Reconsidering the Duchenne Smile: Formalizing and Testing Hypotheses about Eye Constriction and Positive Emotion

Jeffrey M. Girard
University of Kansas

Jeffrey F. Cohn
University of Pittsburgh

Lijun Yin
Binghamton University

Louis-Philippe Morency
Carnegie Mellon University

The common view of emotional expressions is that certain configurations of facial-muscle movements reliably reveal certain categories of emotion. The principal exemplar of this view is the Duchenne smile, a configuration of facial-muscle movements (i.e., smiling with eye constriction) that has been argued to reliably reveal genuine positive emotion. In this paper, we formalized a list of hypotheses that have been proposed regarding the Duchenne smile, briefly reviewed the literature weighing on these hypotheses, identified limitations and unanswered questions, and conducted two empirical studies to begin addressing these limitations and answering these questions. Both studies analyzed a database of 751 smiles observed while 136 participants completed experimental tasks designed to elicit amusement, embarrassment, fear, and physical pain. Study 1 focused on participants' self-reported positive emotion and Study 2 focused on how third-party observers would perceive videos of these smiles. Most of the hypotheses that have been proposed about the Duchenne smile were either contradicted by or only weakly supported by our data. Eye constriction did provide some information about *experienced* positive emotion, but this information was lacking in specificity, already provided by other smile characteristics, and highly dependent on context. Eye constriction provided more information about *perceived* positive emotion, including some unique information over other smile characteristics, but context was also important here as well. Overall, our results suggest that accurately inferring positive emotion from a smile requires more sophisticated methods than simply looking for the presence/absence (or even the intensity) of eye constriction.

Keywords: smiling, facial behavior, FACS, emotion, Bayesian data analysis

Most dictionaries define the "smile" as a pleasant facial expression, and the smiley face has become the *de facto* symbol of positive feelings in digital communication. This view of the smile as synonymous with positive emotion aligns with the belief that certain emotion categories are reliably revealed by certain configurations of facial-muscle movements (called the "common view of emotional expressions" by Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019). However, people also frequently smile when experiencing unpleasant emotions such as embarrassment, pain, and distress (e.g., Keltner, 1995; Kraut & Johnston, 1979; Landis, 1924; Prkachin & Solomon, 2008) and when signaling interpersonal information like dominance or affiliation (Martin, Rychlowska, Wood, & Niedenthal, 2017; Rychlowska et al., 2017). The smile's ubiquity has led some researchers to conclude that it has no reliable meaning and should not be

considered a reflexive expression of positive emotion (e.g., Barrett et al., 2019; Hunt, 1941; Klineberg, 1940; Tagiuri, 1968). Others have argued that there are different types of smiles (e.g., Ekman, 1985; Ekman & Friesen, 1982).

Building on work by Duchenne (1862/1990) and Darwin (1872), argued that a special type of smile, which they named the "enjoyment" or "Duchenne" smile, is involuntarily triggered by positive emotion and is identifiable by its configuration of facial-muscle movements. All smiles involve the *zygomaticus major* muscle, which pulls the lip corners toward the ears. What defines a Duchenne smile is that it also involves the *orbicularis oculi (pars orbitalis)* muscle, which lifts the cheeks, narrows the eyes, and wrinkles the outer eye corners (see Figure 1). We will use "smile" to refer to the former muscle and "eye constriction" to refer to the latter. Ekman, Davidson, and Friesen (1990) argued that eye constriction is difficult to deliberately control and rarely occurs during smiles in the absence of positive emotion; therefore, the Duchenne smile is a reliable signal of genuine positive emotion. In contrast, smiles that lack this muscle (i.e., "nonenjoyment" or "non-Duchenne" smiles) should be regarded as

Correspondence should be addressed to Dr. Jeffrey Girard, Dept. of Psychology, University of Kansas. E-mail: jmgirard@ku.edu

Table 1
Formalized hypotheses regarding the Duchenne smile

H1	Positive emotion reliably triggers the production of Duchenne smiles.
H2	Eye constriction is difficult to deliberately control and therefore Duchenne smiles are difficult to counterfeit.
H3	Duchenne smiles rarely occur in the absence of positive emotion.
H4	Eye constriction reliably distinguishes between positive emotion smiles and false/miserable smiles.
H5	Of all the smile characteristics, eye constriction is the most reliable/diagnostic marker of positive emotion.
H6	The relationship between Duchenne smiles and experienced positive emotion is independent of context.
H7	Duchenne smiles are reliably perceived as more positive than non-Duchenne smiles.
H8	The relationship between Duchenne smiles and perceived positive emotion is independent of context.

voluntary and lacking genuine positive emotion, e.g., "false" smiles that feign positive emotion or "miserable" smiles that mask/express negative emotion (Ekman & Friesen, 1982). Frank and Ekman (1993) argued that eye constriction is "the most reliable, most robust, and most diagnostic marker for an enjoyment smile," and that "the enjoyment smile is lawful behavior and that its features operate more independent of context than other types of smiles" (pp. 21–22). Furthermore, Frank, Ekman, and Friesen (1993) argued that "a person is seen as more positive when they display an enjoyment smile compared to when they display a nonenjoyment smile, again independent of the situation in which the smile is elicited" (p. 92). To facilitate communication and testing, we formalized these claims as numbered hypotheses in Table 1.

To evaluate these hypotheses, we turned to a recent literature review by Gunnery and Hall (2015). This review concluded that people are more likely to produce Duchenne smiles than non-Duchenne smiles when experiencing positive emotion (e.g. Ekman, Friesen, & O'Sullivan, 1988; Jakobs, Manstead, & Fischer, 1999; Matsumoto & Willingham, 2009; Mehu, Grammer, & Dunbar, 2007) and that observers tend to perceive Duchenne smiles as more positive than non-Duchenne smiles (Gunnery & Ruben, 2016). However, it also concluded that Duchenne smiles can be—and often are—produced deliberately and therefore are not reliable signals of genuine positive emotion (e.g., Gosselin, Perron, & Beaupré, 2010; Gunnery, Hall, & Ruben, 2013; Krumhuber & Manstead, 2009). Finally, it suggested that other smile characteristics (e.g., smile intensity and duration) may be more important than eye constriction in distinguishing positive emotion (Gunnery & Ruben, 2016; Krumhuber & Manstead, 2009). Put succinctly, it concluded that "people who feel happy are likely to Duchenne smile, but those who Duchenne smile are not necessarily happier" (p. 130).

Comparing these conclusions to the hypotheses in Table 1, the review supported H1 that positive emotion reliably triggers Duchenne smiles and H7 that Duchenne smiles are reliably perceived as more positive, but contradicted H2 that Duchenne smiles are difficult to deliberately control and H3 that Duchenne smiles rarely occur without positive emotion. It therefore cast doubt on H4 that eye constriction re-

liably distinguishes positive emotion smiles and questioned H5 that eye constriction is the best marker of positive emotion. It did not directly speak to H6 or H8 regarding context-independence (cf., Harris & Alvarado, 2005). There are thus important questions about the Duchenne smile still in need of answers. How accurately can positive emotion be inferred from a Duchenne or non-Duchenne smile? How does eye constriction compare to other smile characteristics in revealing positive emotion? To what extent does the relationship between Duchenne smiles and positive emotion differ across contexts?

There are also limitations to the existing literature. First, previous studies focused on the binary presence or absence of eye constriction, but there are statistical (DeCoster, Iselin, & Gallucci, 2009) and empirical Messinger, Mattson, Mahoor, and Cohn (2012) reasons to suspect that its dimensional intensity may be more informative. Second, previous studies focused more on false smiles than on miserable smiles, despite both being important in H4. Third, previous studies often compared smiles across conditions without accounting for the positive emotion reported in each condition. As participants may vary in their experiences of experimental tasks (e.g., finding a comedy clip underwhelming or delighting in the challenge of convincingly faking a smile), it is important to control for these variations. Finally, previous studies often analyzed images of smiles, which are less informative/representative than videos (Barrett et al., 2019).

To begin answering these questions and addressing these limitations, we conducted two studies using a large database of smiles observed while participants completed experimental tasks designed to elicit amusement, embarrassment, fear, and physical pain. Study 1 focused on participants' self-reported positive emotion and Study 2 focused on how third-party observers would perceive videos of these smiles. Both studies analyzed expert measures of smile intensity, smile duration, eye constriction presence, and eye constriction intensity. Advanced statistical modeling techniques were used to account for shared variance among smile characteristics, to represent the amount of positive emotion reported or perceived, and to assess context-dependence.
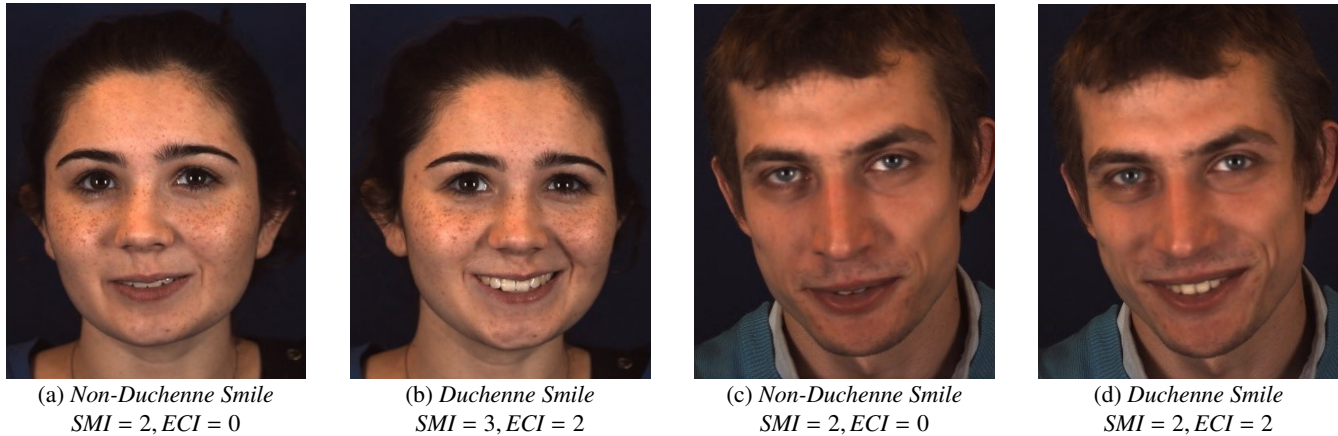
(a) *Non-Duchenne Smile*
SMI = 2, ECI = 0

(b) *Duchenne Smile*
SMI = 3, ECI = 2

(c) *Non-Duchenne Smile*
SMI = 2, ECI = 0

(d) *Duchenne Smile*
SMI = 2, ECI = 2

Figure 1

Example images from Duchenne and non-Duchenne smiles in the BP4D+ dataset with smile intensity (SMI) and eye constriction intensity (ECI) labels. These examples are typical in terms of intensity and the presence of additional action units.

Study 1

Our first study[1] focused on participants' production of facial behavior and sought to answer three research questions. (1) What are the *zero-order* relationships between self-reported positive emotion and the smile characteristic variables: eye constriction presence, eye constriction intensity, smile intensity, and smile duration? That is, what is the strength of these relationships when no other variables are controlled for? (2) Are eye constriction presence and intensity predictive of self-reported positive emotion when smile intensity and smile duration are controlled for? (3) Do the relationships between self-reported positive emotion and the smile characteristic variables differ across emotional contexts/experimental tasks?

Data

To find examples of spontaneous facial behavior, we accessed the BP4D+ multimodal spontaneous emotion database (Zhang et al., 2016), which includes video recordings and metadata from 140 participants during lab tasks intended to elicit different emotions. Although participants experienced ten total tasks, in this study we focused on the four tasks that have expert facial behavior coding. In the joke (amusement) task, the participant was told a joke by the experimenter; in the song (embarrassment) task, the participant was told to improvise a rhyming song and sing it loudly; in the darts (fear) task, darts were thrown by the experimenter at a dartboard located near the participant's head; and in the water (pain) task, the participant submerged their hand into ice water for as long as possible. The BP4D+ study was approved by the governing institutional review board and all participants consented to having their data used in further research and their images published in scientific journals.

After each task, the participant rated how intensely they had felt 14 different emotion categories during the task (and could write in and rate additional, unlisted categories); each category was rated on a six-point ordinal scale ranging from 0 (*not at all*) to 5 (*extremely*). To represent positive emotion, we used participants' self-reported responses to the question, "How much did you feel happy, joyful, or amused?"

Participants' facial behavior in each task video was observationally measured using the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002), which provides detailed rules for coding facial action units (AUs) corresponding to the movement of different facial muscles. Each video was annotated by one of five expert coders who had passed the official FACS final test and had multiple years of coding experience. A period of around 15 s was selected from each task video to be annotated; this period corresponded to when the emotion elicitation was strongest (e.g., the period leading up to when the participant removed their hand from the ice water). The coders annotated each video frame during this period (at 25 fps) for the presence and intensity of AU6 (*orbicularis oculi pars orbitalis*) and AU12 (*zygomaticus major*) . AU presence was annotated using a binary scale where 0 corresponded to the absence of the AU and 1 corresponded to its presence at any intensity level. AU intensity was annotated using a six-point ordinal scale where 0 corresponded to the absence of the AU and 1 through 5 corresponded to the official FACS intensity levels (i.e., *trace*, *slight*, *marked*, *extreme*, and *maximum*, respectively).

As reported by Zhang et al. (2016), a subset of 94 task videos was coded by two or more coders to assess the frame-by-frame reliability of the AU presence annotations. Intercoder agreement was calculated using the generalized *S*
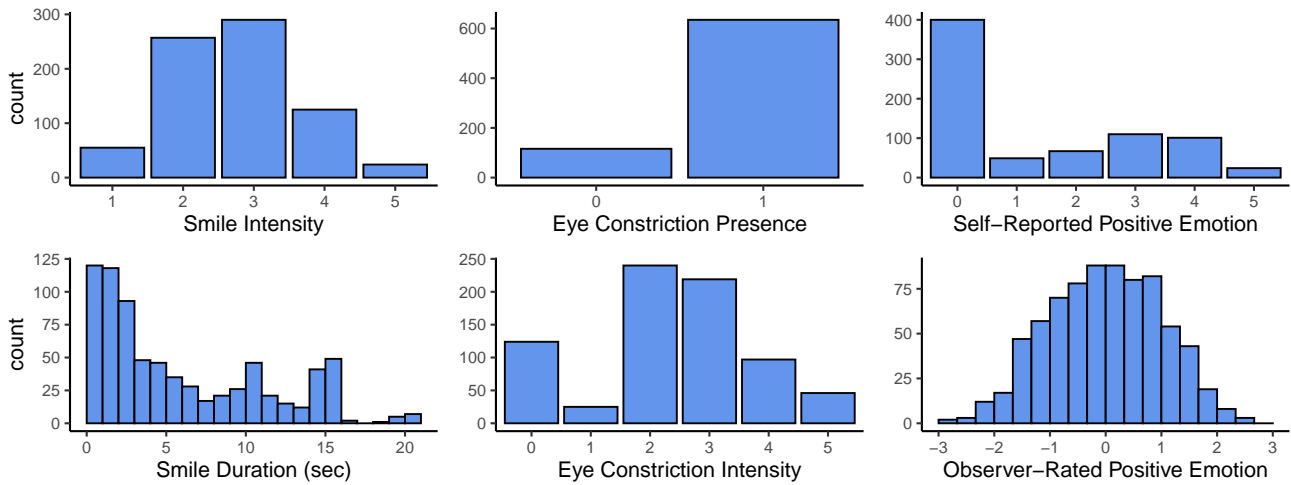
Figure 2

Histograms depicting the distributions of each study variable (Observer-Rated Positive Emotion is described in Study 2)

Table 2

Summary statistics per emotion-elicitation task

Task	Self-Reported PE Mean (SD)	Smile Count	Duchenne Smile Count (%)	EC Intensity Mean (SD)	Smile Intensity Mean (SD)	Smile Duration Mean (SD)
Joke (Amusement)	4.04 (1.17)	210	195 (95%)	3.94 (1.06)	3.14 (0.73)	9.48 (4.92)
Song (Embarrass.)	2.12 (1.52)	249	208 (87%)	3.65 (1.22)	2.99 (0.81)	8.96 (5.29)
Darts (Fear)	1.67 (1.34)	205	174 (85%)	3.36 (1.22)	2.63 (0.81)	5.48 (4.70)
Water (Pain)	1.04 (0.28)	87	58 (69%)	2.64 (1.20)	2.28 (0.90)	4.34 (5.06)

Note. PE = Positive Emotion, EC = Eye Constriction, Self-report and intensity variables range from 0 to 5, Duration is in seconds.

score (also called the Brennan-Prediger kappa coefficient; Gwet, 2014), which adjusts for chance agreement and can accommodate nominal or ordinal categories through different weighting schemes. Scores above 0.60 and 0.80 were interpreted as "good" and "very good," respectively (Gwet, 2014). Inter-coder agreement for AU presence was good for both AU6 ($S_{nominal} = 0.73$) and AU12 ($S_{nominal} = 0.77$). Similarly, a subset of nine task videos was coded by two coders to assess the reliability of the intensity annotations. Inter-coder agreement for AU intensity was good for AU6 ($S_{ordinal} = 0.70$) and very good for AU12 ($S_{ordinal} = 0.84$).

We used the FACS codes from the BP4D+ database to identify all smile events, which we defined as sequences of consecutive video frames during which AU12 was coded as present. We excluded one participant who was an outlier in terms of age (i.e., 66 years old) but retained all others who smiled at least once, resulting in a sample of 751 smile events from 136 participants. The number of smile events per participant ranged from 3 to 11 ($M = 5.5$, $SD = 2.0$), and the duration of events ranged from 0.1 to 20.1 s ($M = 6.0$, $SD = 5.3$). Participants were all students at Binghamton University. The retained sample was 60% Female and 40% Male; 46% White, 34% Asian, 10% Latino/Hispanic, and

7% Black; and ages ranged from 18 to 30 years ($M = 20.2$, $SD = 2.5$).

Model Formulation

In addition to providing descriptive statistics and a heterogeneous correlation matrix that accommodates ordinal variables (Fox, 2019), we built several sets of regression models to explore our research questions. To investigate our first research question (i.e., to quantify the zero-order relationships between self-reported positive emotion and the smile characteristic variables), we estimated separate models in which self-reported positive emotion was regressed on each smile characteristic variable as a single predictor. Slopes in these single-predictor models represent the strength of the overall relationship between the predictor and self-reported positive emotion, including any variance that predictor may share with the other smile characteristic variables.

To investigate our second research question (i.e., to determine if eye constriction presence and intensity predict self-reported positive emotion above-and-beyond the other smile characteristic variables), we estimated two related models with multiple predictors. In Model 1A, we regressed self-

Table 3
Heterogeneous correlation matrix for the study variables

	1. ECP	2. ECI	3. SMI	4. SMD	5. SRPE	6. ORPE
1. Eye Constriction Presence (ECP)		Polychoric	Polychoric	Polyserial	Polychoric	Polyserial
2. Eye Constriction Intensity (ECI)	.97		Polychoric	Polyserial	Polychoric	Polyserial
3. Smile Intensity (SMI)	.67	.69		Polyserial	Polychoric	Polyserial
4. Smile Duration (SMD)	.46	.51	.56		Polyserial	Pearson
5. Self-Reported PE (SRPE)	.21	.27	.34	.26		Polyserial
6. Observer-Rated PE (ORPE)*	.49	.51	.57	.40	.34	

Note. PE = Positive Emotion, *This variable is described in Study 2.

reported positive emotion on an ordinal variable representing the smile's intensity, a continuous variable representing the smile's standardized duration, and a binary variable (i.e., dummy code) representing eye constriction presence. In Model 1B, we replaced the eye constriction presence binary variable with an ordinal variable representing eye constriction intensity. Slopes in these multiple regression models represent the strength of the unique/partial relationship between a predictor and self-reported positive emotion, controlling for all the other predictor variables. For example, the slope of eye constriction presence in Model 1A answers the question: if the smile's intensity and duration are already known, how much does learning whether it included eye constriction help us to predict self-reported positive emotion?

To investigate our third research question (i.e., to determine if the relationships between self-reported positive emotion and the smile characteristic variables differed across tasks), we further modified Models 1A and 1B by adding moderation by a nominal variable representing the task each smile occurred during. In Model 2A, self-reported positive emotion was regressed on task, smile intensity, smile duration, eye constriction presence, and the interactions of task with smile intensity, smile duration, and eye constriction presence. In Model 2B, self-reported positive emotion was regressed on task, smile intensity, smile duration, eye constriction intensity, and the interactions of task with smile intensity, smile duration, and eye constriction intensity. These models allowed us to estimate the effect of each smile characteristic variable in each task (i.e., by adding the smile characteristic variable's main effect to the interaction effect of that same smile characteristic variable and task).

Model Building

In building these models, several aspects of the data required specialized treatment. First, the models included multiple observations (i.e., smile events) from each participant. To accommodate this hierarchical structure, we used multilevel regression models with a two-level structure, nesting smile event observations (level 1) within participants (level 2). Varying effects[2] were estimated to allow participants to

have different average levels for variables (i.e., intercepts) and different relationships between variables (i.e., slopes). In interpreting the results, we focused on the population-level effects[3], which estimate the central tendencies of the distribution of varying effects (e.g., what is the typical intercept or slope in the population?). Varying effects were added for all smile characteristic variables in all models, but not for task or the task-by-smile-characteristic interaction effects—there were too few observations to support the estimation of participant-varying task effects, so population-level task effects were estimated instead.

Second, the outcome variable (i.e., self-reported emotion) was ordinal. To accommodate this non-normal distribution, we used a form of ordinal regression called the cumulative model, which assumes that the ordinal scores come from discretization of a continuous latent variable and estimates K thresholds to partition the latent variable into $K + 1$ observable, ordered categories (Bürkner & Vuorre, 2019). This latent variable then takes the place of the predicted variable in the regression equation and the thresholds take the place of the model intercept.

Third, many predictor variables (e.g., smile intensity) were ordinal. To model these predictors, we used the monotonic effects approach (Bürkner & Charpentier, in press), which represents the relationship between an ordinal predictor and the outcome variable as a piecewise linear curve where all components have the same sign. The parameterization of each monotonic effect has two parts: a *scale* parameter, which is essentially a regression coefficient and can be interpreted as the expected average distance between two adjacent ordinal categories, and a vector of *shape* parameters, which form a simplex[4] and can be interpreted as the expected, normalized distances between each pair of adjacent categories. In addition to providing these intuitive interpretations, this approach also allows (1) the data to inform the sign of the scale parameter, (2) the distance between adjacent cat-

---

[2]Varying effects are also called "random effects."

[3]Population-level effects are also called "fixed effects."

[4]In this context, a simplex is a vector where each element is a real number between 0 and 1 and all the elements add up to 1.
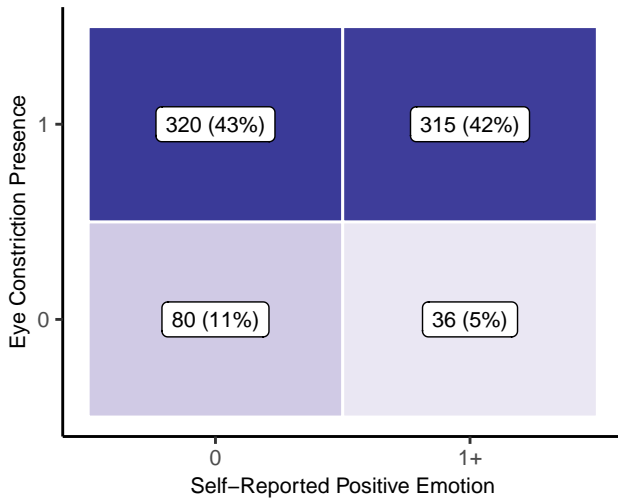
Figure 3

Contingency table and heatmap depicting the count (and proportion) of smiles observed with and without eye constriction and self-reported positive emotion

egories to differ, and (3) interaction terms and varying effects involving one or more ordinal predictors.

Finally, we had to incorporate all these approaches within the same model, which required a particularly flexible modeling framework. To do so, we implemented our models within a Bayesian multilevel modeling framework (Gelman et al., 2014; McElreath, 2016). In brief, Bayesian methods combine existing knowledge about the probability of different parameter values (in the form of prior distributions) with observed data to generate updated knowledge about the parameter values (in the form of posterior distributions). Statistical inferences can then be made using this updated knowledge (e.g., by estimating the central tendency and spread of the posterior distributions). We estimated our models using the *brms* package (Bürkner, 2017, 2018) as a high-level interface to the *Stan* platform for statistical computing (Gelman, Lee, & Guo, 2015). Model estimation was performed through Markov chain Monte Carlo (Neal, 1993) via the No-U-Turn Sampler (Hoffman & Gelman, 2014) algorithm, which converges quickly in high-dimensional models and eliminates the need for any hand-tuning. Full details about model estimation (e.g., number of chains and iterations) are provided in the supplemental materials.

In setting the prior distributions for our model parameters, we strove to exclude unreasonable values without ruling out reasonable values (Gelman et al., 2014). For the slope parameters, we used normal priors ($\mu = 0$, $\sigma = 1$) in order to apply light regularization and deter overfitting. For the intercept parameters, we used Student's t priors ($\nu = 3$, $\mu = 0$, $\sigma = 5$) in order to reflect that we did not have substantive hypotheses about these parameters. For the varying

effects' standard deviations, we used nonnegative Student's t priors ($\nu = 3$, $\mu = 0$, $\sigma = 1$) in order to reflect that negative standard deviations are unreasonable. For the correlations between varying effects, we used LKJ priors ($\eta = 1$) in order to assign equal probability to all valid correlation matrices (Lewandowski, Kurowicka, & Joe, 2009). Finally, for models including monotonic effects, we used Dirichlet priors ($\alpha = 1$) to assign equal probability to all valid shape parameter simplexes (Bürkner & Charpentier, in press).

Model Interpretation

In interpreting our model results, we had two primary goals. First, we wanted to estimate the magnitude (i.e., size and sign) of each important effect and the amount of precision (i.e., certainty) in these estimates. To accomplish these goals within a Bayesian framework, we represented the magnitude of each effect as the central tendency of its posterior distribution and the precision of each effect as the spread of its posterior distribution. Specifically, we used the posterior median as our measure of central tendency and the 89% highest density interval (HDI) as our measure of spread. The posterior median minimizes the expected absolute error and the 89% HDI is the narrowest continuous interval that contains 89% of the posterior density. The 89% HDI has become common in Bayesian data analysis because it is more stable than the 95% HDI (Kruschke, 2014) and because it highlights the arbitrariness of such threshold conventions in the first place (McElreath, 2016). Finally, for each effect, we calculated the probability of direction (*pd*), which varies from 50% to 100% and can be interpreted as the probability that a parameter is strictly positive or negative (i.e., the proportion of the posterior distribution that has the same sign as the median; Makowski, Ben-Shachar, & Lüdecke, 2019). We interpreted effects with *pd* values above 95% as statistically "significant" and effects with *pd* values above 90% as "suggestive." However, we appreciate the arbitrariness of these cutoffs and encourage readers to carefully consider the 89% HDIs and raw *pd* values.

Another goal was to quantify the proportion of variance in the outcome variable explained by each model. To accomplish this goal, we calculated Bayesian R^2 values using the approach described by Gelman, Goodrich, Gabry, and Vehtari (2019). Because our models in Study 1 were not Gaussian, we used the approach of McKelvey and Zavoina (1975) to estimate the error variance given our logit link function (see the supplemental materials for details); this approach yielded pseudo-R^2 values on the scale of the latent variables underlying our ordinal variables. Note that, as described by Gelman et al. (2019), the variance estimates contributing to the Bayesian R^2 come from the model rather than directly from the data (as in frequentist versions of R^2) because, "from a Bayesian perspective, a concept such as 'explained variance' can ultimately only be interpreted in the context of

Table 4

Population-level effects from the single-predictor models predicting self-reported positive emotion

Single Predictor	Median	89% HDI	pd	Pseudo-R^2	89% HDI
None (Intercept Only)				0.24	[0.18, 0.30]
Smile Intensity	0.61	[0.46, 0.77]	**100%**	0.37	[0.29, 0.45]
Smile Duration	0.68	[0.50, 0.86]	**100%**	0.44	[0.37, 0.52]
Eye Constriction Presence	0.84	[0.44, 1.26]	**100%**	0.27	[0.20, 0.33]
Eye Constriction Intensity	0.38	[0.27, 0.51]	**100%**	0.32	[0.25, 0.40]

Table 5

Population-level effects from the multilevel models predicting self-reported positive emotion with covariates

Parameter	Model 1A			Model 1B		
	Median	89% HDI	pd	Median	89% HDI	pd
Smile Intensity	0.47	[0.30, 0.63]	**100%**	0.43	[0.25, 0.62]	**100%**
Smile Duration	0.42	[0.22, 0.63]	**100%**	0.41	[0.20, 0.62]	**99.9%**
Eye Constriction Presence	0.01	[−0.48, 0.49]	51.0%			
Eye Constriction Intensity				0.07	[−0.07, 0.20]	78.8%
Bayesian Pseudo-R^2	0.48	[0.41, 0.55]		0.50	[0.42, 0.58]	

a model" (p. 309). Thus, differences in Bayesian R^2 values between models should be interpreted cautiously given the absence of a fixed denominator.

Finally, to address the possibility of multicollinearity (i.e., near-perfect associations between predictors) leading to instability/imprecision in our primary slope estimates, we calculated variance inflation factors (VIFs) for the eye constriction presence and intensity variables, remaining vigilant for values greater than 5 (Sheather, 2009). These VIFs were calculated from the pseudo-R^2 values of supplemental models in which the eye constriction variables were regressed on the other predictor variables (using the same framework as above, with the one deviation being that binary regression was used to predict eye constriction presence; Bergtold, Spanos, & Onukwugha, 2010).

Results

We first examined the main study variables' distributions (Figure 2) and then calculated their summary statistics by experimental task (Table 2). Despite the limitations of this simple approach (e.g., calculating the mean of ordinal variables and aggregating all smiles observed during a task is not ideal), it provided several useful pieces of information and helped motivate our more sophisticated statistical models. First, the results for self-reported positive emotion match what we would expect based on theory (e.g., that amusement is more positive than fear or pain) and serves as a basic manipulation check. Second, it revealed that mean self-reported positive emotion and all smile characteristic variables ranked the tasks in the same order: joke (amusement), song (embarrassment), darts (fear), and then water (pain). The consis-

tency of the task ordering by different variables suggests that all these variables are indexing similar information. Finally, the percentage of smiles that included eye constriction was much higher than expected in the non-amusement tasks. Indeed, most smiles included eye constriction, even when participants reported feeling little or no positive emotion. As shown in Figure 3, when ignoring task and participant identity, eye constriction presence performed poorly as a diagnostic test of positive emotion with a sensitivity of 0.90 and a specificity of 0.20 (positive predictive value = 0.50, negative predictive value = 0.69).

We next calculated correlations between the primary study variables as further descriptive statistics. We interpreted correlations as "negligible" when less than 0.1 in absolute value, as "small" when between 0.1 and 0.3, as "medium" when between 0.3 and 0.5, as "large" when between 0.5 and 0.9, and as "almost perfect" when greater than 0.9. These results are presented in Table 3 and show that all variables were positively correlated with one another. The correlation between the two eye constriction variables was almost perfect; the correlations with smile intensity were large for both eye constriction variables; the correlations with smile duration were medium for eye constriction presence and large for both eye constriction intensity and smile intensity; and the correlations with self-reported positive emotion were medium for smile intensity and small for the other variables.

In the single-predictor models exploring our first research question, all smile characteristic variables were significantly and positively associated with self-reported positive emotion (Table 4). The slopes are difficult to compare across predictors because they are scaled differently. Instead, we consider the amount of variance explained by each smile character-
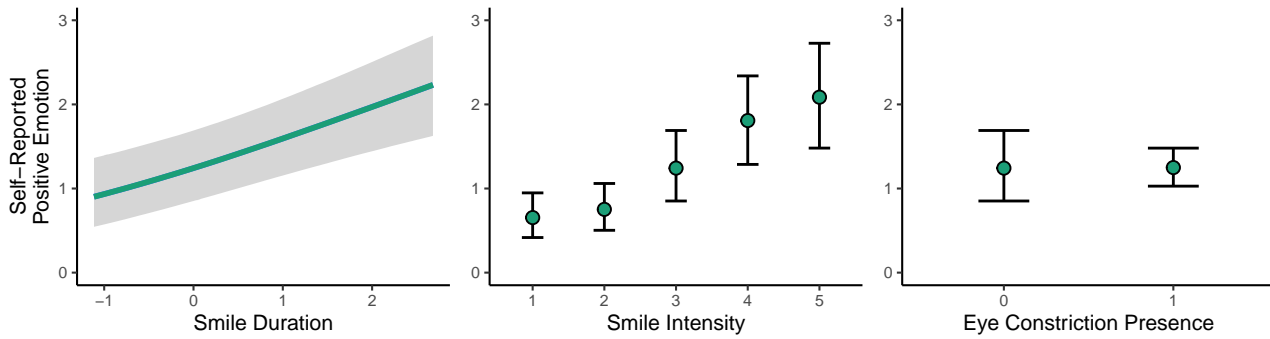
Figure 4

Conditional effects in the prediction of self-reported positive emotion from Model 1A (error bars show 89% HDIs).
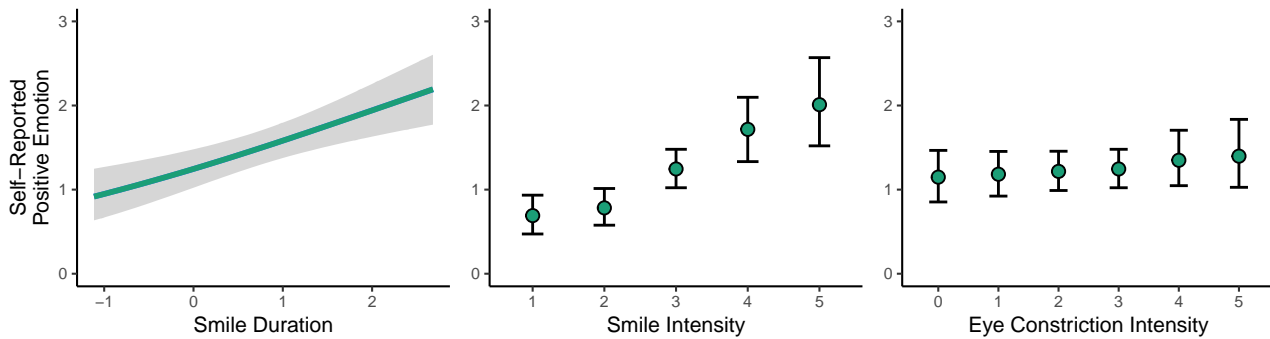
Figure 5

Conditional effects in the prediction of self-reported positive emotion from Model 1B (error bars show 89% HDIs).

istic predictor over-and-above that explained by the varying intercepts alone. This value was roughly 13% for smile intensity, 20% for smile duration, 3% for eye constriction presence, and 8% for eye constriction intensity (noting again the caveat about Bayesian R^2 values having different denominators across models). These results suggest that, in terms of zero-order relationships, (1) Duchenne smiles were associated with higher self-reported positive emotion than non-Duchenne smiles, (2) the more intense the eye constriction in a smile, the higher self-reported positive emotion tended to be, (3) both smile intensity and smile duration were also important indicators of self-reported positive emotion, seemingly even more so than was eye constriction.

In the set of multiple regression models exploring our second research question (Table 5, Figure 4, and Figure 5), the unique effects of smile intensity and duration were significant but the unique effects of eye constriction presence and intensity were nonsignificant. These results suggest that, if smile intensity and smile duration are already known, then learning about the presence or intensity of eye constriction provides very little new information about the participant's self-reported positive emotion. Given that the multicollinearity diagnostic values for eye constriction presence in Model

1A (VIF = 2.97) and eye constriction intensity in Model 1B (VIF = 2.45) were well-below the threshold value of 5.00, we discount the alternative explanation that the eye constriction effects were nonsignificant due to problematic levels of multicollinearity. Finally, even the best performing Model 1B only explained half of the variance in self-reported positive emotion (i.e., Pseudo-R^2 = 0.50), which indicates that inferring felt positive emotion from these smile characteristics was quite difficult.

In the set of multiple regression models exploring our third research question (Table 6), the effects of the smile characteristic variables differed across tasks. In both models, the partial association between smile intensity and self-reported positive emotion was significant and positive in the joke (amusement) task, suggestive and positive in the song (embarrassment) task, nonsignificant in the darts (fear) task, and significant and negative in the water (pain) task. Thus, when listening to a joke, greater smile intensity indicated more positive emotion, but when holding a hand in ice water, greater smile intensity indicated less positive emotion. In both models, the partial association between smile duration and self-reported positive emotion was nonsignificant in the joke (amusement) and song (embarrassment) tasks and sig-

Table 6

Population-level effects from the multilevel models predicting self-reported positive emotion moderated by task

Parameter	Model 2A			Model 2B		
	Median	89% HDI	*pd*	Median	89% HDI	*pd*
Smile Intensity						
Joke (Amusement) Task	0.49	[0.15, 0.86]	**99.1%**	0.53	[0.11, 0.95]	**98.4%**
Song (Embarrassment) Task	0.47	[0.02, 0.90]	93.7%	0.14	[−0.42, 0.69]	65.0%
Darts (Fear) Task	−0.03	[−0.56, 0.50]	53.4%	0.05	[−0.53, 0.67]	55.1%
Water (Pain) Task	−1.20	[−2.12, −0.29]	**99.1%**	−0.95	[−1.92, 0.04]	**95.3%**
Smile Duration						
Joke (Amusement) Task	0.03	[−0.26, 0.32]	56.8%	0.02	[−0.28, 0.33]	53.2%
Song (Embarrassment) Task	0.18	[−0.14, 0.50]	82.2%	0.10	[−0.25, 0.44]	67.3%
Darts (Fear) Task	0.70	[0.25, 1.14]	**99.4%**	0.74	[0.29, 1.21]	**99.5%**
Water (Pain) Task	1.01	[0.17, 1.82]	**97.1%**	1.20	[0.31, 2.08]	**98.4%**
Eye Constriction Presence						
Joke (Amusement) Task	0.11	[−0.59, 0.82]	59.7%			
Song (Embarrassment) Task	−0.58	[−1.42, 0.25]	86.6%			
Darts (Fear) Task	−0.53	[−1.50, 0.41]	81.3%			
Water (Pain) Task	−1.59	[−2.95, −0.13]	**96.4%**			
Eye Constriction Intensity						
Joke (Amusement) Task				0.03	[−0.20, 0.27]	58.9%
Song (Embarrassment) Task				0.15	[−0.18, 0.44]	75.8%
Darts (Fear) Task				−0.24	[−0.61, 0.11]	86.9%
Water (Pain) Task				−1.56	[−2.58, −0.66]	**99.9%**

nificant and positive in the darts (fear) and water (pain) tasks. Thus, when having darts thrown nearby or holding a hand in ice water, longer smiles indicated more positive emotion (and shorter smiles indicated less positive emotion). Finally, both eye constriction presence (in Model 2A) and eye constriction intensity (in Model 2B) had the same pattern of partial association with self-reported positive emotion: nonsignificant in the joke (amusement), song (embarrassment), and darts (fear) tasks but significant and negative in the water (pain) task. Thus, when holding a hand in ice water, eye constriction indicated less positive emotion.

Study 2

Our second study focused on how third-party observers would perceive the smile events from the first study. It sought to answer several research questions: (1) What are the zero-order relationships between observer-rated positive emotion, self-reported positive emotion, eye constriction presence, eye constriction intensity, smile intensity, and smile duration? (2) Are eye constriction presence and intensity predictive of observer-rated positive emotion when smile intensity and smile duration are controlled for? (3) Do the relationships between self-reported positive emotion and the smile characteristic variables differ across emotional contexts/experimental tasks? (4) To what extent did observer-rated positive emotion match the smiling participants' self-

reported positive emotion?

Data

We analyzed the same 751 smile events from Study 1. Video clips of each smile event were created by segmenting the original task videos. This segmentation was accomplished using the FFmpeg software program (FFmpeg Development Team, 2019), and care was taken to encode the clips using settings that would maximize their compatibility with web browsers (see the supplemental materials). Audio tracks were not included in the video clips because we wanted the perceptual ratings to reflect only the visual appearance of the smiles.

Perceptual Ratings

The smile event video clips were separated into 20 roughly equal groups. Observers were recruited using Prolific (www.prolific.co) to view and rate all clips in each group. Observers were required to have USA nationality, normal or corrected to normal vision, English language fluency, no history of mild cognitive impairment or dementia, and an approval rate of 90% or higher across 10 or more previous submissions. Using the *formr* platform (Arslan, Walther, & Tata, 2020), observers signed a consent form, provided some basic information about their own background and personality, and then viewed each clip in their as-

Table 7
Results from the confirmatory factor analysis model

Parameter	Median	89% HDI
Amused Factor Loading	0.92	[0.88, 0.97]
Comfortable Factor Loading	0.87	[0.82, 0.91]
Happy Factor Loading	1.00	[0.96, 1.04]
Amused Residual SD	0.15	[0.14, 0.16]
Comfortable Residual SD	0.25	[0.22, 0.27]
Happy Residual SD	0.00	[0.00, 0.01]

Note. RMSEA = 0.025, $\hat{\Gamma}_{adj}$ = 0.989, Mc = 0.999, CFI = 0.999.

signed group (presented in a randomized order) and provided perceptual ratings about it on three scales. Task instructions were presented before the first video clip and stated "After watching each video, please rate how much the person in the video seemed to be feeling the following: *Amused* (pleasantly entertained or diverted as by something funny), *Comfortable* (free from stress or tension), and *Happy* (enjoying or characterized by well-being and contentment)." Ratings were made on three separate, seven-point ordinal scales from 1 to 7 (see the supplemental materials for further details about the rating scales). Note that the observers were blinded to the fact that the smiles were observed during different experimental tasks.

Four approaches were used simultaneously to ensure data quality. First, observers were prescreened to have high approval ratings as previously stated. Second, the time each observer took to complete the study was recorded and observers who completed the study in less time than it would take to watch all video clips were excluded. Third, observers were asked to watch a brief video that showed a number, some letters, and an image of a food item. Observers were required to answer, in unstructured text, what was shown in the video; those with unintelligible or incorrect answers were excluded. Finally, at two points in the study (halfway and at the end), observers were shown an "attention check" video, which began as a typical smile event video clip but then quickly transitioned to showing text that instructed participants to fill out a specific combination of answers (to prove they were watching the video and paying attention). Observers who failed one or both attention checks were excluded.

For each of the 20 groups of smile events, we recruited six observers to serve as raters; thus, we began with a sample of 120 observers. Fifteen observers (13%) were excluded for one or more of the above reasons and new observers were recruited to replace them. Additionally, six observers enrolled in two different groups. As a result, we had a final sample of 114 included observers. Observers reported on their gender (71 women, 37 men, 2 non-binary, and 4 prefer not to answer), age (*min* = 18, *mdn* = 26, *max* = 69), and education (2 less than high school, 20 high school or equivalent, 38 some college, 42 college degree, and 12 graduate degree).

Inter-Observer Reliability

The inter-observer reliability of the perceptual ratings was estimated per rating scale using the agreement software package (Girard, 2020), which uses the approach described by Gwet (2014) to estimate variance components in the presence of missing data and optionally estimates average-score intraclass correlation coefficients (ICCs) and bootstrapped confidence intervals (CIs). Specifically, we used ICC model 2A[5] and estimated the absolute agreement reliability of the average of all six observers' scores. This "planned missing data" design (Graham, Taylor, Olchowski, & Cumsille, 2006) allowed us to use a two-way ICC model without needing all observers to score all videos.

We considered ICC values above 0.50 to be evidence of "moderate" reliability, values above 0.75 to be evidence of "good" reliability, and values above 0.90 to be evidence of "excellent" reliability (Koo & Li, 2016). Based on their CIs, the inter-observer reliability of the average of all six observers' scores was good for the question about how amused the smiling participant appeared to feel, $ICC = 0.83$, 95% CI: $[0.80, 0.85]$, moderate or good for the question about how comfortable the smiling participant appeared to feel, $ICC = 0.74[0.70, 0.77]$, and good for how happy the smiling participant appeared to feel, $ICC = 0.82[0.79, 0.84]$.

Latent Variable Modeling

After averaging across observers, scores on the rating scales were all highly inter-correlated ($r = 0.80$ for amused and comfortable, $r = 0.87$ for comfortable and happy, and $r = 0.92$ for amused and happy). As such, we estimated a three-indicator confirmatory factor analysis (CFA) model to capture their shared variance[6]. This analysis was conducted in a Bayesian framework using the *blavaan* software package (Merkle & Rosseel, 2018). All latent and manifest variables were standardized to zero mean and unit variance. Normal priors ($\mu = 0$, $\sigma = 5$) were used for manifest variable intercept and factor loading parameters, and gamma priors ($\alpha = 1$, $\beta = 1$) were used for the manifest variable precision (i.e., residual standard deviation) parameters. Like the Bayesian multilevel models, the CFA model was estimated

---

[5]Inter-rater reliability under ICC model 2A equals $\hat{\sigma}_o^2/\left(\hat{\sigma}_o^2 + (\hat{\sigma}_r^2 + \hat{\sigma}_e^2)/k\right)$ where $\hat{\sigma}_o^2$ is the estimated object variance, $\hat{\sigma}_r^2$ is the estimated rater variance, $\hat{\sigma}_e^2$ is the estimated residual variance, and k is the number of raters whose scores are being averaged per object.

[6]McNeish and Wolf (2020) provide compelling arguments for why this approach of estimating a CFA model with freely estimated factor loadings and residuals is preferable to using a simpler approach, such as sum or mean scores, even with highly inter-correlated indicators.

Table 8
Population-level effects from the single-predictor models predicting observer-rated positive emotion

Single Predictor	Median	89% HDI	pd	Pseudo-R^2	89% HDI
None (Intercept Only)				0.10	[0.04, 0.16]
Smile Intensity	0.48	[0.41, 0.56]	**100%**	0.40	[0.34, 0.45]
Smile Duration	0.42	[0.35, 0.48]	**100%**	0.34	[0.27, 0.40]
Eye Constriction Presence	0.91	[0.76, 1.06]	**100%**	0.20	[0.14, 0.26]
Eye Constriction Intensity	0.31	[0.27, 0.36]	**100%**	0.38	[0.32, 0.44]

Table 9
Population-level effects from the multilevel models predicting observer-rated positive emotion with covariates

Parameter	Model 3A			Model 3B		
	Median	89% HDI	pd	Median	89% HDI	pd
Smile Intensity	0.37	[0.28, 0.45]	**100%**	0.32	[0.22, 0.51]	**100%**
Smile Duration	0.16	[0.09, 0.23]	**100%**	0.10	[0.04, 0.18]	**99.1%**
Eye Constriction Presence	0.22	[0.06, 0.36]	**98.9%**			
Eye Constriction Intensity				0.13	[0.08, 0.19]	**100%**
Bayesian R^2	0.49	[0.43, 0.55]		0.54	[0.48, 0.59]	

in *Stan* using the No-U-Turn Sampler (full details in the supplemental materials).

The parameter estimates from the CFA model are provided in Table 7. We used the approach of Garnier-Villarreal and Jorgensen (2020) to evaluate model fit. The Bayesian RMSEA fit index was 0.025 (scores below 0.050 are considered "good"), the adjusted Bayesian $\hat{\Gamma}$ fit index was 0.989 (scores above 0.950 are considered "good"), the Bayesian Mc fit index was 0.999 (scores above 0.900 are considered "good"), and the Bayesian CFI index was 0.999 (scores about 0.950 are considered "good"). Given the evidence of good fit for this model, we extracted factor scores (i.e., latent means) for use in later analyses (Devlieger & Rosseel, 2017).

Model Building and Interpretation

Our approach to model building and interpretation in Study 2 mirrored that from Study 1. That is, we built single-predictor models and two sets of multiple regression models (one without and one with moderation by task) using the Bayesian multilevel modeling framework. There were two main differences to our approach in Study 2.

First, to investigate our fourth research question, we built an additional multilevel model (using the same approach as in Study 1) in which self-reported positive emotion was regressed on observer-rated positive emotion. Second, for the other models, we replaced the ordinal outcome variable representing self-reported positive emotion with a continuous outcome variable representing observer-rated positive emotion (i.e., the factor scores described above). As a result of this change, ordinal regression and pseudo-R^2 values were not necessary and Gaussian regression and standard R^2 val-

ues were used instead. For the standard deviation of the Gaussian likelihood function (i.e., the σ parameter), we used nonnegative Student's t priors ($\nu = 3$, $\mu = 0$, $\sigma = 5$).

Results

The final row in Table 3 provides the correlations involving observer-rated positive emotion. Just as with self-reported positive emotion, all these correlations were positive. Using the same interpretive heuristics as in Study 1, the correlations with observer-rated positive emotion were large for smile intensity and eye constriction intensity and medium for smile duration, eye constriction presence, and self-reported positive emotion.

In the single-predictor models exploring our first research question (regarding zero-order relationships), all smile characteristic variables were significantly and positively associated with observer-rated positive emotion (Table 8). The amount of variance explained by each smile characteristic predictor, over-and-above that explained by the varying intercepts alone, was roughly 30% for smile intensity, 24% for smile duration, 10% for eye constriction presence, and 28% for eye constriction intensity (noting again the caveat about Bayesian R^2 values having different denominators across models). These results suggest that, in terms of zero-order relationships, (1) Duchenne smiles were perceived as more positive than non-Duchenne smiles, (2) more intense eye constriction was perceived as more positive than less intense eye constriction, (3) more intense smiles were perceived as more positive than less intense smiles, and (4) longer smiles were perceived as more positive than shorter smiles.

In the set of multiple regression models exploring our sec-
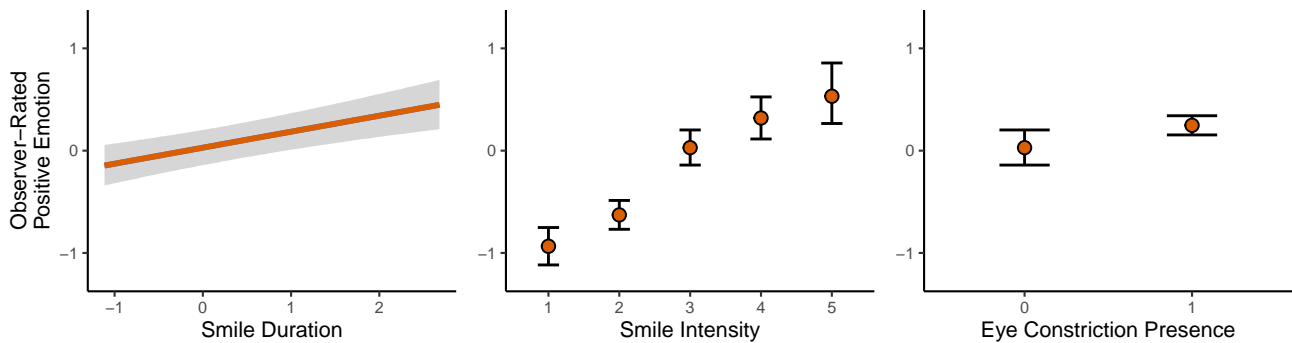
Figure 6

Conditional effects in the prediction of observer-rated positive emotion from Model 3A (error bars show 89% HDIs).
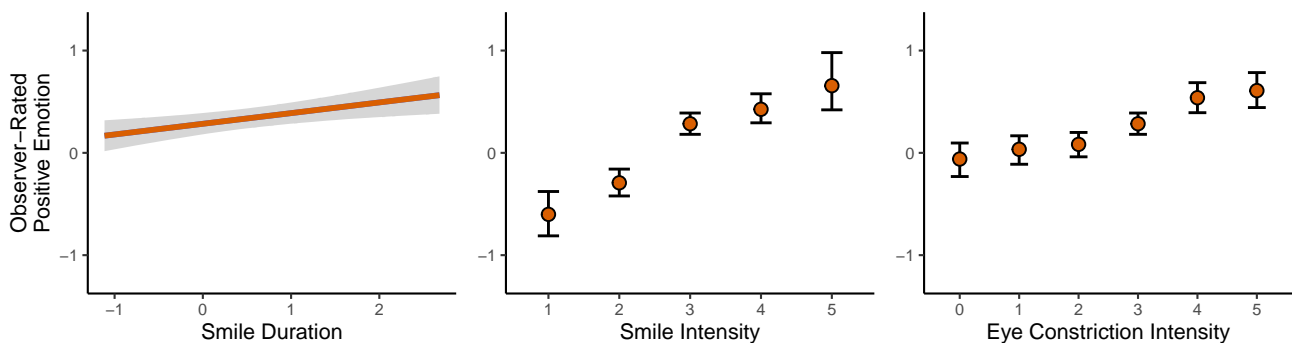
Figure 7

Conditional effects in the prediction of observer-rated positive emotion from Model 3B (error bars show 89% HDIs).

ond research question, the unique/partial effects of smile intensity, smile duration, eye constriction presence, and eye constriction intensity were significant (Table 9, Figure 6, and Figure 7). These results suggest that each of these variables added reliable new information to the prediction of observer-rated positive emotion, even when the other variables were already known. The combination of varying intercepts, smile intensity, smile duration, and eye constriction intensity (i.e., Model 3B) explained a little more than half of the variance in observer-rated positive emotion ($R^2 = 0.54$). Thus, observers' perceptions of positive emotion were being substantially influenced by other, unmeasured cues as well. Multicollinearity was not problematic for the eye constriction slopes in either Model 3A ($VIF = 2.98$) or Model 3B ($VIF = 2.45$).

In the set of multiple regression models exploring our third research question (Table 10), the unique/partial effects of the smile characteristic variables differed across tasks. In both models (i.e., controlling for eye constriction presence and intensity), the partial association between smile intensity and observer-rated positive emotion was significant and positive in the joke (amusement), song (embarrassment), and darts (fear) tasks but was nonsignificant in the water (pain)

task. In Model 4A, the partial association between smile duration and observer-rated positive emotion was significant in the joke (amusement) and song (embarrassment) tasks only; in Model 4B, these effects were nonsignificant, although the effect was suggesstive and positive in the song (embarrassment) task. In Model 4A, the partial association between eye constriction presence and observer-rated positive emotion was nonsignificant in all tasks, although it was suggestive and positive in the song (embarrassment) task. In Model 4B, the partial association between eye constriction intensity and observer-rated positive emotion was significant and positive in the joke (amusement), song (embarrassment), and darts (fear) tasks but significant and negative in the water (pain) task. Thus, more intense smiles were perceived as more positive than less intense smiles when those smiles occurred while listening to a joke, singing a silly song, or having darts thrown nearby but not when those smiles occurred while holding a hand in ice water. Similarly, smiles with more intense eye constriction were perceived as more positive than smiles with less intense eye constriction when those smiles occurred while listening to a joke, singing a silly song, or having darts thrown nearby, but the opposite was true when holding a hand in ice water. Because the ob-

Table 10

Population-level effects from the multilevel models predicting observer-rated positive emotion moderated by task

Parameter	Model 4A			Model 4B		
	Median	89% HDI	*pd*	Median	89% HDI	*pd*
Smile Intensity						
Joke (Amusement) Task	0.41	[0.29, 0.54]	**100%**	0.31	[0.19, 0.44]	**100%**
Song (Embarrassment) Task	0.35	[0.22, 0.48]	**100%**	0.26	[0.13, 0.39]	**99.9%**
Darts (Fear) Task	0.41	[0.25, 0.57]	**100%**	0.34	[0.19, 0.51]	**99.9%**
Water (Pain) Task	0.03	[−0.17, 0.22]	60.6%	0.11	[−0.11, 0.30]	79.4%
Smile Duration						
Joke (Amusement) Task	0.11	[0.01, 0.21]	**95.9%**	0.03	[−0.08, 0.12]	66.7%
Song (Embarrassment) Task	0.15	[0.05, 0.24]	**99.5%**	0.10	[0.00, 0.19]	94.6%
Darts (Fear) Task	0.04	[−0.09, 0.17]	69.8%	−0.02	[−0.14, 0.11]	58.3%
Water (Pain) Task	−0.02	[−0.19, 0.14]	57.5%	−0.06	[−0.22, 0.12]	70.5%
Eye Constriction Presence						
Joke (Amusement) Task	0.22	[−0.08, 0.52]	87.4%			
Song (Embarrassment) Task	0.21	[−0.02, 0.44]	92.2%			
Darts (Fear) Task	0.19	[−0.06, 0.43]	88.9%			
Water (Pain) Task	0.17	[−0.13, 0.47]	81.1%			
Eye Constriction Intensity						
Joke (Amusement) Task				0.18	[0.11, 0.25]	**100%**
Song (Embarrassment) Task				0.14	[0.06, 0.22]	**99.7%**
Darts (Fear) Task				0.14	[0.07, 0.22]	**99.8%**
Water (Pain) Task				−0.20	[−0.35, −0.03]	**97.9%**

servers were blind to what task the smiles came from, these differences were likely due to unmeasured context-specific behavioral cues.

Finally, the Bayesian multilevel model found that the population-level estimate of the effect of observer-rated positive emotion was significant and positive (0.90, 89% HDI: [0.72, 1.09], *pd* = 100%). This model explained a little less than half of the variance in self-reported positive emotion ($R^2 = 0.44[0.36, 0.51]$). Untrained third-party observers thus performed similarly to, though a little worse than, Models 1A and 1B in predicting self-reported positive emotion and they too left a substantial amount of variance unexplained.

General Discussion

The common view of emotional expressions is that certain configurations of facial-muscle movements reliably reveal certain categories of emotion (Barrett et al., 2019). The principal exemplar of this view is the Duchenne smile, a configuration of facial-muscle movements (i.e., smiling with eye constriction) that has been argued to reliably reveal positive emotion. We formalized a list of hypotheses that have been proposed regarding the Duchenne smile (Table 1), briefly reviewed the literature to identify limitations and unanswered questions, and conducted two empirical studies to advance the literature.

Study 1 supported H1 that positive emotion reliably trig-

gers the production of Duchenne smiles: 90% of the smiles that occurred when positive emotion was reported included eye constriction. However, contrary to H3 that Duchenne smiles rarely occur in the absence of positive emotion, eye constriction was also present in 80% of the smiles that occurred when positive emotion was not reported. Furthermore, eye constriction presence only explained 27% of the variance in self-reported positive emotion and eye constriction intensity only explained 32%. Thus, results only weakly supported H4 that eye constriction reliably distinguishes between positive emotion and false/miserable smiles. Eye constriction provided some information about positive emotion, but far less than what we would consider "lawful behavior." These results are consistent with previous findings that Duchenne smiles occurred at similar (and high) rates in both spontaneous and deliberate conditions (Krumhuber & Manstead, 2009).

Study 1 contradicted H5 that eye constriction is a more reliable marker of positive emotion than other smile characteristics. Smile intensity and duration explained more variance in self-reported positive emotion than did eye constriction, and were significant predictors of positive emotion even when controlling for eye constriction presence and intensity. In contrast, the effects of eye constriction presence and intensity were both nonsignificant when controlling for smile intensity and duration. These results suggest that, of the in-

formation that eye constriction provided about self-reported positive emotion, nearly all was shared with smile intensity and duration. These results are consistent with previous findings that Duchenne smile perception studies had smaller effect sizes when controlling for smile intensity (Gunnery & Ruben, 2016) and that training observers to focus on smile duration led to higher accuracy in distinguishing spontaneous and posed smiles than training them to focus on eye constriction (Ruan, Liang, Hong, & Yan, 2020).

Study 1 also contradicted H6 that the positive relationship between Duchenne smiles and experienced positive emotion is context-independent. When examining this relationship in each experimental task separately, the only significant partial effect was *negative*. Thus, when holding a hand in ice water, eye constriction during smiling was associated with *less* self-reported positive emotion. This result is consistent with research linking eye constriction and pain (Kunz, Meixner, & Lautenbacher, 2019) and may be viewed by some to support the theory that eye constriction signals strong emotional intensity/arousal in both positive and negative contexts[7] (e.g., Fridlund, 1994; Malek et al., 2019; Messinger et al., 2012). Interestingly, the effects of smile intensity and duration also varied by task. That none of these smile characteristics was a reliable marker of positive emotion in all tasks suggests that context is critically important when inferring positive emotion from smiles.

Study 2 supported H7 that Duchenne smiles are perceived as more positive. Eye constriction presence and intensity explained 20% and 38% of the variance in observer-rated positive emotion, respectively. When controlling for smile intensity and duration, the associations between observer-rated positive emotion and eye constriction presence and intensity were still significant. This pattern of results—nonsignificant effects in Study 1 (reported emotion) and significant effects in Study 2 (perceived emotion)—is consistent with Fernández-Dols and Ruiz-Belda's (1997) argument that the Duchenne smile is an "artistic truth" (i.e., a widely shared belief/convention borne out by perception studies) but not an "optical truth" (i.e., an empirical association borne out by production studies).

Study 2 only weakly supported H8 that the positive relationship between Duchenne smiles and perceived positive emotion is context-independent. The partial effect of eye constriction presence was nonsignificant in all four tasks (though suggestive and positive in the song task). The partial effect of eye constriction intensity was significant and positive in three tasks but was significant and negative in the fourth. The perceived positivity of a smile with eye constriction thus did depend on its context.

The following limitations should be considered while interpreting these results. First, all participants were students at an American university and all observers were American; future work with other populations is needed. Second, we examined behavior in a controlled laboratory setting, which may differ from behavior in more naturalistic settings. Third, only a single positive context and emotion was examined; future work on other positive contexts and emotions is needed. Finally, experienced positive emotion was measured using a single self-report item after each task, which may have been influenced by subject effects (Weber & Cook, 1972) and may not have described all moments within that task equally well.

In conclusion, we found that most of the hypotheses that have been proposed about the Duchenne smile were either contradicted by or only weakly supported by our data. Eye constriction did provide some information about experienced positive emotion, but this information was lacking in specificity, already provided by other smile characteristics, and highly dependent on context. The best support we found for the Duchenne smile was that it was *perceived* as more positive (although this also depended on context). Overall, our results suggest that accurately inferring positive emotion from a smile will require more sophisticated methods than simply looking for the presence or absence (or even the intensity) of a single facial-muscle movement. We believe that success in this endeavor will require the careful synthesis of additional behavioral and contextual information.

Declarations

Conflicts of interest: None.

Ethics approval: The collection of the BP4D+ dataset was approved and overseen by the Institutional Review Board at Binghamton University. The collection of perceptual ratings of videos from the BP4D+ dataset was approved and overseen by the Institutional Review Board at Carnegie Mellon University.

Consent to participate: All participants in the BP4D+ dataset provided informed consent to participate and to have their data shared with other researchers, shown to participants in further studies, and printed in scientific journals.

Availability of data and material: The BP4D+ dataset, including videos and FACS annotations, is available from Binghamton University. Because access to this data requires signing a license, instead of providing direct access to the data as part of this work, we are providing code to reproducibly derive our data from the distributed dataset. This code, as well as other materials not covered by the BP4D+ license, is available online (`https://osf.io/k3g2e`) under an open source license.

---

[7]Testing this theory is beyond the scope of this paper (and would require looking beyond just smiles). However, we note that the nonsignificant partial effects of eye constriction in the joke task are problematic for this theory.

Code availability: All code used to derive our measures from the BP4D+ dataset and analyze the data is available online at the aforementioned link.

Authors' contributions: JFC and LY designed, collected, and provided consultation about the BP4D+ dataset. JFC and JMG managed the process of annotating the dataset for facial action units. JMG conceived the present studies, wrote the code, collected the perceptual ratings, ran the statistical analyses, and wrote the initial draft of the manuscript under the advisement of LPM. All authors contributed to editing and approved the submitted manuscript.

References

Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, *52*, 376–387. doi: 10/gfx5zz

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. doi: 10/gf4772

Bergtold, J. S., Spanos, A., & Onukwugha, E. (2010). Bernoulli regression models: Revisiting the specification of statistical models with binary dependent variables. *Journal of Choice Modelling*, *3*(2), 1–28. doi: 10/gf2kzw

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10/gddxwp

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. doi: 10/gfxzpn

Bürkner, P.-C., & Charpentier, E. (in press). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*. doi: 10/ggqtkh

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. doi: 10/gfv26q

Darwin, C. (1872). *The expression of emotions in man and animals* (3rd ed.). New York, NY: Oxford University.

DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*(4), 349–366. doi: 10/bh86w7

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology*, *13*, 31–38. doi: 10/gbs2p7

Duchenne, G. B. (1862/1990). *The mechanism of human facial expression* (R. A. Cuthbertson, Ed. & Trans.). Cambridge University Press.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.

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. doi: 10/csk95r

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. doi: 10/fvkmgg

Ekman, P., Friesen, W. V., & Hager, J. (2002). *Facial action coding system: A technique for the measurement of facial movement*. Salt Lake City, UT: Research Nexus.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*(3), 414–414. doi: 10/cfxtd6

Fernández-Dols, J.-M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J.-M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 255–274). New York, NY: Cambridge University Press.

FFmpeg Development Team. (2019). *FFmpeg*.

Fox, J. (2019). *Polycor: Polychoric and polyserial correlations*.

Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *Humor - International Journal of Humor Research*, *6*(1). doi: 10/b787gh

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, *64*(1), 83–93. doi: 10/czz6g4

Fridlund, A. J. (1994). Facial reflexes and the ontogeny of facial displays. In *Human facial expression: An evolutionary view* (pp. 99–122). San Diego, CA: Academic Press.

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, *25*(1), 46–70. doi: 10/gf835p

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, *73*(3), 307–309. doi: 10/gf2hnb

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, *40*(5), 530–543. doi: 10/gfzk2m

Girard, J. M. (2020). *Agreement: An R package for the tidy analysis of agreement and reliability*.

Girard, J. M., Shandar, G., Liu, Z., Cohn, J. F., Yin, L., & Morency, L.-P. (2019). Reconsidering the Duchenne smile: Indicator of positive emotion or artifact of smile intensity? In *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 594–599). Cambridge, UK: IEEE. doi: 10.1109/ACII.2019.8925535

Gosselin, P., Perron, M., & Beaupré, M. (2010). The voluntary control of facial action units in adults. *Emotion*, *10*(2), 266–271. doi: 10/fh3xqn

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. doi: 10/d7v7nv

Gunnery, S. D., & Hall, J. A. (2015). The expression and perception of the Duchenne smile. In A. Kostic & D. Chadee (Eds.), *The*

social psychology of nonverbal communication. New York, NY: Palgrave Macmillan.

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, *37*(1), 29–41. doi: 10/gfzrfz

Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion*, *30*(3), 501–515. doi: 10/gfzrf2

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (Fourth ed.). Gaithersburg, MD: Advanced Analytics.

Harris, C. R., & Alvarado, N. (2005). Facial expressions, smile types, and self-report during humour, tickle, and pain. *Cognition and Emotion*, *19*(5), 655–669. doi: 10/fs3rhh

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Hunt, W. A. (1941). Recent developments in the field of emotion. *Psychological Bulletin*, *38*(5), 249–276. doi: 10/dzc6sd

Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (1999). Social motives, emotional feelings, and smiling. *Cognition & Emotion*, *13*(4), 321–345. doi: 10/fht7rn

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, *68*(3), 441–441. doi: 10/b436ks

Klineberg, O. (1940). *Social psychology*. New York, NY: Holt.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. doi: 10/b84r

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, *37*(9), 1539–1539. doi: 10/bvndw6

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, *9*(6), 807–820. doi: 10/dvcfcz

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Second ed.). Cambridge, MA: Academic Press.

Kunz, M., Meixner, D., & Lautenbacher, S. (2019, March). Facial muscle movements encoding pain—a systematic review. *PAIN*, *160*(3), 535. doi: 10/gf86kb

Landis, C. (1924). Studies of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology*, *4*(5), 447–510. doi: 10/bcjhxb

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. doi: 10/bs7dst

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. doi: 10/gf9pds

Malek, N., Messinger, D., Gao, A. Y. L., Krumhuber, E., Matt-

son, W., Joober, R., . . . Martinez-Trujillo, J. C. (2019). Generalizing Duchenne to sad expressions with binocular rivalry and perception ratings. *Emotion*, *19*(2), 234–241. doi: 10/gf42t2

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences*, 1–14. doi: 10/gcf8p5

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, *96*(1), 1–10. doi: 10/cd7w3s

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. New York, NY: CRC Press.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*(1), 103–120. doi: 10/dqfhpp

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*. doi: 10/ggssqh

Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior*, *28*(6), 415–422. doi: 10/dbxxmx

Merkle, E. C., & Rosseel, Y. (2018). Blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(1), 1–30. doi: 10/gf7fkx

Messinger, D. S., Mattson, W. I., Mahoor, M. H., & Cohn, J. F. (2012). The eyes have it: Making positive expressions more positive and negative expressions more negative. *Emotion*, *12*(3), 430–436. doi: 10/dcn4np

Neal, R. M. (1993). *Probabilistic inference using Markov Chain Monte Carlo methods* (Technical Report No. CRG-TR-93-1). Department of Computer Science: University of Toronto.

Prkachin, K. M., & Solomon, P. E. (2008). The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, *139*(2), 267–274. doi: 10/drrknk

Ruan, Q.-N., Liang, J., Hong, J.-Y., & Yan, W.-J. (2020). Focusing on mouth movement to improve genuine smile recognition. *Frontiers in Psychology*, *11*, 1126. doi: 10/gg9vq8

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. doi: 10/gfrkm2

Sheather, S. J. (2009). Diagnostics and transformations for multiple linear regression. In *A modern approach to regression with R* (pp. 151–225). New York, NY: Springer-Verlag.

Tagiuri, R. (1968). Person perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 395–449). Reading, MA: Addison-Wesley.

Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, *77*(4), 273–295. doi: 10/bccth2

Zhang, Z., Girard, J. M., Wu, Y., Zhang, X., Liu, P., Ciftci, U., . . . Yin, L. (2016). Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3438–3446). Las Vegas, NV: IEEE. doi: 10.1109/cvpr.2016.374