" expressions took longer than "pure" expressions, but only in the expression categorisation condition. Specifically, in the expression categorisation condition, there was a squared effect of emotion type (inverse U-shape), $F(1, 39) = 23.41$; $p < .001$; $\eta^2 = .37$. In the gender categorisation condition, emotion type had no impact on RTs. Overall, this resulted in a interaction between emotion type and categorisation condition, $F(1, 76) = 8.93$; $p = .003$; $\eta^2 = .10$. Post-hoc power analysis of above mentioned interaction, with G-Power software (version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that we achieved over 80% power, using a two-tailed test at $\alpha = .05$ and lower-bound estimate of non-sphericity correction at $\varepsilon = .077$. Data on categorisation accuracy also showed that aside from the middle frames, which presented highly mixed emotions (i.e. morph levels 6, 7, and 9), subjects classify around 80% of expressions as emotions that actually were dominant on the face (Figure 3c).

Trust judgments

Figure 3b shows trust judgments. Trust evaluations depended on facial expression, with happier faces trusted more, $F(1, 76) = 19.6$; $p < .001$; $\eta^2 = .21$. This linear trend was significant in the control group that categorised gender, $F(1, 37) = 8.88$; $p = .005$; $\eta^2 = .19$,

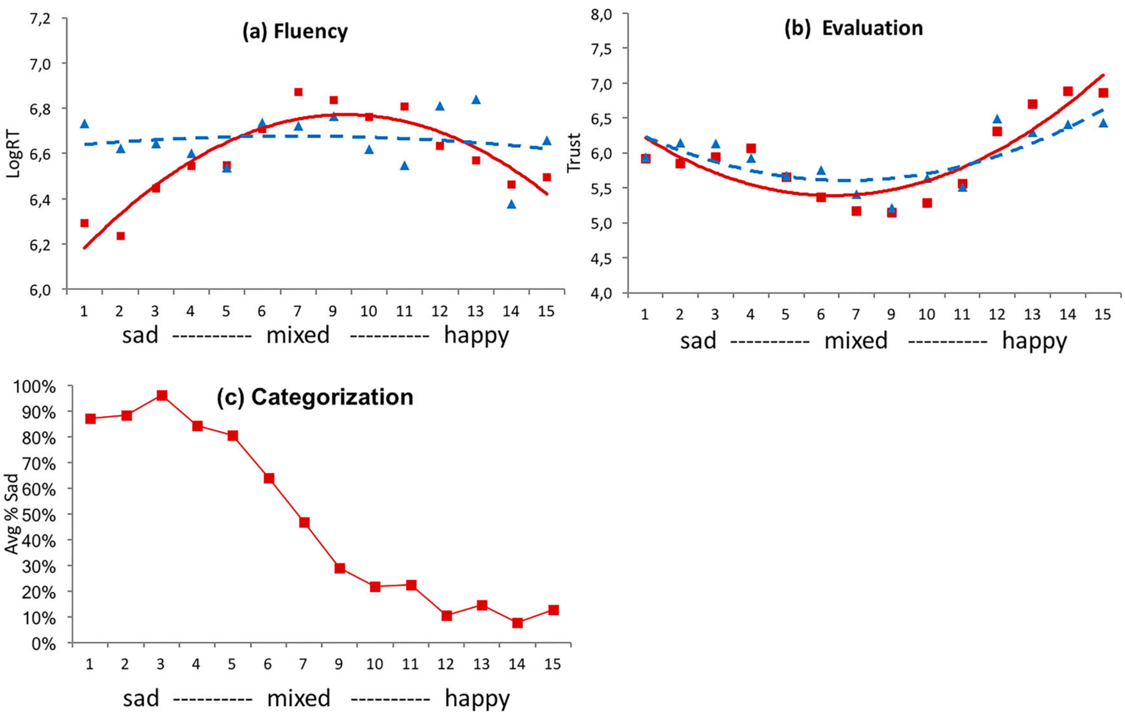

Figure 3. Experiment 1 results for (a) fluency, (b) evaluation (trust), and (c) categorisation. In (a) and (b) responses are shown as a function of target features (sadness and happiness) and (expression as red squares, gender as blue triangles), while lines represent an estimated trend. In (c) only expression categorisation condition is presented, and points show the average percentage of sadness categorisations.

and in the experimental group that categorised expression, $F(1, 39) = 11.06$; $p = .002$; $\eta^2 = .22$. We also found that "pure" expressions (happiness and sadness) were generally trusted more than "mixed" expressions, as reflected in overall quadratic effect, $F(1, 76) = 43.18$; $p < .001$; $\eta^2 = .36$. This quadratic effect occurred in the control (gender) condition, $F(1, 37) = 8.67$; $p = .006$; $\eta^2 = .19$. This effect was even stronger in the expression condition, $F(1, 39) = 40.39$; $p < .001$; $\eta^2 = .51$. This resulted in the significant quadratic interaction with categorisation $F(1, 76) = 5.78$; $p = .019$; $\eta^2 = .08$; achieved interaction power was over 70% ($\alpha = .05$ and $\varepsilon = .077$). As shown in the Figure 3b, the expression categorisation condition particularly enhanced rating for "pure" happy faces.

Readability judgments

We found a different pattern of results for readability judgments. As happiness in the faces increased, they were rated as more readable, as reflected in a linear effect, $F(1, 76) = 77.76$; $p < .001$; $\eta^2 = .51$. There was also a significant quadratic effect, showing a general tendency for higher readability rating for "pure" faces, $F(1, 76) = 18.33$; $p < .001$; $\eta^2 = 0.19$. There was no significant quadratic interaction with the categorisation task.

Mediation

We also examined whether fluency mediates the relationship between expression "mixed-ness" and evaluation within the emotion categorisation condition. Our tests for mediation employed the multilevel approach recommended by Bauer et al. (2006). Our predictor was expression "mixed-ness" (ranging from 1 [pure sadness and pure happiness] to 7 [maximally blended]). Our dependent measure was evaluation of trust, and our mediator was categorisation RT (fluency).

First, we regressed fluency onto "mixed-ness" (*a*-path), which showed that blended faces were less fluent than pure faces, $B = .0838$, $SE = .017$, $p < .001$. Next, we regressed evaluations onto fluency (*b*-path), which showed that less fluent targets received more negative evaluations than more fluent targets: $B = -.3797$, $SE = .079$, $p < .001$. When controlling for the effect of fluency, the path from "mixed-ness" to evaluative judgments was significant, but reduced in magnitude, $B = -.1807$, $SE = .031$, $p < .001$ (c' path), compared to $B = -.2125$ (c-path). In fact, a test for indirect effects for lower level mediation models (Bauer et al., 2006) indicated a statistically significant indirect effect of fluency ($M = -.027$, $SE = .011$, 95% CI [−.0508, −0.0041]). This confirms that lower evaluations of mixed faces and higher evaluations of pure faces occurred partly because of differences in processing fluency.

Discussion

Experiment 1 examined the interplay between features of facial expressions (ranging from happiness to sadness) and categorisation fluency on trust judgments. Consistent with previous work emphasising the importance of the valence dimension, we observed that increasing the happiness percentage within expressions leads to higher trust, consistent with the Oosterhof and Todorov (2008) model. This is interesting in terms of the model because it shows that negative features that impact the trust evaluation do not have to come from anger. Critically, we also found that trust evaluations for pure faces were relatively increased by fluency, whereas evaluations of mixed happy-sad faces appear reduced by categorisation disfluency. This finding fits with our previous work that found such relative (as compared to pure faces) devaluation effects on mixed angry-happy expressions (Winkielman et al., 2015), and it also suggests that disfluency effects can occur even when mixed faces do not imply conflict on a motivational direction (e.g. withdrawal in the case of anger and approach in the case of joy). In short, it supports a model by which fluency influences trust evaluations by modifying the impression of valence.

Experiment 2

This experiment aimed to replicate results from Experiment 1 with a modified procedure and a different subject population. It also addressed an important theoretical issue for understanding the mechanisms by which fluency influences social evaluation. Specifically, on most theoretical models, the valence effects from conceptual fluency manipulations occur in late stages of processing, and thus influence judgments during the process of integration with other pieces of information (Schwarz, 2015). Therefore, conceptual fluency should not influence evaluations of early, more perceptual aspects (such as how positive or negative the target looks). However, it is theoretically possible that fluency directly changes how pleasant the expression appears, via top-down influences on perception (for tests and discussion of such influences, see Carr, Brady, et al., 2017; Winkielman, Huber, Kavanagh, & Schwarz, 2012). To gauge how people evaluate the valence of the faces, we included an additional scale to assess participants' overall ratings of positivity versus negativity. In addition, we slightly changed the rating procedure. In Experiment 2, participants recorded their answers with mouse, by clicking a button (while categorising stimuli) or moving a slider (while evaluating stimuli). Moreover, instead of the USA, Experiment 2 was performed on a different population consisting of Polish students. This is useful, as the influence of the valence dimension on trust judgments may vary by culture (Cuddy et al., 2009).

Method

Participants

Thirty undergraduates (20 females; we did not track each participant's age but the population from which participants were drawn has a mean age of 20 years, SD = 3 years) at SWPS University of Social Sciences & Humanities in Poland participated for course-credit.

Procedure

Given the effect size achieved in Experiment 1, this smaller sample size is reasonable (see below), but we acknowledge its limitations. Note, however, that to have sufficient statistical power, we increased the number of trials to 84. Otherwise, the procedure and materials were the same as in Experiment 1, except that participants responded with a mouse. In the categorisation task, participants indicated the appropriate category by clicking on a virtual button located approximately 50 pixels above and below the centre of computer screen (while the face was centrally located). Next, for each face, participants responded to three questions by moving a 100-point slider (the change was strictly technical, as it allowed for more flexible mouse movement). The first question was "Is this person trustworthy?" (with anchors from *no* to *yes*). The second question was "This person appears … " (with anchors from *negative* to *positive*). The third question was, again, "How clear (readable) are the person's intentions?" (with anchors from *not clear* to *clear*).

Data preprocessing and analysis strategy

As before, fluency was measured by categorisation speed, and data were analyzed with a 2 (categorisation: expression vs. gender [between-subjects design]) × 14 (emotion: sad to happiness [within-subjects design]) repeated-measures ANOVA, which also tested for the presence of quadratic (U-shape) contrasts.

Results

Fluency

As expected, fluency of mixed expressions depended on the categorisation condition (see Figure 4a). In the emotion condition, categorisation of mixed expressions took longer than pure expressions. This quadratic (inverse U-shape) contrast was significant, $F(1, 14) = 22.66$; $p < .001$; $\eta^2 = .62$. In the control (gender) categorisation condition, valence had no impact on RTs. This pattern yielded a quadratic interaction on emotion type with categorisation condition, $F(1, 28) = 8.58$; $p = .007$; $\eta^2 = .24$. Additionally, we assessed the accuracy of categorisation (see Figure 4c). Subjects classified more than 80% of expressions as emotions that actually were more dominant on the face, except the highly mixed middle frames (i.e. morph levels 7, 8, and 9). Despite the relatively small sample, post-hoc power analysis of Experiment 2 indicated that we achieved more than 81% power for the interaction effect ($\alpha = .05$ and $\varepsilon = .077$).

Trust judgments

As in Experiment 1, we found that trust ratings closely followed the valence of the facial expressions, with happier faces trusted more, as reflected in the significant overall linear trend, $F(1, 28) = 41.97$, $p < .001$, $\eta^2 = .60$. "Pure" expressions (i.e. happiness and sadness) were generally trusted more than "mixed" expressions, as reflected in overall quadratic effect, $F(1, 28) = 24.47$, $p < .001$, $\eta^2 = .47$.[1] This quadratic effect was significant in the gender categorisation condition, $F(1, 14) = 6.57$, $p = .023$, $\eta^2 = .32$, and also in the expression categorisation condition $F(1, 14) = 17.94$, $p < .001$, $\eta^2 = .56$. Critically, there was also a significant interaction between emotion type and categorisation condition, with the quadratic effect stronger for the expression categorisation participants, $F(1, 28) = 6.2$, $p = .019$, $\eta^2 = .18$. This effect seems most noticeable in the relatively higher rating of pure sad faces within the expression categorisation condition (see Figure 4b). In short, the result on trust ratings replicated key aspects of Experiment 1, despite using a different population and rating setup. Post-hoc power analysis revealed 67% power for the interaction effect ($\alpha = .05$ and $\varepsilon = .077$).

Positivity and readability ratings

We did not observe any effects of $F(1, 28) = 52.55$; $p < .001$; $\eta^2 = .65$. This suggests that expression positivity ratings basically track the perceptual information in the face. On readability judgments, "pure" faces were rated as more readable than "mixed" faces, as reflected in a quadratic effect, $F(1, 28) = 22.59$; $p < .001$; $\eta^2 = .45$. Unexpectedly, there was significant quadratic interaction of emotion type and categorisation condition, $F(1, 28) = 5.68$; $p = .024$; $\eta^2 = .17$. The quadratic effect was stronger in the expression categorisation condition, $F(1,14) = 17.41$; $p < .001$; $\eta^2 = .55$ than the gender categorisation condition, $F(1,14) = 5.26$; $p = .038$; $\eta^2 = .17$. The observed power of the interaction effect was 63% ($\alpha = .05$ and $\varepsilon = .077$). However, note that this interaction was not observed in Experiment 1, which was run in the USA and with a larger sample size. It was also absent in our previous studies mixing happy and angry faces (Winkielman et al., 2015). This finding may reflect the importance of cultural differences in interpreting the social meaning of expressions (see Matsumoto, Yoo & Fontaine, 2008). However, since this effect is not robust and not central to the theoretical questions addressed here, we will not discuss it further.

Discussion

Despite changes in the rating procedure and participant population, Experiment 2 confirmed the main findings from Experiment 1. Fluency of categorisation increased trust evaluations, whereas disfluency decreased trust evaluations, for expressions ranging from happiness to sadness. Note that these faces varied primarily on valence, but not social motivation. Importantly, this experiment also showed that fluency-based effects on evaluation do not work via changing judgments about the perceptual appearance of the stimuli, since participants rated more smiling faces as more positive regardless of the categorisation condition. This is consistent with models suggesting that valence effects from conceptual fluency arise and enter judgments in late stages of processing, during of integration cognitive experience

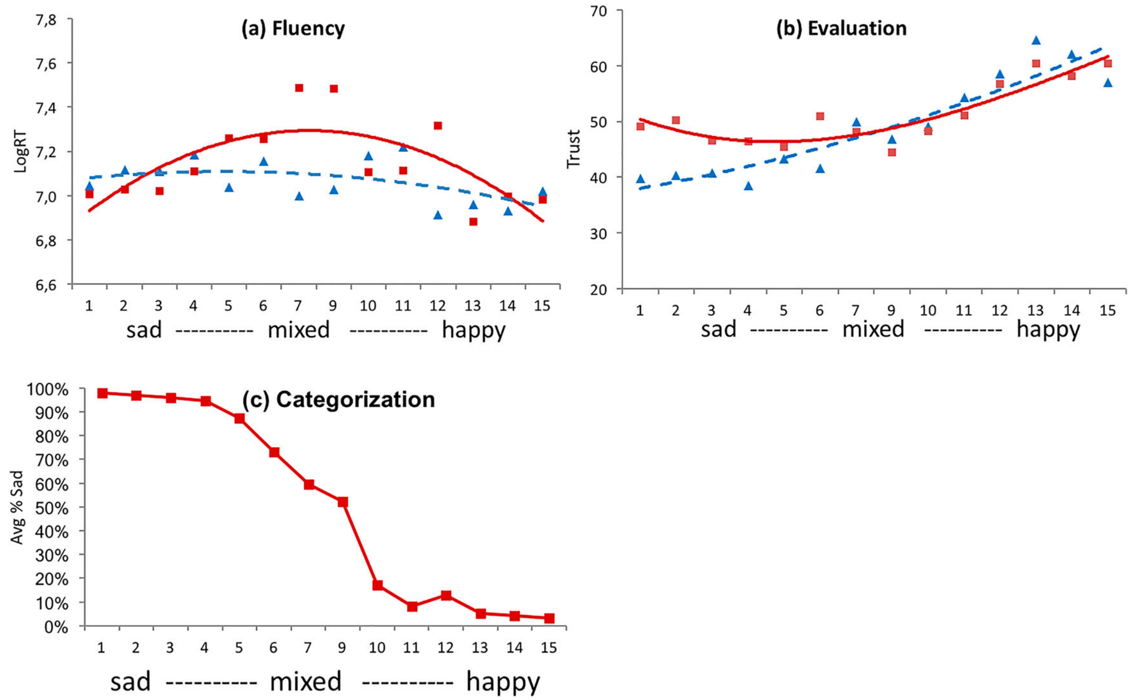

Figure 4. Experiment 2 results for (a) fluency, (b) evaluation (trust), and (c) categorisation. For (a) and (b), responses are shown as a function of target features (sadness and happiness) and categorisation condition (expression as red squares, gender as blue triangles), while lines represent an estimated trend. In (c), only expression categorisation condition is presented, and points show the average percentage of responses for sadness.

with other relevant pieces of information (Schwarz, 2015). The finding that fluency selectively impacted different judgments also argues against the idea that fluency functions as "generic inducer of affect," which influences all evaluations.

Experiment 3

Experiments 1 and 2 showed that fluency impacts trust ratings for faces that vary on valence. However, as discussed in the Introduction, there are several possible interpretations of the current data, along with earlier findings on "happy-angry" mixes that vary on both valence and motivation (Winkielman et al., 2015). First, these findings could indicate that fluency impacts trust judgments when the target contains valence-related features, and thus, the hedonic effects of fluency can be integrated with other information (or be the sole basis of judgment, if there are no relevant features). Second, the findings could reflect a more general, non-specific effect where fluency changes ratings of an associated stimulus. In Experiment 3, we tested these alternative predictions by showing participants faces that ranged from sadness to anger. Recall that such faces do not differ on valence but instead on motivational direction (approach-withdrawal; submission-dominance). Again, on the "generic negative affect" model, trust judgments of such faces should be influenced by disfluency. However, on the model where trust judgments are based on relevant information in the stimulus, participants should be primarily using differences in social motivation implied in the faces, with fluency having minimal impact. The aim of Experiment 3 was to preliminarily explore these predicted relations, and (similar to Experiment 2) we used a smaller sample size of Polish participants.

Participants

Twenty-four undergraduates (15 females; we did not track each participant's age but the population from which participants has a mean age of 20 years, SD = 3 years) from the SWPS University of Social Sciences & Humanities participated for course-credit (13 participants completed the expression categorisation condition).

Procedure

The sad and angry faces came from the WSEFEP set (Olszanowski et al., 2015). The procedure was based on Experiment 2 (which had 84 within-subject trials) but was simplified, in that we asked only for trustworthiness judgments ("Is this person trustworthy?") matched on 100-point slider scale. Again, we acknowledge this smaller sample size, but it is comparable to Experiment 2.

Results

Fluency

Like our previous experiments, our fluency manipulation was successful (see Figure 5a). We observed an interaction of quadratic contrast on emotion type (i.e. angry to sad) with categorisation condition, $F(1, 22) = 8.58$; $p = .007$; $\eta^2 = .21$, with achieved power 68% ($\alpha = .05$ and $\varepsilon = .077$). In the emotion condition, categorisation of mixed expressions took longer than pure expressions (i.e. quadratic, inverse U-shape contrast was significant), $F(1, 12) = 7.26$; $p = .013$; $\eta^2 = .38$), while in the gender categorisation condition, emotion type had no impact on RTs. As in Experiment 2, participants were correct in categorising facial displays, except the middle frames presenting mixed emotions.

Trust judgments

As predicted, sadder faces were trusted more than angrier faces (see Figure 5b). This linear trend of expression type on judgment occurred in both categorisation groups, $F(1, 24) = 32.51$; $p < .001$; $\eta^2 = .60$. Importantly, this linear effect was significant even when calculated separately for each categorisation group (gender categorisation condition: $F(1, 10) = 11.41$; $p = .007$; $\eta^2 = .53$; emotion categorisation condition: $F(1, 12) = 22.69$; $p < .001$; $\eta^2 = .65$). Critically, in contrast to previous experiments, there were no significant quadratic effects, including no interaction of ratings with categorisation. We found no effects on trust for pure or mixed expressions, even when these mixed expressions created disfluency in the categorisation task. Importantly, the achieved power for the linear trend was over 99% ($\alpha = .05$ and $\varepsilon = .077$).

Discussion

Experiment 3 found that categorising expressions that are mixed on motivational implications generates disfluency, just like it did in Experiments 1 and 2 and in our previous research. However, despite this cognitive difficulty, participants in the experimental condition did not rate mixed faces as particularly untrustworthy (or pure faces as particularly trustworthy). It is worth noting that this occurred even though participants were sensitive to information present in the faces, and their judgments followed the well-known pattern of trust impressions (i.e. more negative ratings for angry faces). It is also worth noting that ratings were below the midpoint on the 100-point scale, which again suggests that participants were sensitive to face valence information and rated them as rather negative. Finally, note that the categorisation accuracy data suggest that participants clearly saw a distinction between sad and angry faces. Thus, we can exclude a potential concern for this study that participants based their judgments solely on features of one particular expression (e.g. anger). Nevertheless, a limitation of this study is its small sample size, which prevented detection of a possible quadratic effects. We address this and other issues in the next experiment.

Experiment 4

Experiment 4 was designed as an extended exploration and replication of the pattern observed in Experiment 3, with greater statistical power and addressing several theoretical issues. First, so far, we have separately compared the effects of happy-sad blends (Experiments 1 and 2) or sad-angry blends (Experiment 3), assuming alongside previous literature that the happy-sad blends vary more on valence and the sad-angry blends vary more on motivation. However, recent research suggests that the motivational implications of an expression such as sadness or anger switch from avoidance to approach depending on the context in which it appears (Paulus & Wentura, 2015). This could challenge the interpretation of our results. Thus, in Experiment 4, two sets of stimuli (i.e. mixes of sad-happy and angry-sad) were used within the same experimental block, so that we could verify the previously observed differences in such a context-controlled design. Furthermore, we addressed a possible concern that focusing participants on gender in the control condition diverts them from processing emotion-relevant information (Schyns & Oliva, 1999). In Experiment 4, as a control condition, we asked participants to point to the middle of the presented face. This ensured attention to faces without

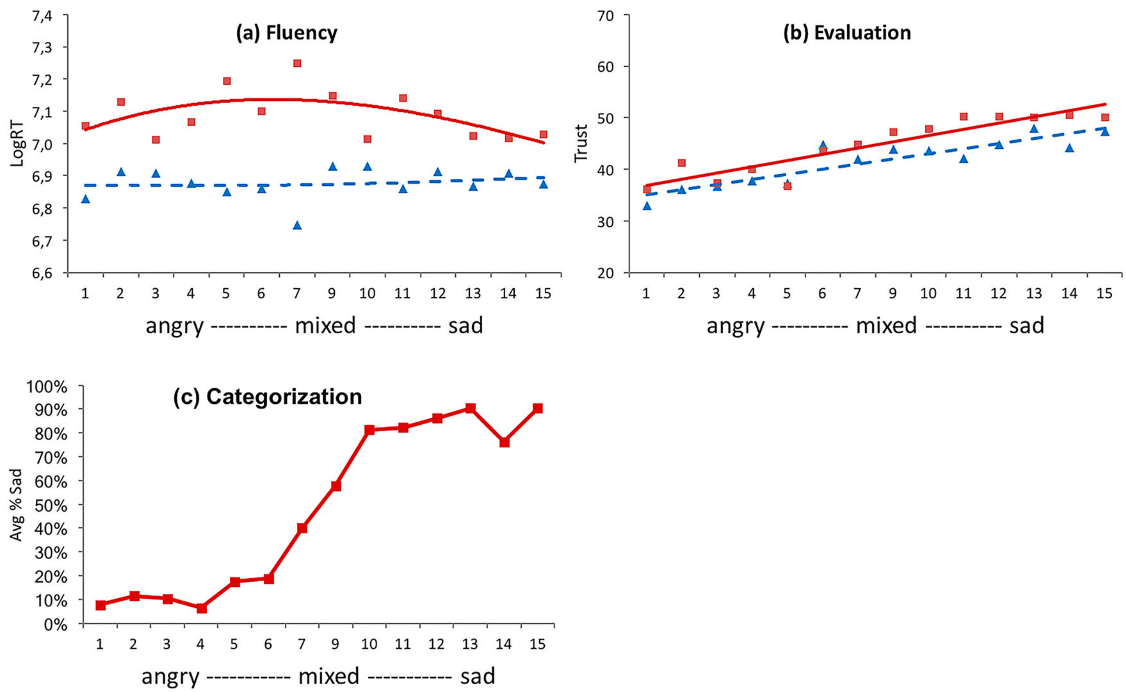

Figure 5. Experiment 3 results for (a) fluency, (b) evaluation (trust), and (c) categorisation. For (a) and (b), responses are shown as a function of target features (sadness and frowning) and categorisation condition (expression as red squares, gender as blue triangles), while lines represent an estimated trend. In (c), only expression categorisation condition is presented, and points show the average percentage of responses for sadness.

the need for binary categorisation. Finally, after the main task in Experiment 4, we directly asked participants for their ratings of valence and motivational implications of each expressions. This was done to directly verify our assumptions about the dimensions that vary across sad-happy and angry-sad blends.

Participants

Two-hundred forty-eight undergraduates from the University of California, San Diego (UCSD) participated for course-credit. We did not track each participant's gender and age but the population from which participants were drawn is 70% female and has a mean age of 21 years, SD = 5 years. Data for fifteen participants were excluded, as they did not complete all three parts of the study.

Procedure

The main parameters of experimental presentation were the same as Experiments 2 and 3, but we used 42 trials in the main (first) part, to keep the overall length of the experiment reasonable. The categorisation condition was between-subjects: participants either (*i*) categorised faces based on the displayed emotion (angry-sad or sad-happy) or (*ii*) clicked the middle of the face. After that, participants assessed the trustworthiness of the person by answering the question "Is this person trustworthy?" on a 100-point slider with extreme points labelled *no* and *yes*. Following the main part, participants performed two additional blocks (also 42 trials in each, in counterbalanced order). One block asked them to judge the person's affect ("Is the person negative or positive?" ranging on 100-point scale from *negative* to *positive*). The second block assessed the motivational dimension ("Would you approach or avoid that person?" with extreme values *avoid* and *approach*). None of these judgments were preceded by any categorisation.

Data analysis strategy

Data were analyzed separately for each type of emotional blend (angry-sad and sad-happy) with a 2-factor model consisting of 2 (task: categorisation vs control [between-subjects]) × 14 (mix-level [within-subject]) design. Additionally, we used a 3-factor model consisting of 2 (task: categorisation vs. control [between-subjects]) × 2 (type of emotional mix: angry to sad vs. sad to happy [within-subjects]) × 14 (emotion/mix-level: angry to sad vs. sad to happy [within-subjects design]) repeated-measures ANOVA, in order to test if the effects on trust rating for angry/sad and sad/happy are significantly different from one another.

Results

Fluency

As before, mixed emotions were difficult to process, but only in the emotion categorisation condition (see Figure 6, top panels). This yielded a quadratic interaction on RTs between the blend level and categorisation condition. Critically, this effect was robust for both types of expression blends, $F_{sad\text{-}happy}(1, 231) = 21.96$; $p < .001$; $\eta^2 = .09$ and $F_{angry\text{-}sad}(1, 231) = 19.88$; $p < .001$; $\eta^2 = .08$, with over 99% power for both types of blends ($\alpha = .05$ and $\varepsilon = .077$). There was no 3-way interaction. That is, emotion-categorisation created comparable disfluency for mixed faces along the angry-sad as well as the sad-happy continuums.

Trust judgments

First, we found a significant 3-way interaction between blend level, blend type, and condition, $F(13, 219) = 2.31$, $p = .012$, $\eta^2 = .01$, with over 99% power ($\alpha = .05$ and $\varepsilon = .077$), which confirms the difference in response patterns between angry-sad and sad happy mixes.

Separate analyses for each type of emotional mix showed that participants' judgments were sensitive to the expression-related information. That is, happy faces were trusted most, but sad faces were also trusted more than angry faces (see Figure 6, middle panels). The linear trend of expression type on judgment occurred in both categorisation groups for both types of mixes, $F_{sad\text{-}happy}(1, 231) = 232.13$; $p < .001$; $\eta^2 = .5$, and $F_{angry\text{-}sad}(1, 231) = 119.22$; $p < .001$; $\eta^2 = .34$. Additionally, the cubic trend was observed for sad-happy mixes, $F_{sad\text{-}happy}(1, 231) = 69.3$; $p < .001$; $\eta^2 = .23$.

The key analyses compared ratings for different types of mixes by categorisation condition. Most importantly, for sad-happy mixes in the emotion categorisation group, we found a quadratic trend, $F_{sad\text{-}happy}(1, 111) = 20.13$; $p < .001$; $\eta^2 = .15$. "Pure" sad and happy expressions were relatively more valued than mixed expressions when participants categorised on emotion (see Figure 6). In the control group, only the linear trend was significant, $F_{sad\text{-}happy}(1, 120) = 139.26$; $p < .001$; $\eta^2 = .54$. This pattern produced a quadratic interaction, $F_{sad\text{-}happy}(1, 231) = 7.72$; $p = .006$; $\eta^2 = .03$, replicating findings from Experiments 1 and 2. Post-hoc analysis of achieved power for the interaction effect was 79% ($\alpha = .05$ and $\varepsilon = .077$).

For angry-sad mixes, the linear trend was significant in the control group, $F_{angry\text{-}sad}(1, 120) = 47.66$; $p < .001$; $\eta^2 = .28$, and also in the emotion categorisation group, $F_{angry\text{-}sad}(1,111) = 73.91$; $p < .001$; $\eta^2 = .4$, resulting in strong linear main effect of mixes, $F_{angry\text{-}sad}(1,231) = 119.22$; $p < .001$; $\eta^2 = .34$; achieved power was over 99% ($\alpha = .05$ and $\varepsilon = .077$). Critically, there was no interaction between blend level and condition, replicating the pattern from Experiment 3.[2]

Mediation

Our theoretical account predicts that fluency mediates the relationship between expression mixed-ness and evaluation of sad-happy morphs in the emotion categorisation condition. Our tests for mediation employed the multilevel approach recommended by Bauer et al. (2006). Our predictor was the expression mixed-ness (ranging from 1 [pure sadness and pure happiness] to 7 [maximally blended]). Our DV was the trust rating (standardised), and the mediator was fluency (log-transformed categorisation RTs).

First, we regressed fluency onto "mixed-ness" (*a*-path), which showed that blended faces were less fluent than pure faces: $B = .046$, $SE = .005$, $p < .001$. Next, we regressed evaluations onto fluency (*b*-path), which showed that less fluent targets received more negative evaluations than more fluent targets: $B = -.263$, $SE = .042$, $p < .001$. When controlling for the effect of fluency, the path from "mixed-ness" to evaluative judgments was significant, but reduced in magnitude, $B = -.035$, $SE = .01$, $p < .001$ (c' path), compared to $B = -.045$ (*c*-path). In fact, a test for indirect effects for lower level mediation models (Bauer et al., 2006) indicated a statistically significant indirect effect of fluency ($M = -.013$, $SE = .004$, 95% CI: $-.022$, -0.004). This confirms that lower evaluations of mixed sad-happy faces and higher evaluations of pure faces occurred partly because of differences in processing fluency. Importantly, there was no significant mediation of fluency on the relation between mixed-ness and trust ratings for angry-sad morphs.

Affect and approach-avoid judgments

Finally, we tested how participants perceived changes in affect and approach ratings across different types of mixes. We found linear patterns on affect judgments for both type of mixes (see Figure 6, bottom panels). As the face moved from sadness to happiness, the more positively the person was

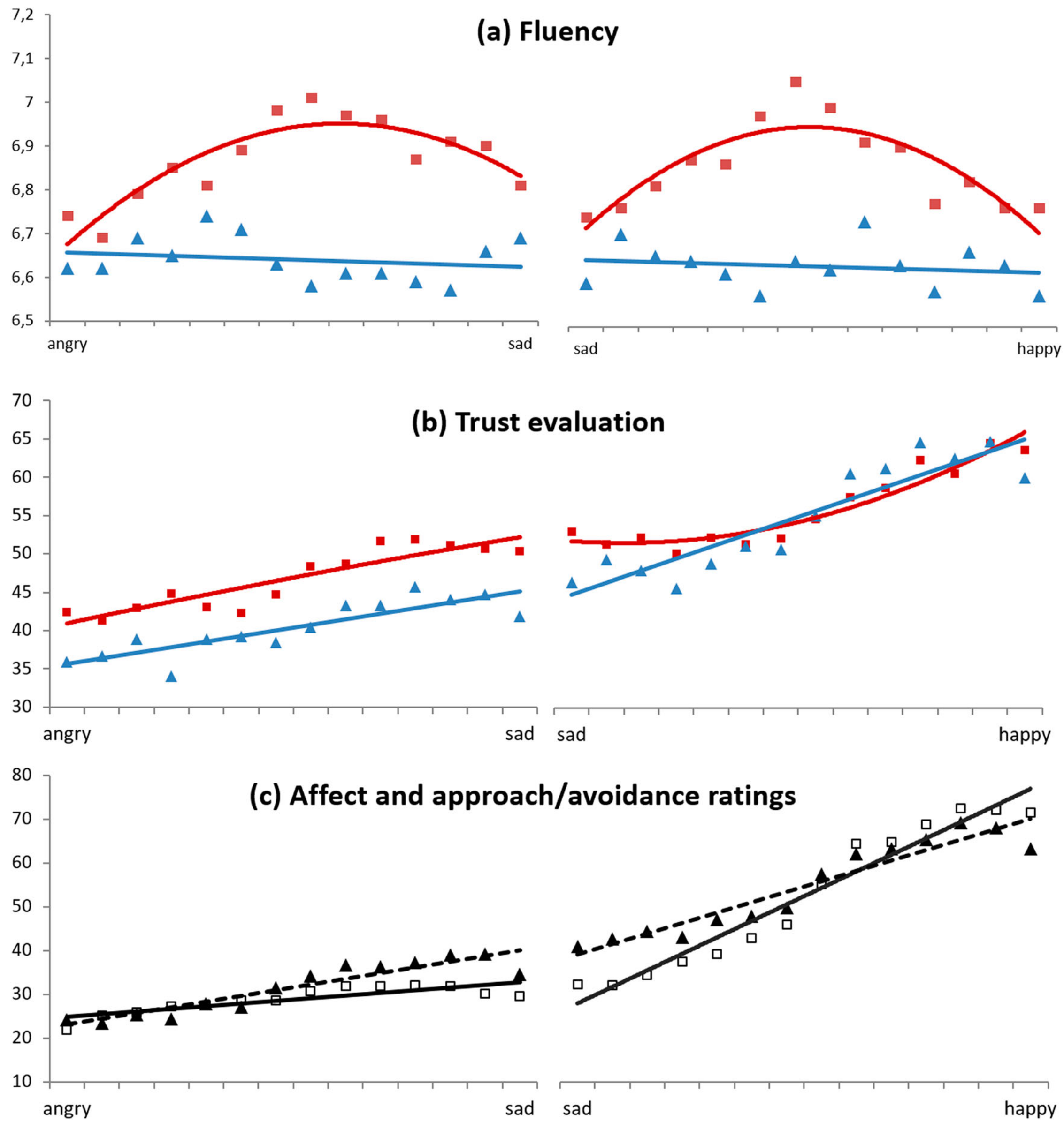

Figure 6. Experiment 4 results for (a) fluency, (b) trust evaluation, (c) affect and approach/avoidance ratings. For (a) and (b), responses are shown as a function of target features (frowning to sadness on the left side and sadness to smile) and categorisation condition (expression as red squares, control as blue triangles), while lines represent an estimated trend. In (c), squares represent ratings of a target face on affect whereas triangles represent its ratings on approach/avoidance, solid and dashed lines represent estimated trends.

evaluated, $F_{sad\text{-}happy}(1,231) = 1156.99$; $p < .001$; $\eta^2 = .84$. Similarly, as the face moved from angry to sad, the more positive the evaluation, $F_{angry\text{-}sad\text{-}}(1,231) = 109.59$; $p < .001$; $\eta_2 = .32$. Note, however, that the former effect was much stronger than the latter effect ($\eta^2 = .84$ vs. .32) and that overall ratings of angry-sad morphs remained low, way below the midpoint of the scale on the dimension of valence (see Figure 5, left panel). We also observed linear patterns for approach-avoid judgments. Happier faces were more approachable than sad faces, $F_{sad\text{-}happy}(1,231) = 475.32$; $p < .001$; $\eta^2 = .67$. Sadder faces were more approachable than angry ones, $F_{angry\text{-}sad\text{-}}(1,231) = 270.17$; $p < .001$; $\eta_2 = .54$.[3]

Importantly, Figure 6 suggests that variation between sad – happy morphs is relatively more related to the valence dimension and angry – sad morphs to the motivational dimension. To test this, we analyzed magnitude of changes separately for each morph type (sad-happy or angry-sad [14 frames]) on two judgments scales (approach/avoidance [which is related to motivation] and affect [which is related to valence]). This analysis revealed an interaction, which shows that affect ratings, as opposed to motivation ratings, were more sensitive to changes from sadness to happiness, $F(13, 221) = 26.27$, $p < .001$, $\eta^2 = .1$. On the other hand, the motivation ratings, as opposed to affect ratings, were more sensitive to changes in angry-sad blends, resulting in an interaction $F(13, 223) = 10.86$, $p = .002$, $\eta^2 = .05$. In short, these results are consistent with the idea that on angry-sad mixes, participants are relatively more attuned to variation in motivational implications.

Meta-analysis of valence-fluency interaction

To estimate the general effects of interaction on trust evaluation of faces mixed on valence dimension (sad-happy) by categorization condition, we ran a meta-analysis by grouping three of our experiments (Experiments 1, 2, and 4). The analysis was conducted using the Comprehensive Meta-Analysis V3 software. The data were computed by Cohens' d (standardized by SD of difference scores) and 95% confidence levels. The results of this analysis are presented in Table 1 and show the robust overall effect of these three experiments. Additionally, we also present results of (see Winkielman, Olszanowski & Gola, 2015, Experiments 1, 2, and 3). These confirm the presence of a strong overall effect of mixed valence on trust ratings, while analogous analyses of angry-sad mixes produce non-significant effects for two experiments presented in this paper (Table 1).

Conclusions

Facial features that resemble emotional expressions contribute to the formation of important social impressions, including trust. In the real world, these expressions often represent a mix of different emotions (Aviezer et al., 2008; Russell, 1997; Sebe et al., 2007). The current research explored the interaction of such expressive features with the perceivers' cognitive goals. We proposed that when emotion categorisation is required, the mixed expressions become disfluent. The four experiments presented here explored when and how such disfluency influences social judgments. This was done to elucidate mechanisms underlying social judgments, including the key judgment of trust. We also wanted to better understand how and when cognitive experiences, such as fluency, are used to form judgments and how such experiences are combined with featural information.

Our findings suggest that trust judgments are influenced by facial features related to valence, as well as features related to social motivation. That is, in Experiments 1 and 2, subjects rated happier faces as more trustworthy than sad faces, suggesting the impact of valence-relevant information. In Experiments 3 and 4, participants rated sad faces as more trustworthy than angry faces, suggesting the impact of motivation-relevant information. These findings replicate previous work on social dimensions underlying trust judgments and extend it by showing that one can empirically separate these valence and motivation dimensions by using different emotional expressions (Oosterhof & Todorov, 2008).

Importantly, our findings show that the impact of features is qualified by fluency. That is, what matters is not only *what* features are perceived but also *how* they are perceived (the mental effort during their processing). Consequently, when expressions varied from sadness to happiness, categorising them on emotion (but not a control dimension) made pure expressions relatively fluent and mixed expressions relatively disfluent. This fluency manipulation influenced trust ratings of those blends, with easier processing predicting higher ratings. This aspect of our study replicates and extends our earlier work with expressions varying from happiness to anger and thus shows the broad nature of the interplay between mental fluency and stimulus features (Winkielman et al., 2015).

The key innovation of the current work, however, lies in specifying how fluency interacts with specific expressive features. The current data suggest that categorisation disfluency impacts trust judgments only when target faces vary on valence. Specifically, in Experiments 1 and 4 using faces on the happy-sad dimension, people found pure expressions relatively more trustworthy and happy-sad mixes relatively less trustworthy, especially when they needed to be first categorised on the emotion dimension. In Experiment 2, we replicated the effect in a different culture and with a different rating procedure. Importantly, we showed that this fluency effect is not driven by simple changes in impressions of positivity, but is driven by more complex inferences underlying trust judgments.

Table 1. Results of meta-analysis for categorization condition & level of expression features (i.e. sad-happy, sad-angry and angry-happy mixes) interaction. Experiments matched with (*) are taken from Winkielman, Olszanowski & Gola (2015).

	Study	*N*	Std diff in means	95% CI Lower	95% CI Upper	*p*-value
	Sad – Happy blends					
1.	Experiment 1	78	0.544	0.099	0.989	.017
2.	Experiment 2	30	0.909	0.188	1.630	.013
3.	Experiment 4	233	0.364	0.106	0.622	.006
Total (1–3)			0.453	0.240	0.666	<.000
	Angry – Happy blends					
4.	Experiment 1*	29	1.013	0.273	1.753	.007
5.	Experiment 2*	53	0.608	0.071	1.145	.027
6.	Experiment 3*	22	1.034	0.195	1.873	.016
Total (1–6)			0.536	0.349	0.722	<.000
	Angry – Sad blends					
7.	Experiment 3	24	0.191	−0.568	0.950	.622
8.	Experiment 4	233	0.052	−0.203	0.307	.690
Total (7–8)			0.066	−0.176	0.308	.592

Interestingly, Experiments 3 and 4 showed no decrease in trustworthiness for mixes of anger and sadness (i.e. emotions that are similar in valence, but different on motivational direction). This occurred even though our fluency manipulation was successful and robustly enhanced difficulty in processing of mixed angry-sad expressions within the emotion-classification condition. Further, the fluency effect on trust judgments was absent even though participants clearly processed the faces and differentiated angry and sad expressions in terms of trust-related features. All in all, this suggests that at least in trust judgments, fluency plays a role in mixtures of valence, but not in mixtures of motivational direction.

Why did fluency fail to impact judgments of sad-angry faces? There are several explanations. We propose an inferential explanation that is consistent with other research on when people use of metacognitive experiences (Schwarz, 2010). When faces themselves vary on valence and participants can base their trust judgment on this variation, fluency, which is itself a source of valence, is perceived as applicable and thus integrated into the final judgment. The same is true when faces are neutral, so participants have no other basis for making their evaluative judgment (e.g. Halberstadt & Winkielman, 2014). However, when faces primarily vary on motivation, participants can derive their trust judgments from evident differences in facial features on that dimension and can ignore any fluency-related valence fluctuations. This could be especially true when both faces have negative valence, like angry-sad, which in our studies participants rated on the negative end of the scale. The idea that the use of fluency is context-sensitive fits with other aspects of our data, such as the observation that disfluency does not lower the ratings of perceived positivity of the face (Experiment 2) or judgments of readability (Experiments 1 and 2). More generally, our findings are consistent with data highlighting the role of "naïve theories" in selective use and interpretation of cognitive experiences (Carr, Hofree, et al., 2017; Cho & Schwarz, 2008; Winkielman, Schwarz, & Belli, 1998).

It is also important to note here that our proposal simply says that when forming impressions of trust, fluency-related affect is more likely to combine with valence-related, rather than motivation-related, features of the face. We do not deny fluency's general role in motivation when it is perceived to be an informative for the judgment at hand. After all, other research shows the fluency can enter into inferences about motivation (Song & Schwarz, 2008), can influence approach-avoidance movements (Carr et al., 2016), and when manipulated via name pronounceability, can even increase trust behaviour in economic games (Zürn & Topolinski, 2017).

Besides our proposed explanation, several other theoretical perspectives are worth considering. One suggests that mixed valence stimuli, as opposed to mixed motivation stimuli, may induce more negative affect from the same amounts of disfluency (Nohlen et al., 2014). A related explanation notes that in social life, individuals sometimes use smiles to conceal negative emotion (Ekman, Friesen, & O'Sullivan, 1988; Niedenthal, Mermillod, Maringer, & Hess, 2010). As such, positive–negative mixes (happy-sad or happy-angry) could contain stronger "wrongness" cues and invite additional skepticism (Alter, 2013). Yet another interpretation could argue that the processing of more "negative blends", including blends of anger with sadness, is more encapsulated from incidental influences (Pinkham, Griffin, Baron, Sasson, & Gur, 2010) or that people are generally more cautious when dealing with negative stimuli (Schwarz, 2015). On the other hand, one could argue that ambiguities related to more negative blends (angry-sad) should be more effective at eliciting elaboration and integration with fluency, compared to more benign blends (happy-sad).

It is also worth noting that in real life, feelings that mix emotions with the same valence occur more often than feelings that mix emotions with opposite valence (Watson & Stanton, 2017). If this holds for emotional expressions, angry-sad faces could be perceived as more common, or at least more sensible than sad-happy mixes. Recall that we found no differences between the size of categorization disfluency effects on different types of blends (angry-sad vs. happy-sad), suggesting that both types of blends were equally difficult to classify. Still, participants may be more likely to consider task fluency as less relevant cue when judging more familiar or sensible angry-sad mixed expression. In other words, for our participants, the equal difficulty of the categorization task may not imply equal "unusualness" of the blends. After all, emotion categorization is much more than a simple assignment into boxes, but a process of making sense of the available information, reconstructing possible appraisal structure, or even making a story of how a given expression came to be (for more elaborate discussion, see Hess, 2017).

Yet another related consideration derives from the literature showing that people use inconsistencies across information channels to detect deception and evaluate trustworthiness (DePaulo et al., 2003). Here, again, mixed valence expressions could be perceived as leaking more inconsistent information and thus leading to lower trust judgments (despite similar disfluency on our simple categorisation task). However, one should note that on a more abstract level, the channel inconsistency account and the fluency account are compatible. After all, consistency is about elements being tied together in a sensible relation, while fluency here is basically a description of a behavioural effect that often results from consistency (i.e. processing is easier and faster; see Winkielman et al., 2012). Thus, on a more abstract level, detecting channel inconsistency involves some disfluency, though perhaps at a different cognitive level, that was measured by our simple categorisation task. All these alternative perspectives should be subject to further studies.

Finally, the current work, along with previous studies (Winkielman et al., 2015), shows that the impact of mixed expressions on key social judgments depends on perceivers' cognitive categorisation goals. Objectively mixed facial expressions lead to small (if any) reduction in trust judgments when participants simply look at the faces (Experiment 4) or first judge them on another dimension (e.g. gender in Experiments 1, 2, and 3; ethnicity in Winkielman et al., 2015, Experiment 4). Importantly, in those control conditions, participants' judgments are still highly sensitive to expression-related information; they show that participants are able to process features, but it is just that their "mixed-ness" matters less. This raises a legitimate question for how often in real-world situations people "categorize" on facial expression first, before formulating trust impression? While a systematic answer requires a future study, evidence suggests that in some conditions, people do spontaneously categorise on emotional expressions (Canadas, Lupianez, Kawakami, Niedenthal, & Rodrıguez-Bailon, 2016). Furthermore, many interpersonal interactions force people into the emotion categorisation context, including daily social and business tasks where quickly classifying an expression is important (Hess, 2017).

In sum, we suggest that this work demonstrates the need for a more specific understanding of the interplay between stimulus features and processing experiences. This is not only theoretically important, but also helps us better understand how people make key social judgments, such as trust.

Notes

1. We also found a relatively weak but significant cubic trend, $F(1, 28) = 4.41$; $p < .05$; $\eta^2 = .14$.
2. There was also a main effect for condition on angry-sad mixes. The expression categorisation group gave overall higher trust ratings than the control group, $F(1, 231) = 25.2$; $p < .001$; $\eta^2 = .1$
3. Sad-happy morphs also showed a cubic trend on affect judgments, $F(1, 231) = 242.75$; $p < .001$; $\eta^2 = .51$ and approach judgments, $F(1, 231) = 109.99$; $p < .001$; $\eta^2 = .33$. Angry-sad morphs show quadratic effects for affect judgments, $F(1, 231) = 61.47$; $p < .001$; $\eta^2 = .21$, and a cubic effect for approach judgments, $F(1, 231) = 43.02$; $p < .001$; $\eta^2 = .16$.

Acknowledgements

We thank Evan Carr and Andy Arnold for their help and research assistants at SWPS and UCSD for running the studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by National Science Center, Poland [grant number 2011/01/B/HS6/0335 to PW & MO] and from the UCSD Academic Senate Grant to PW.

References

AdamsJr, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion*, *30*, 177–186.

Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, *22*, 437–442.

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, *103*, 9369–9372.

Ambady, N., & Weisbuch, M. (2011). Facial expressions: Culture and context. In A. J. Calder, G. Rhodes, J. V. Haxby, & M. H. Johnson (Eds.), *The handbook of face perception* (pp. 479–488). Oxford: Oxford University Press.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732.

Bach, D. R., Guitart-Masip, M., Packard, P. A., Miró, J., Falip, M., Fuentemilla, L., & Dolan, R. J. (2014). Human hippocampus arbitrates approach-avoidance conflict. *Current Biology*, *24*, 541–547.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, *40*, 35–55.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Brownlow, S., & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior*, *14*, 51–60.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*, 1179–1208.

Canadas, E., Lupianez, J., Kawakami, K., Niedenthal, P. M., & Rodrıguez-Bailon, R. (2016). Perceiving emotions: Cueing social categorization processes and attentional control through facial expressions. *Cognition and Emotion*, *30*, 1149–1163.

Carr, E. W., Brady, T. F., & Winkielman, P. (2017). Are you smiling or have I seen you before? Familiarity makes faces look happier. *Psychological Science*, *28*(8), 1087–1102.

Carr, E. W., Hofree, G., Sheldon, K., Saygin, A. P., & Winkielman, P. (2017). Is that a human? Categorization (dis)fluency drives evaluations of agents ambiguous on human-likeness. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 651–666.

Carr, E. W., Rotteveel, M., & Winkielman, P. (2016). Easy moves: Perceptual fluency facilitates approach-related action. *Emotion*, *16*, 540–552.

Cho, H., & Schwarz, N. (2008). Of great art and untalented artists: Effort information and the flexible construction of judgmental heuristics. *Journal of Consumer Psychology*, *18*, 207–213.

Clark, M. S., & Taraban, C. B. (1991). Reactions to and willingness to express emotion in communal and exchange relationships. *Journal of Experimental Social Psychology*, *27*, 324–336.

Cuddy, A. J. C., Fiske, S. T., Kwan, V., Glick, P., Demoulin, S., Leyens, J.-P., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*, 1–33.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118.

Dimberg, U., & Ohman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion*, *20*, 149–182.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, *1*, 0001.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, *51*, 665–697.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*, 414–420.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, *30*, 527–554.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology*, *85*, 276–290.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Halberstadt, J., & Winkielman, P. (2014). Easy on the eyes, or hard to categorize: Classification difficulty decreases the appeal of facial blends. *Journal of Experimental Social Psychology*, *50*, 175–183.

Hess, U. (2017). Emotion categorization. In C. Lefebvre & H. Cohen (Eds.), *Handbook of categorization in cognitive science* (2nd ed., pp. 107–126). Amsterdam: Elsevier.

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion (Washington, D.C.)*, *3*, 150–166.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. W: H. L. Roediger, F. I. M. Craik (red.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (s. 391–422). Hillsdale, NJ: Erlbaum.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*, 165–182.

Lee, A. Y., & Labroo, A. (2004). The effect of conceptual and perceptual fluency on brand evaluation. *Journal of Marketing Research*, *41*, 151–165.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion*, *5*, 119–124.

Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus

collectivism. *Journal of Cross-cultural Psychology*, *39*(1), 55–74.
Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*, 237–254.
Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, *9*, 5. doi:10.1167/9.1.5
Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*, 417–433.
Nohlen, H. U., van Harreveld, F., Rotteveel, M., Lelieveld, G.-J., & Crone, E. A. (2014). Evaluating ambivalence: Social-cognitive and affective brain regions associated with ambivalent decision-making. *Social Cognitive and Affective Neuroscience*, *9*, 924–931.
Olszanowski, M., Pochwatko, G., Kukliński, K., Ścibor-Rylski, M., Lewinski, P., & Ohme, R. (2015). Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*, *5*, 1516. doi:10.3389/fpsyg.2014.01516
Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092.
Owen, H. E., Halberstadt, J., Carr, E. W., & Winkielman, P. (2016). Johnny Depp, reconsidered: How category-relative processing fluency determines the appeal of gender ambiguity. *PLoS ONE*, *11*(2), e0146328.
Paulus, A., & Wentura, D. (2015). It depends: Approach and avoidance reactions to emotional expressions are influenced by the contrast emotions presented in the task. *Journal of Experimental Psychology. Human Perception and Performance*, *2*, 197–212.
Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: Anger superiority when using real faces and multiple identities. *Emotion (Washington, D.C.)*, *10*, 141–146.
Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE*, *7*(3), e34293.
Russell, J. A. (1997). Reading emotion from and into faces: Resurrecting a dimensional-contextual perspective. In J. A. Russell, & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 295–320). New York: Cambridge University Press.
Schneirla, T. C. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. Jones (Ed.), *Nebraska symposium on motivation* (pp. 1–42). Lincoln, NE: University of Nebraska Press.
Schwarz, N. (2010). Meaning in context: Metacognitive experiences. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 105–125). New York: Guilford.
Schwarz, N. (2015). Metacognition. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology: Attitudes and social cognition* (pp. 203–229). Washington, DC: APA.
Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*, 243–265.
Sebe, N., Lew, M., Sun, Y., Cohen, I., Gevers, T., & Huang, T. S. (2007). Authentic facial expression analysis. *Image and Vision Computing*, *25*, 1856–1863.
Smith, F. W., & Schyns, P. G. (2009). Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science*, *20*, 1202–1208.
Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, *19*, 986–988.
Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*, 349–354.
Topolinski, S., Likowski, K. U., Weyers, P., & Strack, F. (2009). The face of fluency: Semantic coherence automatically elicits a specific pattern of facial muscle reactions. *Cognition and Emotion*, *23*, 260–271.
van Kleef, G. A., de Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, *42*, 45–96.
Van Kleef, G. A., Oveis, C., van der Löwe, I., LuoKogan, A., Goetz, J., & Keltner, D. (2008). Power, distress, and compassion: Turning a blind eye to the suffering of others. *Psychological Science*, *19*, 1315–1322.
Watson, D., & Stanton, K. (2017). Emotion blends and mixed emotions in the hierarchical structure of affect. *Emotion Review*, *9*(2), 99–104.
Wilkowski, B. M., & Meier, B. P. (2010). Bring it on: Angry facial expressions potentiate approach-motivated motor behavior. *Journal of Personality and Social Psychology*, *98*, 201–210.
Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, *31*, 121–135.
Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation increases positive affect. *Journal of Personality and Social Psychology*, *81*, 989–1000.
Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, *17*, 799–806.
Winkielman, P., Huber, D. E., Kavanagh, L., & Schwarz, N. (2012). Fluency of consistency: When thoughts fit nicely and flow smoothly. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 89–111). New York: Guilford Press.
Winkielman, P., Olszanowski, M., & Gola, M. (2015). Faces In-between: Evaluations reflect the interplay of facial features and task-dependent fluency. *Emotion*, *15*, 232–242.
Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science*, *9*, 124–126.
Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch, & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion*. (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251–1263.
Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*, 175–189.
Zürn, M., & Topolinski, S. (2017). When trust comes easy: Articulatory fluency increases transfers in the trust game. *Journal of Economic Psychology*, *61*, 74–86.

Appendix A

MLM model

All repeated-measures analyses (including classification RTs and ratings) used multilevel modelling (MLM) via maximum likelihood. The MLM procedure used on a fixed effects structure that included Condition (2 levels: gender and expression categorisation) and Sad-Happy or Angry-Sad Morph Level (14 values: SH_{morph} or AS_{morph}). SH_{morph} and AS_{morph} was modelled using continuous fixed effects for both the linear component (SH^1_{morph} or AS^1_{morph}) and quadratic component (SH^2_{morph} or AS^2_{morph}), as well as their respective interactions with Condition. Random effects for the model included intercepts fit across subjects and face of individuals used as stimuli.

Experiment 1: On Rts we observed main effects on SH^1_{morph}, $F(1, 2183.91) = 19.83$, $p < .001$, and SH^2_{morph}, $F(1, 2183.32) = 27.68$, $p < .001$. We also detected the predicted quadratic Condition × SH^2_{morph} interaction, $F(1, 2183.32) = 13.44$, $p < .001$ and smaller in size linear interaction of Condition × $SH^1_{morph}F(1, 2183.91) = 8.69$, $p < .01$.

For trust ratings analysis showed main effect on SH^1_{morph}, $F(1, 2690.55) = 140..4$, $p < .001$, and SH^2_{morph}, $F(1, 2692.05) = 115.98$, $p < .001$. We also observed significant lineal interaction of Condition and SH^1_{morph}, $F(1, 2690.55) = 5.84$, $p < .05$; while quadratic interaction was near to significant – Condition and SH^2_{morph}, $F(1, 2808) = 4.37$, $p <.05$.

For readability we found significant only two of main effect on SH^1_{morph}, $F(1, 2683.44) = 75.72$, $p < .001$, and SH^2_{morph}, $F(1, 2685.11) = 2692$, $p < .001$.

Experiment 2: With this experiment, we used the same MLM strategy. On RTs we observed significant main effect on SH^1_{morph}, $F(1, 386) = 21.46$, $p < .001$, and SH^2_{morph}, $F(1, 386) = 19.87$, $p < .001$, as well as significant lineal interaction on Condition and SH^1_{morph}, $F(1, 386) = 9.91$, $p < .01$ and quadratic interaction on Condition and SH^2_{morph}, $F(1, 386) = 10.36$, $p < .01$.

For trust ratings MLM analysis we found significant main effect on SH^1_{morph}, $F(1, 338.77) = 58.04$, $p < .001$, and SH^2_{morph}, $F(1, 343.95) = 24.71$, $p < .001$, as well as significant interaction on Condition and SH^2_{morph}, $F(1, 343.95) = 6.24$, $p < .05$.

Readability judgments showed main effects on SH^1_{morph}, $F(1, 381.73) = 103.87$, $p < .001$, and SH^2_{morph}, $F(1, 359.85) = 70.52$, $p < .001$ but also on Condition $F(1, 76.96) = 4.48$, $p < .05$, as well as significant interactions on Condition and SH^1_{morph}, $F(1, 381.73) = 11.44$, $p < .01$, Condition and SH^2_{morph}, $F(1, 359.85) = 17.75$, $p < .001$.

Finally on expression positivity we found main effect on SH^1_{morph}, $F(1, 165.02) = 70.79$, $p < .001$, and SH^2_{morph}, $F(1, 348.87) = 21.27$, $p < .001$. As it is worth noticing repeated measures ANOVA did not showed significance of such quadratic main effect.

Experiment 3: For RTs MLM analysis showed significant linear interaction on Condition and AS^1_{morph} interaction – $F(1, 1987.54) = 4.79$, $p < .05$ and near to significant quadratic interaction on Condition and AS^2_{morph} – $F(1, 1987.87) = 3.78$, $p = .052$.

Probably due to relatively small experimental sample multilevel modelling did not fully confirm expected patterns of results on trust ratings, as linear effect on AS^1_{morph} was only near to significant – $F(1, 1988.54) = 3.74$, $p = .053$, with no significant interactions. However worth noticing is that at the same time quadratic effect on AS^2_{morph}, as we expected, was not significant ($F(1, 1988.53) = .38$, $p = .54$).

Experiment 4: On RT we observed significant main effect on SH^1_{morph}, $F(1, 4749.48) = 28.81$, $p < .001$, and SH^2_{morph}, $F(1,4749.49) = 29.33$, $p < .001$ and also interaction Condition and SH^1_{morph}, $F(1, 4749.48) = 23.34$, $p < .001$, and SH^2_{morph}, $F(1,4749.49) = 24.02$, $p < .001$. Significant main effect was also

Table A1. Descriptive Statistics, Experiment 1.

Condition	Emotion			Gender		
Measure	*Log RT M (SD)*	*Trust M (SD)*	*Intention M (SD)*	*Log RT M (SD)*	*Trust M (SD)*	*Intention M (SD)*
Frame_1 (sad)	6.29 (1.07)	5.92 (2.49)	4.57 (1.76)	6.73 (0.99)	5.94 (1.87)	4.79 (1.55)
Frame_2	6.23 (1.03)	5.84 (2.35)	4.69 (1.93)	6.62 (0.99)	6.14 (1.88)	4.87 (1.62)
Frame_3	6.44 (0.91)	5.94 (2.07)	5.01 (1.54)	6.64 (0.96)	6.13 (1.79)	4.97 (1.68)
Frame_4	6.54 (1.07)	6.06 (2.08)	4.77 (1.61)	6.60 (0.93)	5.92 (1.83)	4.82 (1.62)
Frame_5	6.55 (0.97)	5.65 (2.16)	4.71 (1.44)	6.53 (0.91)	5.68 (1.89)	4.77 (1.56)
Frame_6	6.71 (0.99)	5.36 (2.22)	4.93 (1.66)	6.74 (0.88)	5.75 (1.86)	4.93 (1.43)
Frame_7	6.87 (1.08)	5.17 (2.17)	4.80 (1.64)	6.72 (1.02)	5.42 (2.04)	4.83 (1.49)
Frame_9	6.83 (1.04)	5.15 (2.08)	5.01 (1.69)	6.76 (0.90)	5.22 (1.72)	4.81 (1.65)
Frame_10	6.76 (1.10)	5.28 (2.13)	5.41 (1.73)	6.62 (1.11)	5.63 (1.64)	5.43 (1.58)
Frame_11	6.81 (1.21)	5.56 (2.08)	5.45 (1.88)	6.55 (0.87)	5.52 (1.80)	5.32 (1.52)
Frame_12	6.63 (1.13)	6.30 (1.99)	5.76 (1.74)	6.81 (0.76)	6.49 (1.78)	5.93 (1.71)
Frame_13	6.56 (0.98)	6.69 (1.75)	6.00 (1.57)	6.84 (0.87)	6.30 (1.77)	5.84 (1.65)
Frame_14	6.49 (0.94)	6.88 (1.87)	6.24 (1.93)	6.37 (1.13)	6.41 (1.63)	5.79 (1.55)
Frame_15 (happy)	6.46 (1.08)	6.86 (1.97)	6.33 (1.71)	6.66 (0.69)	6.43 (1.84)	5.72 (1.95)

Table A2. Descriptive Statistics, Experiment 2.

Condition	Emotion				Gender			
Measure	*Log RT M (SD)*	*Trust M (SD)*	*Affect M (SD)*	*Readability M (SD)*	*Log RT M (SD)*	*Trust M (SD)*	*Affect M (SD)*	*Readability M (SD)*
Frame_1 (sad)	7.00 (0.41)	49.4 (23.1)	45.4 (26.0)	61 (25.3)	7.05 (0.33)	39.0 (23.0)	37.3 (21.8)	45.5 (24.2)
Frame_2	7.03 (0.4)	48.9 (20.9)	45.0 (25.0)	57.2 (24.8)	7.12 (0.40)	40.1 (25.0)	39.6 (22.5)	46.6 (27.6)
Frame_3	7.02 (0.4)	45.7 (22.6)	41.6 (23.8)	51.9 (26.3)	7.11 (0.39)	41.4 (23.3)	40.6 (20.1)	44.8 (26.1)
Frame_4	7.11 (0.49)	46.7 (27.0)	42.3 (24.6)	52.3 (30.8)	7.18 (0.39)	38.7 (21.8)	38.4 (19.8)	48.2 (24.8)
Frame_5	7.26 (0.52)	45.7 (24.7)	44.7 (24.4)	48.9 (26.4)	7.04 (0.31)	43.2 (21.8)	41.9 (21.2)	46.4 (24.7)
Frame_6	7.25 (0.49)	51.1 (24.3)	49.4 (21.0)	47.1 (26.1)	7.16 (0.38)	41.3 (22.3)	42.0 (20.5)	45.4 (23.8)
Frame_7	7.49 (0.56)	47.7 (22.0)	50.2 (23.5)	47.0 (25.5)	7.00 (0.43)	49.2 (22.6)	48.4 (22.7)	49.0 (21.7)
Frame_9	7.48 (0.62)	44.3 (20.7)	49.5 (22.4)	45.9 (22.6)	7.03 (0.35)	46.5 (22.8)	50.7 (23.0)	48.4 (22.5)
Frame_10	7.11 (0.48)	48.3 (24.4)	54.9 (24.0)	47.6 (27)	7.18 (0.40)	49.2 (25.8)	56.4 (24.5)	51.5 (24.4)
Frame_11	7.11 (0.5)	51.1 (23.6)	61.9 (25.5)	55.8 (26.4)	7.22 (0.53)	53.8 (24.4)	59.5 (24.6)	55.0 (22.7)
Frame_12	7.32 (0.39)	57.0 (24.5)	66.2 (26.1)	60.3 (25.8)	6.91 (0.33)	58.8 (24.8)	63.1 (25.3)	57.9 (27.9)
Frame_13	6.88 (0.32)	60.3 (26.2)	67.1 (26.6)	63.8 (22.5)	6.96 (0.28)	64.8 (23.2)	70.0 (22.3)	64.9 (21.9)
Frame_14	6.99 (0.41)	58.0 (25.6)	70.1 (13.8)	66.8 (13.6)	6.93 (0.34)	61.45 (12.95)	67.5 (26.3)	64.9 (25.7)
Frame_15 (happy)	6.98 (0.34)	60.9 (23.4)	71.8 (14.8)	67.8 (13.1)	7.02 (0.34)	57.51 (14.75)	64.8 (28.3)	61.0 (25.6)

Table A3. Descriptive Statistics, Experiment 3.

Condition	Emotion		Gender	
Measure	*LogRT M (SD)*	*Trust M (SD)*	*LogRT M (SD)*	*Trust M (SD)*
Frame_1 (angry)	7.06 (0.36)	36.3 (18.7)	6.83 (0.32)	32.9 (20.9)
Frame_2	7.13 (0.44)	41.2 (21.9)	6.91 (0.37)	36.1 (22.3)
Frame_3	7.01 (0.39)	37.3 (18.3)	6.91 (0.40)	36.8 (21.3)
Frame_4	7.07 (0.42)	40.0 (20.5)	6.88 (0.34)	37.8 (23.4)
Frame_5	7.19 (0.49)	36.7 (17.0)	6.85 (0.34)	37.3 (21.3)
Frame_6	7.10 (0.46)	43.6 (20.0)	6.86 (0.37)	44.8 (20.9)
Frame_7	7.25 (0.50)	44.5 (20.2)	6.75 (0.33)	41.6 (21.4)
Frame_9	7.15 (0.42)	46.9 (21.2)	6.93 (0.37)	43.8 (22.4)
Frame_10	7.01 (0.34)	48.0 (19.4)	6.93 (0.36)	43.6 (20.7)
Frame_11	7.14 (0.50)	50.3 (19.6)	6.86 (0.33)	42.1 (21.8)
Frame_12	7.09 (0.46)	50.3 (19.9)	6.91 (0.40)	44.7 (23.7)
Frame_13	7.02 (0.44)	50.1 (21.3)	6.87 (0.34)	47.9 (23.6)
Frame_14	7.02 (0.44)	50.5 (18.7)	6.91 (0.41)	44.3 (23.4)
Frame_15 (sad)	7.03 (0.43)	50.0 (20.7)	6.87 (0.35)	47.4 (25.0)

on AS^1_{morph} – $F(1, 4679.89) = 11.93$, $p < .01$ and AS^2_{morph} – $F(1, 4676.73) = 17.54$, $p < .001$, as well as interaction on Condition and AS^1_{morph} – $F(1, 4679.89) = 9.82$, $p < .01$ and AS^2_{morph} – $F(1, 4676.73) = 17.79$, $p < .001$.

For trust ratings on Sad-Happy morphs analysis showed linear effect- SH^1_{morph}, $F(1, 4734.46) = 82.67$, $p < .001$, and quadratic effect – SH^2_{morph}, $F(1,4733.08) = 13.76$, $p < .001$. We also observed significant interaction of Condition and SH^1_{morph}, $F(1, 4734.46) = 6.07$, $p < .05$; and Condition and SH^2_{morph}, $F(1, 4733.08) = 11.8$, $p < .01$. While for Angry-Sad morphs there was significant main effects on AS^1_{morph} – $F(1, 4669.89) = 7.55$, $p < .01$ and main effect on Condition $F(1, 938.09) = 19.61$, $p < .001$.

As both experimental condition contained the same type of affective valence and approach-avoidance judgments procedure (without any task proceeding evaluation) we did not calculate aforementioned dimensions in MLM procedures.

Table A4. Descriptive Statistics, Experiment 4.

Condition	Emotion		Control	
Measure	*LogRT M (SD)*	*Trust M (SD)*	*LogRT M (SD)*	*Trust M (SD)*
Frame_1 (angry)	6.74 (0.48)	42.1 (16.8)	6.62 (0.53)	36.8 (17.5)
Frame_2	6.69 (0.45)	41.5 (17.1)	6.62 (0.56)	36.5 (16.2)
Frame_3	6.79 (0.40)	42.6 (15.1)	6.69 (0.48)	38.7 (15.4)
Frame_4	6.85 (0.52)	44.5 (16.2)	6.65 (0.63)	33.6 (15.1)
Frame_5	6.81 (0.42)	42.1 (18.2)	6.74 (0.69)	37.5 (13.8)
Frame_6	6.89 (0.44)	42.6 (16.9)	6.71 (0.54)	38.3 (15.4)
Frame_7	6.98 (0.53)	44.3 (14.8)	6.63 (0.67)	38.0 (15.6)
Frame_9	7.01 (0.49)	48.3 (15.5)	6.58 (0.62)	39.4 (15.1)
Frame_10	6.97 (0.47)	49.0 (16.0)	6.61 (0.48)	43.9 (13.9)
Frame_11	6.96 (0.38)	52.1 (14.)	6.61 (0.56)	43.4 (13.8)
Frame_12	6.87 (0.43)	51.6 (15.5)	6.59 (0.67)	46.0 (15.8)
Frame_13	6.91 (0.41)	51.2 (16.7)	6.57 (0.61)	43.8 (13.8)
Frame_14	6.90 (0.39)	51.2 (14.3)	6.66 (0.61)	44.7 (15.4)
Frame_15 (sad)	6.81 (0.34)	50.9 (17.8)	6.69 (0.62)	42.6 (17.0)
Frame_1 (sad)	6.75 (0.40)	53.2 (15.5)	6.60 (0.64)	46.8 (15.5)
Frame_2	6.77 (0.34)	50.8 (15.7)	6.71 (0.63)	49.8 (12.7)
Frame_3	6.82 (0.36)	52.7 (14.2)	6.66 (0.54)	47.7 (12.0)
Frame_4	6.88 (0.41)	50.3 (17.6)	6.65 (0.56)	45.6 (14.1)
Frame_5	6.87 (0.36)	52.5 (15.2)	6.62 (0.63)	47.9 (12.5)
Frame_6	6.98 (0.49)	52.1 (13.6)	6.57 (0.67)	51.9 (13.9)
Frame_7	7.06 (0.48)	52.5 (13.3)	6.65 (0.59)	50.2 (13.8)
Frame_9	7.00 (0.44)	55.5 (15.2)	6.63 (0.60)	53.9 (15.4)
Frame_10	6.92 (0.45)	58.6 (15.7)	6.74 (0.56)	60.7 (14.6)
Frame_11	6.91 (0.41)	59.1 (15.8)	6.64 (0.57)	62.6 (16.2)
Frame_12	6.78 (0.37)	63.6 (14.9)	6.58 (0.63)	64.0 (16.8)
Frame_13	6.83 (0.43)	61.4 (18.2)	6.67 (0.67)	63.8 (15.5)
Frame_14	6.77 (0.38)	66.0 (15.6)	6.64 (0.58)	65.2 (17.1)
Frame_15 (happy)	6.77 (0.33)	63.9 (18.2)	6.57 (0.71)	59.8 (19.2)