Evidence for Distinct Facial Signals of Reward, Affiliation, and Dominance from Both Perception and Production Tasks

Jared D. Martin[1] · Adrienne Wood[2] · William T. L. Cox[3] · Scott Sievert[4] · Robert Nowak[4] · Eva Gilboa-Schechtman[5] · Fangyun Zhao[3] · Zachary Witkower[6] · Andrew T. Langbehn[3] · Paula M. Niedenthal[3]

Abstract

Smiles are nonverbal signals that convey social information and influence the social behavior of recipients, but the precise form and social function of a smile can be variable. In previous work, we have proposed that there are at least three physically distinct types of smiles associated with specific social functions: reward smiles signal positive affect and reinforce desired behavior, affiliation smiles signal non-threat and promote peaceful social interactions, dominance smiles signal feelings of superiority and are used to negotiate status hierarchies. The present work advances the science of the smile by addressing a number of questions that directly arise from this smile typology. What do perceivers *think* when they see each type of smile (study 1)? How do perceivers *behave* in response to each type of smile (study 2)? Do people *produce* three physically distinct smiles in response to contexts related to each of the three social functions of smiles (study 3)? We then use an online machine learning platform to uncover the labels that lay people use to conceptualize the smile of affiliation, which is a smile that serves its social function but lacks a corresponding lay concept. Taken together, the present findings support the conclusion that reward, affiliation, and dominance smiles are distinct signals with specific social functions. These findings challenge the traditional assumption that smiles merely convey whether and to what extent a smiler is happy and demonstrate the utility of a social–functional approach to the study of facial expression.

Keywords Smiles · Facial expression · Social–functionalism · Nonverbal behavior

Introduction

Smiles are versatile and powerful social signals (Jensen, 2015; Kraus & Chen, 2013; Kunz et al., 2009; Scharlemann et al.,

✉ Jared D. Martin
  jm5788@nyu.edu

1 Department of Psychology, New York University, 6 Washington Pl., New York, NY 10003, USA

2 Department of Psychology, University of Virginia, Charlottesville, VA, USA

3 Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

4 Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI, USA

5 Department of Psychology, Bar-Ilan University, Ramat Gan, Israel

6 Department of Psychology, University of British Columbia, Vancouver, BC, Canada

2001) that are used to accomplish a diverse set of social goals across interpersonal contexts (Johnston et al., 2010; Stewart et al., 2015). Mirroring the heterogeneity of the contexts in which they are encountered, the morphology of expressions labeled as a smile is also highly variable (Ambadar et al., 2009; Harris & Alvarado, 2005). Although all smiles involve upturned lip corners from activation of the *zygomaticus major* muscle (Ekman & Friesen, 1978), they differ in the extent to which that activation is symmetrical and whether other muscles participate in the expression (Ekman, 2009; Ekman & Friesen, 1982). Thus, smiles are highly variable in their physical form and the situations in which they are encountered. In the present article, we use a social–functional approach to explain this variability (Martin et al., 2017; Niedenthal et al., 2010).

The Classical Smile Perspective

Historically, the dominant approach to smile classification was based on the presence versus absence of "crow's feet" around the eyes, termed the *Duchenne marker* (Duchenne,

1876). Early findings showed that the Duchenne marker occurs relatively automatically during states of positive emotion (Ekman et al., 1990; Ekman et al., 1988). Thus, smiles that include the Duchenne marker (termed Duchenne smiles) were thought to indicate positive affect. Conversely, smiles lacking the Duchenne marker (termed non-Duchenne smiles) were thought to indicate the broad range of internal states outside positive affect, including—but not limited to—politeness and feigned pleasure (Ekman & Friesen, 1982).

Because Duchenne smiles have been held to signal felt positive emotion, the distinction between Duchenne and non-Duchenne smiles is often conflated with a similar approach that separates smiles into those that are true (or *genuine*) versus false, based on whether or not the smiler is simultaneously experiencing positive emotion (for further discussion, see Martin et al., 2017). It should be noted, however, that these two approaches are not necessarily identical: a smile can be a true indicator of positive feelings without necessarily involving the Duchenne marker, and people are capable of producing Duchenne smiles in the absence of positive affect (Gunnery et al., 2012; Krumhuber & Manstead, 2009). Thus, whereas the Duchenne/non-Duchenne distinction relies on physical features to categorize smiles, the true/false distinction relies on the internal state of the smiler (inferred or self-reported) to categorize smiles.

Both the Duchenne/non-Duchenne and true/false distinctions have been useful for advancing research on the smile. However, recent work shows that these categories do not account for the physical variability and social nuances of smiles. For example, spontaneous smiles are not a morphologically homogeneous category (Krumhuber & Manstead, 2009). Furthermore, the physical features that drive perceptions of smile genuineness vary depending on contextual factors, such as culture-based facial expression "dialects" (Maringer et al., 2011; Thibault et al., 2012). Taken together, recent work suggests that perceivers likely mentally represent the social and physical complexity of smiles in ways not wholly captured by the Duchenne/non-Duchenne or true/false distinctions.

A Social–Functional Approach to Smiles

The inherent sociality of smiles (Fridlund, 1991; Kraut & Johnston, 1979) suggests that smiles are first and foremost important social signals, lending them to a *social–functional* analysis (see, for example, Fischer & Manstead, 2008; Keltner & Haidt, 1999; van Dijk et al., 2008). We began to build a social–functional account of smiles by assuming that facial expressions are communicative signals that coordinate social interactions and help individuals accomplish tasks fundamental to successful social living (Martin et al., 2017; Niedenthal et al., 2010). We then considered the many contexts and internal states associated with smiles and distilled them to a smaller set of foundational social tasks (Kenrick et al.,

2016). As an example of this type of analysis, some smiles are thought to convey embarrassment, polite motives, and friendly intentions (Hoque et al., 2011; Keltner, 1995). The underlying social task uniting these disparate contexts and motivations is that of signaling positive social goals. Thus, a social–functional approach to smile categorization would consider these smiles as instances of a larger category of *affiliation smiles*.

Employing this social–functional approach, we have previously argued that smiles can accomplish *at least* three fundamental tasks of social living, including (i) reinforcing desired behavior in the self and others (reward), (ii) indicating non-threat and openness to non-agonistic interaction (affiliation), and (iii) negotiating social status (dominance; Martin et al., 2017; Niedenthal et al., 2010; Rychlowska et al., 2017). Consistent with this framework, perceiver-generated mental prototypes of the three smiles involve unique facial movements (Rychlowska et al., 2017), described in terms of "action units" or AUs in the Facial Action Coding System ("FACS"; Ekman & Friesen, 1978). Findings revealed that reward smiles included AUs 1, 2, 13, and 14; affiliation smiles included AUs 14 and 24; and dominance smiles included AUs 5, 6, 9, and 10, and asymmetrical AU12 activation. A table of AUs encountered in this paper and their corresponding physical descriptions is included in Table 1 for convenience.

Table 1 An index of action unit (AU) codes and descriptive labels for the AUs referenced in this manuscript

Action unit	Descriptive label
AU1	Inner brow raiser
AU2	Outer brow raiser
AU4	Brow lowerer
AU5	Upper lid raiser
AU6	Cheek raiser
AU7	Lid tightener
AU9	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depressor
AU16	Lower lip depressor
AU17	Chin raiser
AU20	Lip stretcher
AU23	Lip tightener
AU24	Lip pressor
AU25	Lips part
AU26	Jaw drop
AU27	Mouth stretch
AU43	Blink

Previous research on the facial movements involved in reward, affiliation, and dominance smiling (Rychlowska et al., 2017) is incomplete for several reasons. First, past research relied on computer-generated avatars; although providing a controlled environment for precise manipulation of facial movements, they lack ecological validity. Second, computer-generated avatars used in past research were only capable of closed-mouth smiles, which is problematic given that separation of the lips (i.e., AU 25) is often present in smiles (Ambadar et al., 2009). Third, prior work has relied on perceiver's forced-choice categorizations of smile stimuli, so we do not yet know how people interpret or naturally generate smiles in free-response contexts. Fourth, little to no work explains how people's behavior differs depending on how their interaction partner smiles which is critical in light of our social–functional account's focus on the *social consequences* of smiles.

In the present research, we improve upon the limitations of previous research on reward, affiliation, and dominance smiles and sought to understand how people produce, think about, and respond to reward, affiliation, and dominance smiles. In addition, we also sought to understand why perceivers commonly mis-classify affiliation smiles as reward smiles. We reasoned that perceivers may not have a readily accessible label for affiliation smiles, in part due to lack of social consensus around how to describe such smiles. To investigate this possibility, we employed a crowd-sourced judgment task to map the structure of the labels perceivers mentally apply to affiliation smiles.

Method

Overview of the Present Studies

Across four studies, we test how people think (studies 1 & 4) and behave (study 2) in response to reward, affiliation, and dominance smiles and whether people use distinct facial movements to encode social motives proposed as being related to each type of smile (study 3). In study 1, we analyze text descriptions of reward, affiliation, and dominance smiles for evidence that each smile type calls to mind divergent social contexts related to their proposed social functions. In study 2, we test for distinct behavioral consequences of receiving reward, affiliation, or dominance smiles during an economic game, an indication that each smile sends a different social signal. In study 3, we examine whether experimentally naïve producers use physically distinct facial movements when encoding social contexts related to the proposed functions of smiles. In study 4, we use the findings from studies 1–3 to dive deeper into what perceivers think about affiliation smiles by analyzing the relationship between facial movements and perceiver ratings via a crowd-sourced machine learning

approach. Overall, we expect that reward, affiliation, and dominance smiles will elicit divergent mental representations, lead to different behavioral responses, and have distinct physical features.

Study 1: Narrative Descriptions of Smile Context

Study 1 considers whether the perception of reward, affiliation, and dominance smiles calls to mind social contexts that map onto the theorized functional distinctions between the smile signals. To address this question, we created a set of smile stimuli portrayed by professional actors. Then, we used these stimuli to probe the social situations brought to perceivers' minds by each category of smile, asking perceivers to freely generate descriptions of social contexts in which the smiles might occur. All data, analysis scripts, and stimuli for this and all other studies in this paper—with the exception of participant videos from study 3—are publicly available (https://osf.io/qs35g/).

Study 1: Method

Stimulus Creation and Selection Fifteen professional actors (four African-American men, three African-American women, four White men, four White women) ranging in age from young-adult to middle-aged were recruited from two Midwestern US cities. Two researchers familiar with previous work on the facial movements involved in reward, affiliation, and dominance smiles (Rychlowska et al., 2017) guided the actors through a standardized procedure. Before encoding a given expression, the researchers coached the actors about its physical appearance and the social contexts in which the expression is thought to occur (see Supplemental Materials for stimulus prompts). For each type of smile, the actors were also shown the corresponding computer-generated smile model of Rychlowska and colleagues (see Rychlowska et al., 2017, study 3). In the event that an actor produced a smile that was noticeably different from the smile model, he or she was coached to display greater involvement of the facial actions shown in the model. It should be noted that due to heterogeneity in particular actor encodings of the video stimuli and the fact that the stimuli were encoded by actors rather than by computer-generated avatars, the stimuli do not completely replicate AUs reported by Rychlowska and colleagues.

Examples of smile stimuli from one of the 15 actors are depicted in Fig. 1. In addition to the three types of smiles, actors also encoded a larger set of facial expressions of emotion: surprise, anger, disgust, sadness, and regret (surprise and regret expressions were recorded for other purposes and are not discussed here). All videos show the actor directly facing the camera with the head and shoulders fully visible. Videos were trimmed such that the actors began with approximately

Fig. 1 Examples of smile stimuli. Reward (left), affiliation (center), and dominance (right) smiles are displayed for one of the 15 actors. Although still images are displayed here, participants viewed video versions of these expressions. Actor-produced stimuli were used in studies 1 and 2

1 s of neutral affect and ended at the peak of the full expression. Each video is roughly 3 s long.

Using the full set of 15 actors, we collected data from three different experimental tasks and selected the six actors who produced stimuli with the highest recognition accuracy across tasks. Findings for these three tasks for the entire 15-actor stimulus set are presented in Supplemental Materials. Videos from the six best actors in our stimulus set (one black female, one white female, two black males, two white males) served as stimuli for studies 1 and 2.

Participants A total of 473 participants completed the study on Amazon's Mechanical Turk. We report analyses on responses from participants who passed a simple attention check where they were explicitly directed to select one of four multiple-choice response options ($N = 434$, $M_{age} = 35.77$, 200 female, 225 male, 9 unspecified; 4 Native American, 40 Asian, 33 African-American, 38 Hispanic or Latino, 291 Caucasian, 27 other).

Procedure Participants viewed three smile videos (one of each type) presented in random order. Participants responded by writing a short description of a situation in which they believed the expression might commonly occur.

Dictionary Coding Using results from a pilot study (described in full in Supplemental Materials), we generated three mutually exclusive sets of words (i.e., "dictionaries"), one for each smile type. The dictionaries represent words that are more likely to appear in descriptions of one type of smile than in descriptions of either of the others. Note that the dictionaries are not intended to be exhaustive of all the possible words that are likely to be used in response to reward, affiliation, or dominance smiles. Rather, they represent a subset of the larger set of words called to mind in response to each smile type.

For each text response, if at least one word from any of the three dictionaries appeared in that response, the response was coded as present (i.e., given a 1 in place of a 0) for that dictionary. The three dictionaries are provided below.

Reward dictionary: picture, genuine, close, chuckle, laugh, compliment, flatter, joy, love, sincere, real, true, unfeigned

Affiliation dictionary: pass, customer, listen, service, polite, pretend, bore, acquaintance, unfunny, acknowledge, serve, serving

Dominance dictionary: gossip, flirt, proud, win, deal, smug, best, business, prove, argument, wrong, incorrect

Thematic Response Coding We also coded participants' responses in terms of thematic content. In order to do this, we used the free responses collected in the pilot study to generate a thematic coding scheme for testing in study 1. Based on the responses in the pilot study, we created six thematic codes, two for each smile type. As was the case with the three dictionaries created from the pilot study, the six thematic categories are not exhaustive of the complete set of themes that may be more likely to occur in response to a given smile type. The six thematic categories are provided below.

i) expressing happiness, contentment, cheerfulness, or joy (reward)
ii) occurring during shared laughter or in a humorous context (reward)
iii) expressing fakeness, expressing politeness, being forced/pretending, or hiding other feelings (affiliation)
iv) occurring when walking past, greeting, acknowledging, or meeting someone new (affiliation)
v) expressing contempt, disapproval, sarcasm, or smugness (dominance)
vi) occurring during sexual interest or flirting (dominance)

Three research assistants naïve to the experimental design and research hypotheses coded the text responses to study 1. Coders received the participant text responses in scrambled order and referred to the above criteria to assign values for each of the six codes for every text response. Coders assigned a 1 to a text response when that theme was present and a 0

otherwise. Text responses could receive six 1s, six 0s, or any combination of 1s and 0s. The modal response across the three coders was used in subsequent analyses.

Study 1: Results and Discussion

Analyses for all studies reported in this manuscript were conducted in the R statistical environment (2013). We first tested whether words in each of the three dictionaries were more likely to appear in response to their expected smile type than in response to either of the other two types. Recall that for each of the three dictionaries, a participant's response was coded as a 1 if one or more words from that dictionary were present, and as a 0 otherwise. We fit generalized linear mixed-effects models (i.e., multilevel logistic regression) separately to each of the three variables reflecting presence of reward, affiliation, and dominance dictionary words. We dummy-coded the smile stimulus type for each text response, with the reference group corresponding to the dictionary being tested (e.g., reward smiles were the reference group in models testing for presence of reward dictionary words). We included by-participant and by-actor random intercepts and correlations between random effects.

Results supported our hypotheses. Participants used reward dictionary words more often in response to reward compared to affiliation smiles ($b = -0.85$, 95% CI [$-1.26, -0.45$], $z = -4.11$, $p < .001$, $RR = 0.56$) or dominance smiles ($b = -1.53$, 95% CI [$-2.03, -1.06$], $z = -6.22$, $p < .001$, $RR = 0.31$); participants used affiliation dictionary words more often in response to affiliation compared to reward smiles ($b = -1.29$, 95% CI [$-1.89, -0.74$], $z = -4.43$, $p < .001$, $RR = 0.34$) or dominance smiles ($b = -0.67$, 95% CI [$-1.16, 0.2$], $z = -2.73$, $p = .006$, $RR = 0.59$); participants used dominance dictionary words more often in response to dominance compared to reward smiles ($b = -1.21$, 95% CI [$-1.92, -0.51$], $z = -3.36$, $p < .001$, $RR = 0.30$) or affiliation smiles ($b = -1.34$, 95% CI [$-2.05, -0.64$], $z = -3.73$, $p < .001$, $RR =$

0.30). Dictionary counts for each smile type are displayed in Fig. 2.

We next tested the extent to which the thematic elements were more likely to be present in response to their expected smile type compared to the other types. In order to test this, we ran identical generalized linear mixed-effects models as those for dictionaries. For example, affiliation smiles were coded as the reference group when testing for the presence of codes for fakeness/politeness. With the exception of only two of the twelve tested contrasts, results supported our hypotheses. In general, reward smiles were seen as happy, affiliation smiles were seen as fake/pretending, and dominance smiles were seen as contemptuous and flirtatious. A full table of results is reported in Table 2.

Study 2: Behavioral Responses to Smiles in an Economic Game

Study 1 documented the divergent, function-consistent content of perceivers' mental representations of reward, affiliation, and dominance smiles. Beyond the perceptions that each type of smile engenders, another core component of our social–functional framework (Martin et al., 2017; Niedenthal et al., 2010) is that each of the three smile types serves different social functions. If that is true, then displaying them in a meaningful social context should result in divergent behaviors from perceivers. To test whether perceivers demonstrate distinct behaviors in response to each smile type, participants engaged in a stimulated economic game where they imagined behaviorally responding to a set of facial expressions that included the three types of smiles.

For study 2, we again used stimuli from the six actors chosen for study 1, this time also including the expressions of anger, sadness, and disgust in order to decrease the likelihood of between-smile stimulus contrast effects. In line with previous work documenting greater perceptions of trustworthiness in response to Duchenne compared to non-Duchenne
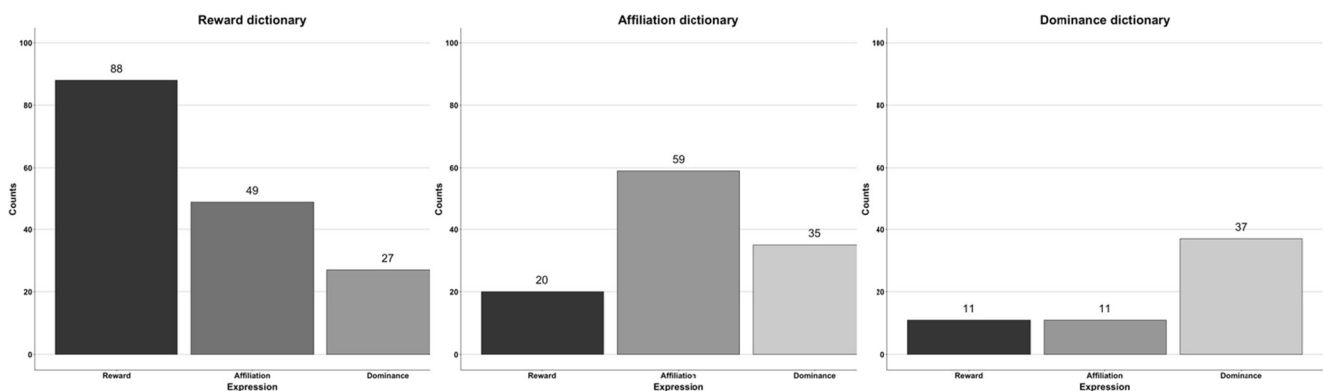
Fig. 2 Reward, affiliation, and dominance dictionary counts between smile types in study 1. We coded each text response for the presence of words from the three smile type dictionaries created from the pilot data. The number of responses for which a dictionary was coded as present, broken down by the type of expression the text was describing. Note that although count data are shown here, summing the number of times a dictionary word was present for each category, responses were modeled via logistic regression

Table 2 Response coding analysis of six *a priori* thematic codes in study 1. Three experimentally naïve research assistants coded all text responses for whether or not they fit into one of the six response types. Responses were coded as a 1 if the response fit the code and coded 0 otherwise. Z tests for the significance of inter-rater reliability as indicated by Fleiss' kappa are reported for each of the six response codes. Responses were allowed to receive 1s for all, some, or none of the codes. The reported results are from generalized linear mixed-effects models estimating the likelihood of a given code in response to the three smile types, where the reference group represents the smile type expected to best fit the code

	vs. reward	vs. affiliation	vs. dominance	Risk ratio	se	95% CI	z	p
Happy/cheerful								
$\kappa = 0.606$; $z = 37.5$; $p < .001$		− 1.57		0.59	0.18	[− 1.93, − 1.23]	− 8.77	< .001
			− 2.41	0.36	0.20	[− 2.82, − 2.02]	− 11.83	< .001
Laughing with								
$\kappa = 0.784$; $z = 48.4$; $p < .001$		− 1.03		0.47	0.22	[− 1.47, − 0.61]	− 4.72	< .001
			− 0.07	0.98	0.19	[− 0.49, 0.29]	− 0.40	.69
Fake/pretending								
$\kappa = 0.586$; $z = 26.2$; $p < .001$	− 1.54			0.30	0.23	[− 2.01, − 1.11]	− 6.69	< .001
		− 1.58		0.30	0.23	[− 2.04, − 1.14]	− 6.89	< .001
Greeting/acknowledging								
$\kappa = 0.347$; $z = 21.4$; $p < .001$	− 1.24			0.34	0.25	[− 1.74, − 0.78]	− 5.07	< .001
		− 0.35		0.76	0.20	[− 0.74, 0.03]	− 1.80	.07
Contempt/sarcasm								
$\kappa = 0.426$; $z = 26.4$; $p < .001$	− 3.26			0.05	0.54	[− 5.07, − 2.71]	− 6.70	< .001
		− 1.51		0.37	0.28	[− 2.14, − 1.00]	− 5.45	< .001
Sexual interest/flirting								
$\kappa = 0.598$; $z = 37.0$; $p < .001$	− 1.37			0.26	0.40	[− 2.15, − 0.58]	− 3.39	< .001
		− 2.13		0.13	0.55	[− 3.17, − 1.08]	− 4.00	< .001

smiles (Gunnery & Ruben, 2016), we expected participants to show behavioral indicators of trust more in response to reward smiles as compared to affiliation or dominance smiles. Similarly, based on our theorizing that affiliation smiles promote positive social encounters, we expected participants to place greater trust in those showing affiliation as compared to dominance smiles. These predictions were derived from theory as well as the results of a pilot study (reported in Supplemental Materials).

Study 2: Method

Participants A total of 189 participants from Amazon's Mechanical Turk completed study 2. We report analyses on responses from participants who passed a simple attention check where they were explicitly directed to select one of four multiple-choice response options ($N = 169$, $M_{age} = 36.24$, 64 female, 103 male, 1 unspecified; 1 Native American, 11 Asian, 21 African-American, 15 Hispanic or Latino, 115 Caucasian, 6 other).

Procedure Participants engaged in a simulated version of a standard one-shot "trust game" (Berg et al., 1995). Participants were invited to imagine that they were "playing a game" with the person displayed in each of six videos—participants were only shown one instance of each of the six

stimulus types: reward smile, affiliation smile, dominance smile, anger expression, disgust expression, sadness expression. Stimuli were presented in random order, and the actor portraying each expression was randomly selected from one of the six actors.

For each video, the participants' goal was to collect as many points as possible. Participants allocated a number of points (out of 100) that they would share with the partner. The number of points was then tripled and given to the partner. For example, if a participant decided to allocate 50 of her 100 points to her partner (i.e., the person in the video for that round), the partner received 150 points. Participants were also told to imagine that partners could return a percentage of tripled points to the participant. Participants reported the percentage of the tripled points they expected to receive back from the partner. Thus, for each of the six videos, participants provided data on (i) how many points they would send, and (ii) what percentage of the tripled points they expected to receive in return.

Study 2: Results

The number of points sent and the number expected in response to the stimuli were highly correlated ($r = .73$). Thus, we created a general trust index by summing the z-scores for the two indicators. We analyzed responses with linear mixed-

effects models, focusing on comparisons between smiles. Specifically, we regressed the trust index on dummy-coded smile type, by-subject random intercepts and slopes for both dummy codes, and by-actor random intercepts and slopes for both dummy codes, and allowed all random effects to correlate. Satterthwaite's degrees of freedom were implemented. The models did not converge. Therefore, we followed established guidelines and fixed the covariance between random effects to zero (Barr et al., 2013). Results of the modified models revealed that participants showed higher trust scores in response to reward compared to affiliation smiles ($b = -0.25$, 95% CI $[-0.43, -0.07]$, $t(154.90) = -2.86$, $p = .005$, Cohen's $D = -0.16$), reward compared to dominance smiles ($b = -0.61$, 95% CI $[-0.88, -0.36]$, $t(3.91) = -5.07$, $p = .008$, Cohen's $D = -0.32$), and affiliation compared to dominance smiles ($b = -0.36$, 95% CI $[-0.63, -0.09]$, $t(4.288) = -2.84$, $p = .043$, Cohen's $D = -0.19$). All Cohen's Ds for mixed-effects models in this manuscript are calculated as per the modifications for mixed-effects models as suggested by Westfall et al. (2014). Results of study 2 are depicted in Fig. 3.

Study 3: Smiles Produced in Response to Hypothetical Social Contexts

Studies 1 and 2 showed that perceivers think about and respond to people differently depending on whether they are displaying a reward, affiliation, or dominance smile. Furthermore, perceivers' thoughts about and behavioral responses to the smiles are consistent with the proposed social function of each smile. For reward, affiliation, and dominance

smiles to serve their proposed social functions, it is not sufficient for perceivers to mentally represent each smile along functional lines. At minimum, for reward, affiliation, and dominance smiles to have their effects, each must involve a distinct set of facial movements that constitutes a clear signal. Thus, study 3 examines the production rather than perception of smiles by testing the extent to which hypothetical social contexts related to the reward, affiliation, and dominance functions of smiles are encoded with distinct sets of facial movements by naïve producers.

Study 3: Method

Participants and Procedure College-aged participants ($N = 121$, 50% female; 8 African or African-American, 22 Asian, 6 Hispanic, 85 Caucasian) in an introductory psychology course took part in a larger stimulus collection study in return for course credit. Participants were not aware that the study involved being videotaped when they signed up for the study in order to prevent them from altering their appearance before the session. However, the recording was explained, and consent was obtained at the beginning of the experimental session itself.

Participants were provided with an instruction sheet with a short conceptual description of each of the three types of smiles along with brief examples of situations in which the smile might occur. When participants felt ready, they looked at the camera with a neutral facial expression and then encoded a smile that corresponded to the signal they felt would appear in the three types of social situations, either focusing on producing a laugh or just a smile. Participants

Fig. 3 Estimated trust in response to each expression type in study 2. This figure displays (number of points sent + points expected)/2 in response to each expression type as general indicator of interpersonal trust. Standard errors are standard error of the mean, adjusted for clustering within actor. Note that for the analyses, we first z-scored points before computing trust scores
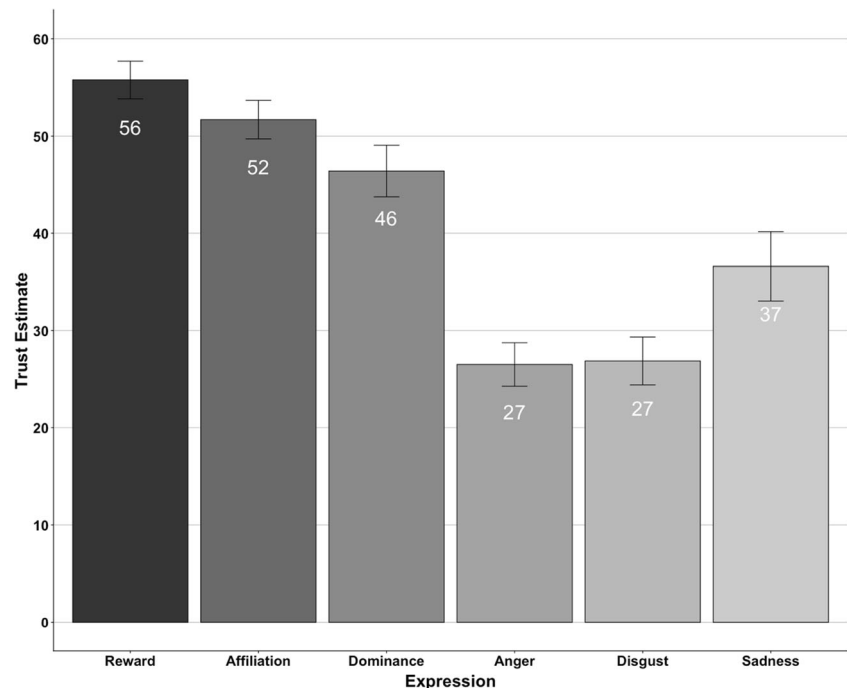
were recorded with their face centered in the frame, from the shoulders up. The laughter data were collected for other purposes and are not reported here. The order of smile type encoding was randomized across participants. For additional information about experimental procedure and participant instructions, see the Supplemental Materials.

FACS Coding A certified FACS coder manually coded a random sample of 1/3 of the videos (via stratified random sampling to maintain class proportions; 4 African or African-American, 10 Asian, 3 Hispanic or Latino, 39 Caucasian). The coder produced presence/absence scores for 21 AUs (see Supplemental Materials for AU presence/absence counts by smile type) at what he considered to be the apex frame of each video. In all subsequent analyses using these data, we analyze only the 76 videos (26 reward, 25 affiliation, 25 dominance) that were FACS coded. In order to ensure that each video had the core component of the smile expression (AU12), we excluded videos in which AU12 was not present; this was the case for only two of the 76 stimuli. Furthermore, AU9 was not present in any of the remaining 74 videos. Thus, given that AUs 9 and 12 showed no variability in the present stimuli, we analyzed the remaining 19 AUs from the original set of 21.

We assessed the reliability of the FACS codes by comparing them against the results of a computer-assisted facial expression analysis using OpenFace (Baltrušaitis et al., 2018). We did this by extracting the average number of frames in which each AU was present and correlating those averages with the human-produced FACS codes. Results of this reliability analysis are reported in Supplemental Materials. Results suggest that reliability of FACS codes may vary between AUs, but that we can place relatively strong confidence in the codes for AUs 1, 2, 4, 6, 14, and 25. We note that although a reliability analysis of this sort is not a substitute for direct comparison of human ratings, it nonetheless provides some indication of which AUs may have been more reliably detected.

Study 3: Results and Discussion

In order to test which AUs were more likely to be present in one smile type compared to the others, we fit a series of logistic regression models. We first created three new variables, where each variable coded the comparison between one smile type versus the other two (e.g., for the reward comparison, reward smiles were coded as 1, whereas affiliation and dominance smiles were coded as 0). We then fit generalized linear mixed-effects models (multilevel logistic regression) separately for the three variables reflecting the comparisons. In these models, we separately regressed each comparison code on each of the 19 AUs, including a by-participant random intercept.

AUs 16, 25, and 26 were more likely to be present in reward smiles as compared to the other two smile types (all $ps < .02$). Furthermore, AU14 was less likely to be present in reward smiles as compared to the other two types ($p < .01$). Similarly, AUs 14 and 24 were more likely to be present in affiliation smiles as compared to the other two types of smiles ($ps < .02$), and the presence of AU 25 was less likely for affiliation smiles compared to the other types ($p = .02$). For dominance smiles, no significant differences were detected. A full table of results from these analyses is presented in Table 3.

In summary, study 3 demonstrated that when experimentally naïve participants are prompted with social contexts that match the social functions of smiles, they largely use a distinct set of AUs to signal each context. The present findings for AUs 14 and 24 replicate aspects of previous work using reverse correlation to document AUs associated with mental representations of each type of smile, where AUs 14 and 24 were present in the mental prototype models of affiliation smiles developed by Rychlowska et al. (2017). The overlap between the present findings and those of Rychlowska et al. (2017) strongly suggests that affiliation smiling is characterized by the presence of AUs 14 and 24.

The present findings also uncover previously undocumented action units relevant to reward smiles. The presence of AUs 25 and 26 in reward smiles is consistent with the action units frequently found in smiles associated with the experience of positive affect (e.g., Crivelli et al., 2015). AU16, which is activated across cultures during posed expressions of happiness (Cordaro et al., 2018), was also displayed more during reward smiles. Taken together, findings from the present study suggest that whereas mouth opening is more likely to be present in reward smiles, smile controls—for example, pressing the lips together and tightening the lip corners—are more likely to be present in affiliation smiles.

Although no significant differences were detected for dominance smiles in the present study, we believe a true lack of physical differences in dominance smiles to be unlikely. Given the strength of the findings for dominance smiles in studies 1 and 2 and other work documenting considerable differences in biological responses to dominance smiles compared to other smiles (Martin et al., 2018), we suspect that the lack of findings in the present work is due to methodological concerns. Specifically, previous work documents the importance of asymmetrical AU12 activation in dominance smiles (Rychlowska et al., 2017), which was not quantified here. Furthermore, since head position (i.e., yaw, pitch, and roll) contains important information related to dominance (e.g., Witkower & Tracy, 2019), future work on the physical features of dominance smiling should include more nuanced measures such as head position and expression asymmetry.

In summary, results of the present studies complement and extend previous findings (Rychlowska et al., 2017): reward smiles are characterized by AUs 6, 25, and 26, and the relative

Table 3 Multilevel logistic regression results from study 3. Results presented in this table reflect findings when predicting smile type comparison (e.g., reward comparison: reward = 1, affiliation = 0, dominance = 0) from each of 19 AUs. Models were linear mixed-effects models with by-participant random intercepts. In the event that an AU was absent from every stimulus in either the comparison group or the other two groups, estimates could not be derived for that particular AU. This is because logistic regression breaks down under cases of complete class separation. Two codes are placed in this table to indicate AUs for which that was the case: [a]not present in any reference group stims. [b]Not present in either of the two comparison groups. The interested reader is also pointed to the Supplemental Materials for a table of raw counts of AU presence for each smile type

Reference	AU	Estimate	se	z	p
Reward	AU1	− 0.45	1.18	− 0.38	.70
	AU2	0.69	1.44	0.48	.63
	AU4[a]				
	AU5[b]				
	AU6	0.13	0.57	0.24	.81
	AU7	− 1.00	1.12	− 0.89	.37
	AU10	0.66	0.54	1.23	.22
	AU13[a]				
	AU14	− 3.22	1.06	− 3.04	< .01
	AU15	− 0.02	1.25	− 0.02	.99
	AU16	1.60	0.63	2.54	.01
	AU17	− 1.39	1.10	− 1.26	.21
	AU20	0.69	1.44	0.48	.63
	AU23	− 0.45	1.18	− 0.38	.70
	AU24[a]				
	AU25	2.29	0.62	3.68	< .01
	AU26	1.85	0.67	2.76	.01
	AU27[b]				
	AU43[b]				
Affiliation	AU1	0.71	1.03	0.69	.49
	AU2[a]				
	AU4[a]				
	AU5[a]				
	AU6	0.13	0.57	0.24	.81
	AU7	− 1.00	1.12	− 0.89	.37
	AU10	− 1.36	0.68	− 1.99	.047
	AU13	− 0.02	1.25	− 0.02	.99
	AU14	1.65	0.53	3.09	< .01
	AU15[a]				
	AU16	− 0.75	0.70	− 1.07	.29
	AU17	0.76	0.75	1.01	.31
	AU20	0.69	1.44	0.48	.63
	AU23	1.88	1.18	1.59	.11
	AU24	1.44	0.53	2.70	.01
	AU25	− 1.21	0.52	− 2.33	.02
	AU26	− 0.63	0.71	− 0.89	.37
	AU27[a]				
	AU43[a]				
Dominance	AU1	− 0.38	1.18	− 0.32	.75
	AU2	0.76	1.44	0.53	.60
	AU4[b]				
	AU5[a]				
	AU6	− 0.27	0.56	− 0.48	.63
	AU7	1.57	0.91	1.73	.08
	AU10	0.46	0.55	0.84	.40
	AU13	1.49	1.25	1.19	.23
	AU14	0.42	0.51	0.81	.42
	AU15	1.49	1.25	1.19	.23
	AU16	− 1.25	0.81	− 1.54	.12
	AU17	0.25	0.78	0.32	.75
	AU20[a]				
	AU23[a]				
	AU24	0.71	0.52	1.35	.18
	AU25	− 0.83	0.51	− 1.64	.10
	AU26	− 1.98	1.07	− 1.85	.06
	AU27[a]				
	AU43[a]				

absence (compared to affiliation or dominance smiles) of smile control muscles such as AU14, and affiliation smiles are characterized by AUs 14 and 24. The present study neither supports nor extends previous findings of the presence of AUs 9 and 10, and asymmetrical AU12 in dominance smiling, likely due to factors relating to the social context in which participants encoded these smiles.

Study 4: Crowd-Sourcing Labels for the Affiliation Smile

Studies 1–3 provide strong evidence that reward, affiliation, and dominance smiles are distinct social signals. However, across experiments, there is inconsistency in the strength of evidence for the affiliation smile, with some measured responses to affiliation smiles being similar to responses to reward smiles (Rychlowska et al., 2017, study 3; this manuscript, Supplemental Materials for study 1). One challenge for the investigation of affiliation smiles is that the intention of the smile—its social function—is positive or prosocial, and the function of the reward smile is *also* positive. Significant semantic overlap therefore arises when explicit language-based descriptions and reports are used to investigate these smiles.

A second challenge is that the concept of the affiliation smile is not a lay concept about which people communicate with language. Indeed, the lay concept of the word "smile" is an expression of happiness or positive emotion (LaFrance, 2011; Nettle, 2006), mapping well onto what we classify as a reward smile.[1] Therefore, since the lay concept of a "smile" means someone is happy, then if someone is smiling but not happy, the smile is de facto "fake." However, from our social–functional perspective, all smiles are "true" insofar as they are honest and effective social signals that each solves different social tasks, which can be the case even without people having a consensual way to talk about a given type of smile (Martin et al., 2018). Given that naive participants likely do not have consensual labels for a smile for which there is no lay concept—*affiliation smiles*—we cannot be sure that the language we use to label these smiles, as researchers, is the best way to probe how perceivers think about this signal. Study 4 attempts to address this communication gap between researchers and participants.

We first derived six candidate labels for assessment in this study: reassuring, agreeable, trustworthy, polite, fake, and happy. "Reassuring" and "agreeable" are relatively simple words that represent the functional claim that affiliation smiles signal non-threat (the word "appeasing" is not as commonly used or understood by lay people). "Trustworthy" was

included in order to better understand the place of affiliation smiles in behavioral settings similar to study 2. The inclusion of "fake" and "polite" was based on findings from study 1 of this manuscript as well as considerations discussed above. And finally, the label "happy" was used in order to probe the similarity previously seen in some responses to affiliation and reward smiles.

With the six labels selected, we designed an online machine learning task to take participant judgments on a set of affiliation smiles as input and return a ranked list of those same stimuli as output. To do this, we randomly selected 103 of the affiliation smiles produced by participants in study 3 and used them as stimuli in six between-subjects labeling tasks. We specifically did not include reward and dominance smiles in order to avoid stimulus comparison effects. After obtaining the rankings for the 103 stimuli on each of the six labels, we then conducted two general types of analyses: within-label analyses to consider how AU presence relates to labels and between-label/between-AU analyses to consider how labeling and AUs relate to one another.

Study 4: Method

Participants Online participants, recruited through Amazon's Mechanical Turk, took part in the study in exchange for $.75–$1.00. In order to participate, an individual had to currently reside in the US, have completed at least 100 HITs on Mechanical Turk, and have at least a 75% approval rate. A total of 3039 participants began the study ($N_{happy} = 516$, $N_{fake} = 481$, $N_{agreeable} = 503$, $N_{polite} = 501$, $N_{trustworthy} = 527$, $N_{reassuring} = 511$) with 2100 of them fully completing it. Data collection was terminated when the six between-subjects label conditions were each fully completed by 350 participants. In total, including trials from participants who did not finish the study, 55,351 responses were collected ($N_{happier} = 9423$, $N_{fake} = 9102$, $N_{agreeable} = 9415$, $N_{polite} = 9200$, $N_{trustworthy} = 9223$, $N_{reassuring} = 9078$). Of the 2100 participants who completed the study, 1850 provided demographic information (49% male, 51% female; 21% in their 20s, 39% in their 30s, 20% in their 40s, 11% in their 50s, 7% in their 60s, 1% in their 70s; 80% Caucasian, 10% African-American, 9% Asian, 1% multiracial or other).

Smile Stimuli We randomly selected 103 of the affiliation smiles (5 African or African-American, 18 Asian, 4 Hispanic or Latino, 76 Caucasian) from the full set of affiliation smiles produced by the 121 participants in study 3. We aimed to have reliable data for 100 stimuli, so we included a few more in case data from some of the stimuli were clear outliers. We chose to use the affiliation smile videos from study 3 rather than those from studies 1 and 2 because the videos from study 3 were more numerous and considerably more variable. Specific sources of variability include greater

---

[1] We also note that a lay concept for the dominance smile exists in the concept of a "smirk" which is defined as a smug or condescending smile (Merriam-Webster, 2020).

ethnic/racial diversity as well as participants' varying degrees of ability to produce clear affiliation smiles on command and their comfort in front of a camera. Thus, the stimuli in this study are constrained in their social–contextual instructions (all were produced with affiliation directions) but highly variable in important factors that might influence smile perception (e.g., race/ethnicity, gender). The constrained variability of these smiles allows for a nuanced test of the facial actions that predict concept-based labels related to the signal of affiliation.

Procedure After participants read consent information, they were assigned to one of six label conditions and directed to the task page. Participants were only allowed to complete the study once, ensuring that label condition was fully between-subjects. On each trial, participants viewed two videos side-by-side (videos could be shown more than once) with one of six labels presented directly above the two videos. The algorithm used to determine which two videos appeared on a given trial is explained below (the NEXT platform). Depending on label condition, participants were directed to select "which of the two expressions looks": (i) happier, (ii) more fake, (iii) more agreeable, (iv) more polite, (v) more trustworthy, (vi) more reassuring. Participants entered their responses by either clicking directly on the video or by pressing the right/left arrow key on their keyboard. The entire session lasted approximately 5 min, and participants were given 25 trials.

The NEXT Platform We used the online machine learning platform NEXT (Jamieson et al., 2015) to collect participant responses. NEXT is an open-source software system designed to make data collection easier for users from diverse academic backgrounds (Sievert et al., 2017). From the suite of algorithms provided by NEXT, we chose to use the KLUCB algorithm (Tanczos et al., 2017). The KLUCB algorithm is designed to find the stimuli with the highest means in the multi-armed bandit problem and is currently being used to find the funniest captions in The New Yorker (TNY) Cartoon Caption Contest from NYC's list of user submitted captions (Tanczos et al., 2017). For our application, the KLUCB algorithm takes comparison judgments between two stimuli and uses those comparisons to estimate a score on that judgment between 0 and 1 with a corresponding confidence interval for each stimulus. The innovation behind this algorithm is that instead of sampling randomly, it adaptively determines which stimuli to present in each trial as a function of the scores and confidence intervals as estimated from all data collected up to that point in time.

A handful of technical points regarding our usage of the KLUCB algorithm as implemented in NEXT are worth noting here. First, the KLUCB algorithm uses all participant responses to produce rankings and scores for each of the six label conditions—even those from participants who did not fully complete the study. To help ensure the validity of the KLUCB rankings, we used random sampling on about 15% of the trials to generate scores and confidence intervals. For each of the six label conditions, we conducted basic verification to make sure KLUCB and random sampling both performed as expected. Verification results confirmed good performance: for each of the six label conditions, the 99% confidence interval for KLUCB and the 99% confidence interval for random sampling overlapped at chance (0.99*0.99 = 98% of the time) levels or better. Note that in both KLUCB trials and randomly sampled trials, sampling was conducted with replacement which means that participants could be presented with the same stimulus on multiple trials.

FACS Codes The 103 stimuli used in this study come from participant videos produced in study 3 of the present manuscript. Since only 25 affiliation smiles were FACS coded in study 3, FACS codes are available for only 25 of the 103 stimuli used in the present study. All analyses in study 4 are conducted on the 25 affiliation videos with FACS codes in order to make direct comparisons to the findings from study 3. Note that only analyzing a subset of the 103 stimuli decreases power to detect an effect but is likely to have little effect on the type I error rate. Thus, although the sample for analysis is smaller than the complete set of stimuli presented to participants, we do not believe analyses on this smaller set of stimuli will significantly increase spurious findings.

We excluded from analysis any AU that was either absent from all 25 videos or present in all 25 videos. This is because, since there is no between-stimulus variability on those action units, including them in statistical analysis yields no added value. Under this condition, 13 FACS codes were included in analyses for the present study: AU1, AU6, AU7, AU10, AU13, AU14, AU16, AU17, AU20, AU23, AU24, AU25, AU26.

Semantic Similarity Analysis via Singular Value Decomposition In order to analyze between-label/between-AU semantic similarity, we adopted an analytic approach to analyze latent semantic structure called singular value decomposition (SVD; Golub & Reinsch, 1971), which is a technique used in many disciplines, including computer vision and natural language processing. SVD is a dimensionality reduction procedure that can be loosely thought of as a two-way factor analysis where a single matrix is decomposed into three matrices: left singular vectors, right singular vectors, and a set of non-negative singular values. When used in psychology, the ultimate goal of an SVD analysis is usually to model how two sets of features (judgments, concepts, ratings, etc.) jointly map into a latent semantic space in order to describe the relationships between those features (for

applications in psychology and an accessible tutorial, see Kosinski et al., 2016).

As an example, a researcher could use SVD in the following way to represent the semantic content of a set of 500 different books. First, the researcher compiles a (w × b) word frequency matrix where each row (w) is a word and each column (b) is a book. After pre-processing this matrix, SVD is applied to the matrix in order to decompose it into three other matrices. In addition to the matrix of singular values, a second matrix maps words to a shared latent space and a third maps books to that same shared latent semantic space. If the 500 books were drawn from three general domains (such as accounting, science fiction, and astronomy), SVD would likely return three latent clusters corresponding to these three general domains. Furthermore, since science fiction and astronomy share more words in common (i.e., planet, rocket, star) than do accounting and science fiction or accounting and astronomy, the two clusters representing science fiction and astronomy would be closer to each other in semantic space than either are to accounting. In this way, researchers can use SVD to jointly map two sets of features (e.g., words and books) into a shared, lower-dimensional, semantic space.

In the present analysis, we treat the presence of facial actions as "words" and label categories as "books" in the matrix input into our SVD analysis. The goal of the analysis is to map facial movements (i.e., *words*) and label categories (i.e., *books*) into a shared latent semantic space. Mapping into a shared space allows for comparison of the semantic similarity of labels and the semantic similarity of AUs, both with reference to a common space. All SVD analyses in this paper were conducted with the function "svd" from the base package in R (Team, 2013).

Study 4: Results and Discussion

Summary of Analyses Study 4 was designed to mathematically represent how perceivers think about affiliation smiles (i.e., the labels that observers apply to such smiles) and how particular facial actions guide those labels. Based on study design, there are two types of evidence that a particular label (e.g., "fake") is associated with affiliation smiles. First, if a label constitutes part of the semantic content of the affiliation smile, *within-label* correlations should show that higher scores on that label are positively correlated with the presence of AUs previously documented as being involved in affiliation smiling (AU 14 and 24). Conversely, if a label does not constitute part of the semantic content of the affiliation smile, *within-label* scores on that label should be negatively correlated with those same features (AU 14 or 24). Second, when labels and AUs are plotted in a shared semantic space, labels associated with affiliation smiling should occupy similar positions in semantic space as do AUs involved in affiliation smiling (AUs 14 and 24). To test these predictions, we

conducted two general types of analyses on the six label scores: (i) *within-label analyses* correlate a given label score with the presence of a given AU and (ii) *between-label analyses* use SVD to map the semantic space common to both labels and AUs.

Within-Label Analyses We tested the extent to which the presence of 13 AUs was correlated with scores for each of the six labels. Both happy and polite labels were positively correlated with the presence of AU6 (ps = .03 & .01, respectively). Polite and reassuring labels were positively correlated with AU23 (ps = .03). Polite and trustworthy labels were positively correlated with AU25 (ps = .03 & .01, respectively). And, most directly relevant to affiliation smiles, trustworthy labels were *negatively* related to the presence of AU1 and AU14 (ps = .02 & .03, respectively). The negative relationship between the presence of AU14 and trustworthiness judgments suggests that insofar as AU14 is a core component of affiliation smiling, affiliation smilers may not be perceived as highly trustworthy. This conclusion is in line with previous work showing that Duchenne smilers (AU6 + AU12) are perceived as more trustworthy than non-Duchenne smilers (Gunnery & Ruben, 2016). A full table of correlations between the scores for the six labels with the 13 AUs is presented in Table 4.

Between-Label Analyses We next examined how labels are related to one another, that is, the semantic similarity of the six labels tested here. In order to do this, we constructed a weighted AU × label matrix. We created six label columns representing the 25 stimuli with FACS codes and their six label scores. We weighted the contribution of each stimulus to each label column by multiplying the AU presence/absence (0/1) value for that stimulus by the score on that label (standardized within label) for that stimulus. For example, if a stimulus had AU6 present and received a happiness score of 0.86, then that stimulus would contribute 1*0.86 to the AU6 cell of the happiness label column. After weighting the contribution of each stimulus to each label column based on the presence/absence of a given AU, we then summed weighted scores across all stimuli for every cell in the 13 × 6, AU × label matrix. This 13 × 6, weighted AU sums × label matrix served as the input for all SVD analyses in this paper.

We decomposed the weighted AU × label matrix into three matrices via SVD. Based on visual inspection of the scree plot of singular values using the elbow method, we retained two singular values. To further quantify between-label semantic similarity, we calculated the Euclidean distance between each label in the latent 2D semantic space. Standardized Euclidean distances (between 0 and 1) are displayed in the form of a dissimilarity matrix between all labels (Fig. 4, panel b). Visual inspection of the dissimilarity matrix and the 2D output from the SVD (Fig. 4, panel a, blue points) reveals that agreeable, reassuring, and happy labels group together, polite and

Table 4 Correlations between label scores and AU presence for six labels and 13 AUs in study 4. A sample size of 25 applies for all correlations reported in this table

Label	AU	r	95% CI	t	se	p
Happier	AU1	− 0.14	[− 0.5, 0.27]	− 0.66	0.21	.52
	AU6	0.44	[0.06, 0.71]	2.37	0.19	.03
	AU7	0.30	[− 0.11, 0.62]	1.52	0.20	.14
	AU10	0.12	[− 0.29, 0.49]	0.56	0.21	.58
	AU13	0.30	[− 0.11, 0.62]	1.52	0.20	.14
	AU14	0.08	[− 0.32, 0.46]	0.39	0.21	.70
	AU16	0.08	[− 0.33, 0.46]	0.38	0.21	.71
	AU17	− 0.10	[− 0.47, 0.31]	− 0.47	0.21	.64
	AU20	− 0.22	[− 0.57, 0.19]	− 1.09	0.20	.29
	AU23	0.32	[− 0.09, 0.63]	1.61	0.20	.12
	AU24	0.04	[− 0.36, 0.43]	0.19	0.21	.85
	AU25	0.29	[− 0.12, 0.61]	1.44	0.20	.16
	AU26	0.06	[− 0.34, 0.44]	0.28	0.21	.78
Fake	AU1	0.00	[− 0.4, 0.39]	− 0.01	0.21	.99
	AU6	− 0.18	[− 0.54, 0.23]	− 0.90	0.20	.38
	AU7	0.04	[− 0.36, 0.42]	0.17	0.21	.87
	AU10	− 0.04	[− 0.43, 0.36]	− 0.18	0.21	.86
	AU13	0.04	[− 0.36, 0.42]	0.17	0.21	.87
	AU14	0.12	[− 0.29, 0.49]	0.58	0.21	.57
	AU16	0.03	[− 0.37, 0.42]	0.14	0.21	.89
	AU17	− 0.25	[− 0.59, 0.16]	− 1.24	0.20	.23
	AU20	0.16	[− 0.25, 0.53]	0.80	0.21	.43
	AU23	− 0.30	[− 0.62, 0.1]	− 1.52	0.20	.14
	AU24	0.11	[− 0.3, 0.48]	0.53	0.21	.60
	AU25	− 0.15	[− 0.51, 0.26]	− 0.71	0.21	.49
	AU26	0.22	[− 0.19, 0.56]	1.07	0.20	.30
Agreeable	AU1	− 0.24	[− 0.58, 0.17]	− 1.17	0.20	.25
	AU6	0.28	[− 0.12, 0.61]	1.42	0.20	.17
	AU7	0.09	[− 0.31, 0.47]	0.44	0.21	.66
	AU10	0.23	[− 0.18, 0.57]	1.12	0.20	.27
	AU13	0.09	[− 0.31, 0.47]	0.44	0.21	.66
	AU14	0.07	[− 0.34, 0.45]	0.31	0.21	.76
	AU16	0.10	[− 0.3, 0.48]	0.50	0.21	.62
	AU17	− 0.05	[− 0.44, 0.35]	− 0.25	0.21	.80
	AU20	− 0.02	[− 0.41, 0.38]	− 0.11	0.21	.91
	AU23	0.26	[− 0.15, 0.59]	1.29	0.20	.21
	AU24	− 0.28	[− 0.61, 0.12]	− 1.42	0.20	.17
	AU25	0.26	[− 0.15, 0.6]	1.32	0.20	.20
	AU26	0.03	[− 0.37, 0.42]	0.15	0.21	.88
Polite	AU1	− 0.19	[− 0.55, 0.22]	− 0.94	0.20	.36
	AU6	0.50	[0.13, 0.75]	2.74	0.18	.01
	AU7	0.21	[− 0.21, 0.56]	1.01	0.20	.32
	AU10	0.16	[− 0.25, 0.52]	0.79	0.21	.44
	AU13	0.21	[− 0.21, 0.56]	1.01	0.20	.32
	AU14	− 0.19	[− 0.55, 0.22]	− 0.95	0.20	.35
	AU16	0.19	[− 0.22, 0.54]	0.92	0.20	.37
	AU17	− 0.29	[− 0.61, 0.12]	− 1.43	0.20	.17
	AU20	0.11	[− 0.3, 0.48]	0.51	0.21	.62
	AU23	0.44	[0.06, 0.71]	2.37	0.19	.03
	AU24	− 0.11	[− 0.48, 0.3]	− 0.51	0.21	.61
	AU25	0.43	[0.04, 0.71]	2.29	0.19	.03
	AU26	0.12	[− 0.29, 0.49]	0.58	0.21	.57
Trustworthy	AU1	− 0.47	[− 0.73, − 0.09]	− 2.54	0.18	.02
	AU6	0.20	[− 0.22, 0.55]	0.96	0.20	.35
	AU7	0.08	[− 0.32, 0.46]	0.40	0.21	.69
	AU10	0.19	[− 0.23, 0.54]	0.91	0.20	.37
	AU13	0.08	[− 0.32, 0.46]	0.40	0.21	.69
	AU14	− 0.44	[− 0.71, − 0.06]	− 2.37	0.19	.03
	AU16	0.23	[− 0.18, 0.58]	1.15	0.20	.26
	AU17	− 0.10	[− 0.48, 0.31]	− 0.49	0.21	.63
	AU20	0.11	[− 0.3, 0.48]	0.53	0.21	.60
	AU23	0.25	[− 0.16, 0.59]	1.26	0.20	.22
	AU24	− 0.07	[− 0.45, 0.34]	− 0.32	0.21	.76

Table 4 (continued)

Label	AU	r	95% CI	t	se	p
	AU25	0.54	[0.18, 0.77]	3.04	0.18	.01
	AU26	0.19	[− 0.22, 0.54]	0.92	0.20	.37
Reassuring	AU1	− 0.22	[− 0.56, 0.19]	− 1.07	0.20	.30
	AU6	0.21	[− 0.21, 0.56]	1.01	0.20	.32
	AU7	0.22	[− 0.19, 0.57]	1.11	0.20	.28
	AU10	0.09	[− 0.32, 0.47]	0.42	0.21	.68
	AU13	0.22	[− 0.19, 0.57]	1.11	0.20	.28
	AU14	0.08	[− 0.32, 0.46]	0.39	0.21	.70
	AU16	0.23	[− 0.18, 0.57]	1.12	0.20	.28
	AU17	− 0.34	[− 0.65, 0.06]	− 1.75	0.20	.09
	AU20	− 0.02	[− 0.41, 0.38]	− 0.07	0.21	.94
	AU23	0.44	[0.06, 0.71]	2.38	0.19	.03
	AU24	− 0.30	[− 0.62, 0.11]	− 1.50	0.20	.15
	AU25	0.08	[− 0.32, 0.46]	0.39	0.21	.70
	AU26	0.20	[− 0.21, 0.55]	0.99	0.20	.33

trustworthy are somewhat distant from the others, and fake is considerably dissimilar from all others.

We next assessed between-AU semantic similarity using an identical approach to the between-label analysis. We obtained the facial action similarity matrix by multiplying the label similarity matrix by the original weighted facial action × label matrix (see Kosinski et al., 2016). Using this AU semantic similarity matrix in two dimensions, we calculated the Euclidean distance between each AU in the latent semantic space. As before, standardized Euclidean distances (between 0 and 1) are displayed in the form of a dissimilarity matrix between all AUs (Fig. 4, panel c). Visual inspection of the dissimilarity matrix and the 2D visualization of the SVD results (Fig. 4, panel a, red points) reveals a tight clustering of AUs, most notably AUs 1, 14, 17, and 24. Comparing the 2D visualizations for AUs and labels in the shared semantic space, we see that the findings make intuitive sense. For example, perceptions of happiness are known to be related to the presence of AU6, and the happy label and AU6 occupy similar positions in the latent semantic space. Given that AUs 14 and 24 occupy a similar position to that of the fake label, the present findings suggest that insofar as affiliation smiles are characterized by the presence of AUs 14 and 24, they are likely also to be labeled as fake. This finding, in conjunction with the findings from study 1, suggests that affiliation smilers are conceptualized as fake and/or pretending.

General Discussion

Across 4 studies spanning production- and perception-based tasks, we find evidence for distinct mental and physical representations of reward, affiliation, and dominance smiles. Participants in these studies use function-consistent words to describe each type of smile, enact different behaviors in response to them, and employ distinct facial movements to encode social contexts related to each of the three social functions.

One notable finding is that affiliation smilers are judged as fake/pretending (studies 1 & 4). However, study 2 suggests that despite being labeled as fake, affiliation smiles nonetheless elicit greater trust behaviors in perceivers than do dominance smiles and some other classes of emotional facial expressions. How do we reconcile affiliation smiles' positive social outcomes (i.e., eliciting trust behaviors) with the more negative label (i.e., being fake) that is associated with the physical features of affiliation smiling (AU14)?

A simple answer, which awaits further study, is that seeing a smile of affiliation out of context looks fake in that it seems not to contain signs of positive emotions that lay concepts of smiles demand "true" smiles possess. Without corresponding lay concepts and consensual labels, affiliation smiles seen out of context may be called fake. When displayed in an appropriate social context such as that used in study 2, or in the study by Martin et al. (2018), the signal may serve its positive function. Indeed, when occupants of a crowded elevator display smiles of affiliation to a person about to enter the elevator, the smiles serve as honest signals of non-threat. The person does not board the elevator worrying that the smiles are "fake" in the sense of not arising from underlying happiness. In fact, smiles of reward from occupants of a crowded elevator might well be seen as inappropriate and off-putting.

Although promising, the present findings are constrained in how far they can be applied in support of our social–functional framework (Martin et al., 2017; Niedenthal et al., 2010). First, although the present research constitutes an improvement upon previous work in which smiling stimuli were computer-generated and viewed apart from any social context, the present stimuli are still somewhat removed from the richness of interpersonal encounters. To the extent that we embedded smiles in an artificial social context (study 2) and investigated the social contexts brought to mind when perceiving each
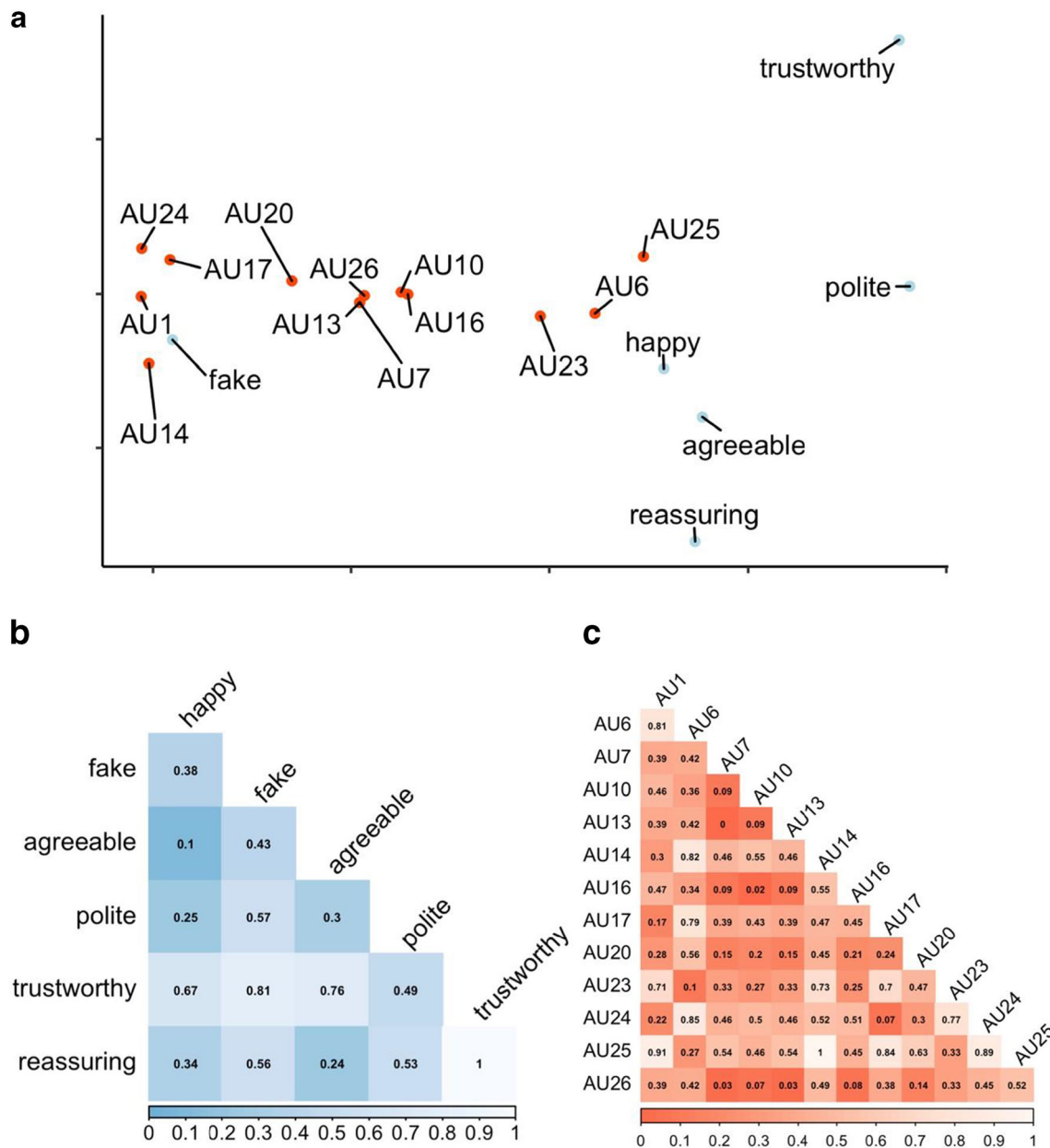
Fig. 4 Between-label and between–facial movement visualizations of semantic similarity. SVD was conducted on a 13 × 6 weighted AU × label matrix. After retaining two dimensions, we plot the labels and AUs in the 2D shared semantic space (panel **a**). To further aid the reader in interpreting the relationships between different labels and between different AUs, dissimilarity matrices via standardized Euclidean distances (minimum = 0, maximum = 1) are plotted for labels (panel **b**) and AUs (panel **c**). In both panels **b** and **c**, lighter colors indicate less dissimilarity (i.e., greater similarity); conversely, darker colors indicate greater dissimilarity (i.e., less similarity). Note that AUs 14 and 24, key components of affiliation smiles, are both relatively close to the "fake" label

smile type (study 1), we have preliminary information about the social–contextual meaning of the smile types. To date, few studies attempt to investigate the effects of reward, affiliation, and dominance smiles during in vivo social interactions (Martin et al., 2018; Rychlowska et al., under review). A necessary avenue for further investigation of these smile types is their consideration within the context of "real-life" social situations (for parallel work on natural laughter in social contexts, see Wood, under review). Work of this sort will undoubtedly help further clarify the social functions of these smiles as well as elucidate whether, and to what extent, the meaning and social consequence of each smile is contingent upon the interpersonal context.

Second, the results from study 2 are reports of how perceivers *expect* they would behave, not actual behavioral data. At minimum, the present results indicate that perceivers *think* they would behave differently in response to each type of smile, which suggests that each is mentally represented as

communicating distinct social information. Although only sparse data exists on the topic of whether and how different kinds of smiles embedded in economic games impact trust behaviors, extant evidence suggests that smile type does likely influence trust (Centorrino et al., 2015; Mehu et al., 2007; Reed et al., 2012). Future studies will be necessary in order to fully document how smile type (particularly beyond the Duchenne/non-Duchenne distinction) influences interpersonal behaviors. Since facial expression perception is influenced by perceivers' race and ethnicity (Friesen et al., 2019), future studies should be particularly careful to consider the potential ways in which the race and ethnicity of the smile perceiver or producer affect interpersonal behaviors.

Findings from the studies presented in this manuscript provide evidence that perceivers hold divergent mental representations of reward, affiliation, and dominance smiles (studies 1, 2, & 4) and describe what these smiles might look like when elicited in relevant contexts (study 3). It remains to be determined, however, whether and how people actually use reward, affiliation, and dominance smiles in daily living and how these smiles are perceived within a dynamic social context. Although we leave these questions open at present, it is our position that documenting how people produce and perceive reward, affiliation, and dominance smiles in a laboratory setting is a necessary precondition for fully understanding how these signals influence the flow of social events if and when they are encountered in the real world. As such, the present manuscript builds upon a long history of approaches to carving up the variability inherent to smiles, bringing the science of the smile one step closer to the ultimate goal of understanding how people make sense of such a contextually and physically heterogeneous expression.

Additional Information

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33*(1), 17–34. https://doi.org/10.1007/s10919-008-0059-5.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition.*

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

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. https://doi.org/10.1006/game.1995.1027.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. *Evolution and Human Behavior, 36*(1), 8–16.

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*(1), 75–93. https://doi.org/10.1037/emo0000302.

Crivelli, C., Carrera, P., & Fernández-Dols, J.-M. (2015). Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evolution and Human Behavior, 36*(1), 52–58. https://doi.org/10.1016/j.evolhumbehav.2014.08.009.

Duchenne, G. B. (1876). Mécanisme de la physionomie humaine: où, Analyse électro-physiologique de l'expression des passions.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (revised ed.). New York, N.Y: WW Norton & Company.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology. II. *Journal of Personality and Social Psychology, 58*(2), 342–353 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2319446.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system (FACS): A technique for the measurement of facial action. In. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6*(4), 238–252. https://doi.org/10.1007/bf00987191.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54*(3), 414–420 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3361418.

Fischer, A. H., & Manstead, A. S. R. (2008). Social functions of emotion. *Handbook of emotions, 3*, 456–468.

Fridlund, A. J. (1991). Evolution and facial action in reflex, social motive, and paralanguage. *Biological Psychology, 32*(1), 3–100. https://doi.org/10.1016/0301-0511(91)90003-Y.

Friesen, J. P., Kawakami, K., Vingilis-Jaremko, L., Caprara, R., Sidhu, D. M., Williams, A., Hugenberg, K., Rodríguez-Bailón, R., Cañadas, E., & Niedenthal, P. (2019). Perceiving happiness in an intergroup context: The role of race and attention to the eyes in differentiating between true and false smiles. *Journal of Personality and Social Psychology, 116*(3), 375–395.

Golub, G. H., & Reinsch, C. (1971). Singular value decomposition and least squares solutions. *Linear Algebra, 186,* 134–151.

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2012). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior, 37*(1), 29–41. https://doi.org/10.1007/s10919-012-0139-4.

Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition & Emotion, 30*(3), 501–515. https://doi.org/10.1080/02699931.2015.1018817.

Harris, C., & Alvarado, N. (2005). Facial expressions, smile types, and self-report during humour, tickle, and pain. *Cognition & Emotion, 19*(5), 655–669. https://doi.org/10.1080/02699930441000472.

Hoque, M., Morency, L.-P., & Picard, R. W. (2011). Are you friendly or just polite?–Analysis of smiles in spontaneous face-to-face interactions. *6974,* 135-144. https://doi.org/10.1007/978-3-642-24600-5_17.

Jamieson, K. G., Jain, L., Fernandez, C., Glattard, N., & Nowak, R. (2015). NEXT: A system for real-world development, evaluation, and application of active learning. *Advances in Neural Information Processing Systems, 28.*

Jensen, M. (2015). Smile as feedback expressions in interpersonal interaction. *International Journal of Psychological Studies, 7*(4), 95. https://doi.org/10.5539/ijps.v7n4p95.

Johnston, L., Miles, L., & Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *British Journal of Social Psychology, 49*(Pt 1), 107–127. https://doi.org/10.1348/014466609X412476.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*(3), 441–454. https://doi.org/10.1037/0022-3514.68.3.441.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion, 13*(5), 505–521. https://doi.org/10.1080/026999399379168.

Kenrick, D. T., Maner, J. K., Butner, J., Li, N. P., Becker, D. V., & Schaller, M. (2016). Dynamical evolutionary psychology: Mapping the domains of the new interactionist paradigm. *Personality and Social Psychology Review, 6*(4), 347–356. https://doi.org/10.1207/s15327957pspr0604_09.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological methods, 21*(4), 493.

Kraus, M. W., & Chen, T. W. (2013). A winning smile? Smile intensity, physical dominance, and fighter performance. *Emotion, 13*(2), 270–279. https://doi.org/10.1037/a0030745.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37*(9), 1539–1553. https://doi.org/10.1037/0022-3514.37.9.1539.

Krumhuber, E. G., & Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion, 9*(6), 807–820. https://doi.org/10.1037/a0017844.

Kunz, M., Prkachin, K., & Lautenbacher, S. (2009). The smile of pain. *Pain, 145*(3), 273–275. https://doi.org/10.1016/j.pain.2009.04.009.

LaFrance, M. (2011). *Why smile?: The science behind facial expressions.* WW Norton & Company.

Maringer, M., Krumhuber, E. G., Fischer, A. H., & Niedenthal, P. M. (2011). Beyond smile dynamics: Mimicry and beliefs in judgments of smiles. *Emotion, 11*(1), 181–187. https://doi.org/10.1037/a0022596.

Martin, J. D., Abercrombie, H. C., Gilboa-Schechtman, E., & Niedenthal, P. M. (2018). Functionally distinct smiles elicit different physiological responses in an evaluative context. *Scientific Reports, 8*(1), 3558. https://doi.org/10.1038/s41598-018-21536-1.

Martin, J. D., Wood, A., Rychlowska, M., & Niedenthal, P. M. (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences., 21,* 864–877. https://doi.org/10.1016/j.tics.2017.08.007.

Mehu, M., Grammer, K., & Dunbar, R. I. (2007). Smiles when sharing. *Evolution and Human behavior, 28*(6), 415–422.

Merriam-Webster. (2020). https://www.merriam-webster.com/dictionary/smirk.

Nettle, D. (2006). *Happiness: The science behind your smile.* USA: Oxford University Press.

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(6), 417–433; discussion 433-480. doi:https://doi.org/10.1017/S0140525X10000865.

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior, 33*(3), 200–209.

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional smiles: Tools for love, sympathy, and war. *Psychological Science, 28*(9), 1259–1270. https://doi.org/10.1177/0956797617706082.

Rychlowska, M., van der Schalk, J., Niedenthal, P. M., Martin, J. D., Carpenter, S. M., & Manstead, A. R. (under review*).* Reward, affiliative, and dominance smiles communicate different social motives following trust violations. https://doi.org/10.31234/osf.io/58gpb.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*(5), 617–640. https://doi.org/10.1016/s0167-4870(01)00059-9.

Sievert, S., Ross, D., Jain, L., Jamieson, K., Nowak, R., & Mankoff, R. (2017). NEXT: A system to easily connect crowd sourcing and adaptive data collection. *Proceedings of the 16th Python in Science Conference (SCIPY).*

Stewart, P. A., Bucy, E. P., & Mehu, M. (2015). Strengthening bonds and connecting with followers. *Politics and the Life Sciences, 34*(1), 73–92. https://doi.org/10.1017/pls.2015.5.

Tanczos, E., Nowak, R., & Mankoff, B. (2017). A kl-lucb algorithm for large-scale crowdsourcing. *Advances in Neural Information Processing Systems, 30.*

Team, R. C. (2013). R: A language and environment for statistical computing.

Thibault, P., Levesque, M., Gosselin, P., & Hess, U. (2012). The Duchenne marker is not a universal signal of smile authenticity–but it can be learned! *Social Psychology, 43*(4), 215–221. https://doi.org/10.1027/1864-9335/a000122.

van Dijk, E., van Kleef, G. A., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology, 94*(4), 600–614. https://doi.org/10.1037/0022-3514.94.4.600.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014.

Witkower, Z., & Tracy, J. (2019). A facial-action imposter: How head tilt influences perceptions of dominance from a neutral face. *Psychological Science, 30*(6), 893–906. https://doi.org/10.1177/0956797619838762.

Wood, A. (under review). Social context influences the acoustic properties of laughter. https://doi.org/10.31234/osf.io/npk8u.