

# Reconsidering the Duchenne Smile: Indicator of Positive Emotion or Artifact of Smile Intensity?

Jeffrey M. Girard

Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA  
jmgirard@cmu.edu

Gayatri Shandar

Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA  
gshandar@cs.cmu.edu

Zhun Liu

Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA  
zhunl@cs.cmu.edu

Jeffrey F. Cohn

Department of Psychology  
University of Pittsburgh  
Pittsburgh, PA  
jeffcohn@pitt.edu

Lijun Yin

Department of Computer Science  
Binghamton University  
Binghamton, NY  
lijun@cs.binghamton.edu

Louis-Philippe Morency

Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA  
morency@cs.cmu.edu

**Abstract**—The Duchenne smile hypothesis is that smiles that include eye constriction (AU6) are the product of genuine positive emotion, whereas smiles that do not are either falsified or related to negative emotion. This hypothesis has become very influential and is often used in scientific and applied settings to justify the inference that a smile is either true or false. However, empirical support for this hypothesis has been equivocal and some researchers have proposed that, rather than being a reliable indicator of positive emotion, AU6 may just be an artifact produced by intense smiles. Initial support for this proposal has been found when comparing smiles related to genuine and feigned positive emotion; however, it has not yet been examined when comparing smiles related to genuine positive and negative emotion. The current study addressed this gap in the literature by examining spontaneous smiles from 136 participants during the elicitation of amusement, embarrassment, fear, and pain (from the BP4D+ dataset). Bayesian multilevel regression models were used to quantify the associations between AU6 and self-reported amusement while controlling for smile intensity. Models were estimated to infer amusement from AU6 and to explain the intensity of AU6 using amusement. In both cases, controlling for smile intensity substantially reduced the hypothesized association, whereas the effect of smile intensity itself was quite large and reliable. These results provide further evidence that the Duchenne smile is likely an artifact of smile intensity rather than a reliable and unique indicator of genuine positive emotion.

**Index Terms**—Duchenne smile, facial expression, nonverbal behavior, emotion, FACS, Bayesian data analysis

## I. INTRODUCTION

Ekman and colleagues [1]–[4] famously hypothesized that genuine positive emotion (e.g., amusement) is revealed by the *Duchenne smile*, a facial expression involving “smiling eyes” as well as a smiling mouth. In the language of the Facial Action Coding System (FACS) [5], which is often used to

This material is based upon work partially supported by the National Science Foundation (1629716, 1629898, 1722822, 1734868) and National Institutes of Health (MH096951). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of National Science Foundation or National Institutes of Health, and no official endorsement should be inferred.

study facial expressions, the Duchenne smile is composed of the *lip corner puller* action unit (AU12) and the *cheek raiser* action unit (AU6). All smiles involve AU12, which pulls the lip corners toward the ears, but Duchenne smiles also involve AU6, which lifts the cheeks, narrows the eyes, and wrinkles the outer eye corners (hence the “smiling eyes”). Smiles without AU6 are called non-Duchenne smiles and are hypothesized to be disconnected from genuine positive emotion (e.g., false, miserable, or masking negative emotion [1]).

The Duchenne smile hypothesis has been very influential and is still commonly adopted in the social/medical sciences [6], [7] and in affective computing [8], [9]. It is often used to classify smiles as “real” or “fake” and to infer participants’ level of positive emotion from their face. It is also central to discrete emotion theory [10], which considers the Duchenne smile to be the universally recognized and produced facial expression (i.e., prototype) of positive emotion.

However, empirical support for the hypothesis has been mixed [11]. Some studies have supported it, finding that Duchenne smiles were more common in response to pleasant stimuli [2], [12], [13] and more associated with self-reported positive emotion [4], [14], [15] than non-Duchenne smiles. But others have cast doubt on it, finding that Duchenne smiles occur at high rates when participants feel negative emotion [2], [14], [16]–[22] or feign positive emotion [11], [23]–[27].

In attempting to explain these mixed results, Krumhuber and Manstead [11] proposed that AU6 may just be an *artifact* (i.e., side-effect) of high AU12 intensity. That is, intense smiles may be more likely to include AU6 regardless of what the smiling person is feeling. If this hypothesis is correct, then previously observed associations between AU6 and positive emotion may have actually been caused by an unmeasured association between AU12 intensity and positive emotion [28]. Later research supported the plausibility of this hypothesis. Gosselin et al. [29] instructed participants to activate AUs one-at-a-time and found that many of them added AU6 while

trying to activate AU12. Moreover, the likelihood of adding AU6 increased with the intensity of AU12. This phenomenon also operated in reverse: many participants added AU12 while trying to activate AU6. These results suggest that there is a dependency between these AUs that encourages them to occur together, especially when the intensity of AU12 is high (and even in the absence of felt positive emotion).

The most direct test of the artifact hypothesis examined the facial behavior of 32 participants during two conditions: viewing amusing stimuli and trying to feign amusement [11]. Contrary to the Duchenne smile hypothesis, rates of Duchenne and non-Duchenne smiles were very similar in both conditions. So, AU6 did not discriminate between felt and false smiles. However, it did discriminate between weak and intense smiles; across both conditions, Duchenne smiles were much higher in AU12 intensity than non-Duchenne smiles. Thus, as predicted by the artifact hypothesis, the occurrence of AU6 was related to AU12 intensity but not to genuine positive emotion.

Several important questions still remain, however. First, this work only compared smiles during expressed amusement and feigned amusement; it did not compare smiles during positive and negative emotions. Thus, it is currently unknown whether Krumhuber and Manstead's [11] finding will generalize to other contexts that the Duchenne smile hypothesis makes predictions about (e.g., negative emotions). Second, this work did not directly compare predictors in the same model (i.e., it did not statistically control for AU12 intensity). Thus, the relative importance of these variables and the exact extent to which AU12 intensity confounds these relationships is currently unknown. Finally, nearly all previous work dichotomized smiles as Duchenne or non-Duchenne without considering the intensity of AU6. However, Messinger et al. [30] argued that the intensity of AU6 contains important information about felt emotion. Thus, it is possible the relationship between AU6 and positive emotion would be strengthened by increasing the granularity with which AU6 is measured.

The current study addressed these gaps in the literature and made several novel contributions. First, we examined smiles during four different emotional contexts (i.e., amusement, embarrassment, fear, and pain), which allowed us to test whether the artifact hypothesis can generalize beyond feigned positive emotion to negative emotion. Including multiple emotions also allowed us to better assess the generalizability of our findings. Second, we directly compared predictors in two related models (i.e., predicting AU6 and amusement) by statistically controlling for AU12 intensity. We also did so using a much larger sample than previous research ( $N = 136$ ) and a more sophisticated modeling approach (i.e., Bayesian multilevel regression with interactions). Finally, we represented AU6 as a continuum rather than a dichotomy, which allowed us to leverage the full extent of this variable's predictive power [31].

We explored two related research questions. First, does AU6 predict how amused a smiling participant feels, even when controlling for AU12 intensity? If the Duchenne smile hypothesis is correct, then the answer should be yes. However, if the artifact hypothesis is correct, then the answer should

be no (or at least the predictive power of AU6 should be largely reduced by controlling for AU12 intensity). Second, does self-reported amusement predict whether a smile contains AU6 (and to what degree), even when controlling for AU12 intensity? If the Duchenne smile hypothesis is correct, then amusement should be uniquely predictive of AU6. However, if the artifact hypothesis is correct, then AU12 intensity should be very related to AU6 and amusement should not be.

## II. METHODS

### A. Data

1) *BP4D+ Database*: To find examples of spontaneous facial behavior, we accessed the BP4D+ database [32], which includes video recordings and metadata from 140 participants during tasks meant to elicit different positive and negative emotions. In the amusement task, the participant was told a joke; in the embarrassment task, the participant was told to improvise a rhyming song and sing it loudly; in the fear task, darts were thrown at a dartboard located near the participant's head; and in the pain task, the participant submerged their hand into ice water for as long as possible.

After each task, the participant rated how intensely they had felt 14 different emotions [32]; each emotion was rated on a six-point scale ranging from 0 (not at all) to 5 (extremely). In this study, we focused on participants' ratings of amusement, which were their responses to the question, "How much did you feel happy, joyful, or amused?"

Facial actions in each task video were annotated by one of five expert coders who had passed the official FACS final test [5]. 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 intensity of AU6 and AU12. Intensity was annotated using a six-point scale where 0 corresponded to the absence of the AU and 1 through 5 corresponded to the official FACS intensity levels (i.e., "trace" through "maximum").

A subset of task videos was coded by two or more coders to assess the reliability of the codes [32]. Inter-coder agreement was calculated using the ordinal-weighted S score (also called the  $\kappa_{BP}$  coefficient) [33], which adjusts for chance agreement and can accommodate scales with ordinal/ordered categories. Scores above 0.60 are often interpreted as "good" and scores above 0.80 are often interpreted as "very good" [33]. Inter-coder agreement was good for AU6 intensity ( $S = 0.70$ ) and very good for AU12 intensity ( $S = 0.84$ ).

2) *Smile Events*: We used the BP4D+ FACS codes to identify all smile events, i.e., sequences of consecutive frames during which AU 12 was coded as having an intensity greater than zero. We excluded events from one participant who was an outlier in terms of age (i.e., 66 years old) but retained all other events, 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 each event ranged from 0.1 s to 20.1 s ( $M = 6.0$ ,  $SD = 5.3$ ).

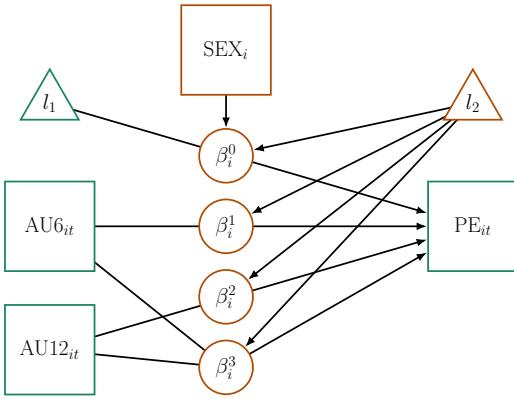


Fig. 1. Path diagram of Model 1. Squares are observed variables, circles are random effects, triangles are intercepts, and arrows are regression paths. Level 1 includes the AU intensity variables as well as the amusement/positive emotion rating (PE) for smile  $t$  from participant  $i$ . Level 2 includes the sex (SEX) and random effects ( $\beta$ ) for participant  $i$ .

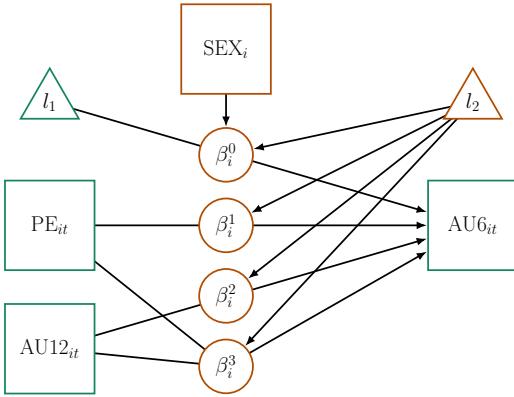


Fig. 2. Path diagram of Model 2. Squares are observed variables, circles are random effects, triangles are intercepts, and arrows are regression paths. Level 1 includes the AU intensity variables as well as the amusement/positive emotion rating (PE) for smile  $t$  from participant  $i$ . Level 2 includes the sex (SEX) and random effects ( $\beta$ ) for participant  $i$ .

The sample was 60 % Female, 46 % White, 34 % Asian, 10 % Latino/Hispanic, and 7 % Black. Ages ranged from 18 to 30 years ( $M = 20.2$ ,  $SD = 2.5$ ). All participants were currently living in the USA and most had lived in the USA for many years ( $M = 15.3$ ,  $SD = 7.1$ ).

Consistent with previous research [11], we calculated each AU's intensity level as its peak (i.e., maximum) value in each smile event. In order to facilitate interpretation and comparisons between predictors, the AU intensity variables and the amusement ratings were standardized to have zero mean and unit variance prior to analysis and visualization.

## B. Statistical Modeling

1) *Model Building*: To explore our research questions, we estimated two linear regression models. Model 1 regressed self-reported amusement on AU6 intensity, AU12 intensity, and the interaction of the two. This model allowed us to quantify how well amusement can be inferred from these

variables and what the nature (i.e., magnitude and direction) of these relationships was. Model 2 regressed AU6 intensity on self-reported amusement and AU12 intensity; it allowed us to quantify how important each variable was in explaining the presence and intensity of AU6 in a smile. Both models also controlled for participant sex.

The regression models were estimated using a Bayesian multilevel modeling framework [34], [35]. Multilevel models were necessary to account for the hierarchical structure of the data (i.e., there were multiple observations/events from each participant). Using a single-level model could lead to incorrect conclusions because observations from the same participant are more similar than observations from different participants. As a result, the standard error estimates (which assume observations are independent) would be biased [36].

The models are depicted in Figures 1 and 2 using multilevel path diagrams [37]. Both models had the same two-level structure with observations (Level 1) clustered within participants (Level 2). The only difference between them was that AU6 was predicting amusement in Model 1 and amusement was predicting AU6 in Model 2. On Level 1, relationships between AU intensities and self-reported amusement were captured using regression pathways. In addition to direct pathways for each predictor, we also included the interaction of both predictors to allow for conditional effects (i.e., the effect of one predictor depending on the value of the other). On Level 2, intercepts were regressed on participant sex to control for sex differences in average levels of amusement and AU6. Additionally, random (i.e., participant-varying) effects were estimated to allow participants to have different average levels (i.e., intercepts) and different relationships between variables (i.e., slopes). Multilevel models that include random effects are more complex than models that do not, but are also more accurate and generalizable [38]. The regression coefficients that make up the primary results of these models capture the central tendency and variability in the random effects.

Bayesian methods were used to estimate the models because they offer many practical and interpretive advantages such as the ability to incorporate prior knowledge about parameters into the model and to estimate the probability of different parameter values [34], [35]. In brief, Bayesian methods combine existing knowledge about the probability of different model parameters (in the form of prior distributions) with observed data to generate updated knowledge about the parameters (in the form of posterior distributions). Statistical inferences can then be made using this updated knowledge. We estimated our models using the *brms* package [39] as an interface to the *Stan* platform [40]. Complete model details, including syntax, are provided in the supplemental materials.

2) *Model Interpretation*: In interpreting our model results, we had two primary goals: to estimate the magnitude (i.e., size and sign) of each important effect, and to quantify the amount of precision (i.e., certainty) in these estimates. To accomplish these goals within a Bayesian framework, we followed recent guidelines from Kruschke [41], [42]. For each effect (i.e., regression coefficient), we represented its magnitude using a

TABLE I  
STANDARDIZED REGRESSION COEFFICIENTS FROM MODEL 1

Parameter	Estimate	95 % HDI	ROPE
Intercept	−0.07	[ −0.20, 0.06 ]	.386
Sex (Male)	0.07	[ −0.12, 0.26 ]	.320
AU6 Intensity	0.08	[ −0.02, 0.17 ]	.285
AU12 Intensity	0.25	[ 0.15, 0.33 ]	.000
Interaction Effect	0.07	[ 0.01, 0.14 ]	.235

*Note.* The outcome variable in this model was amusement.

*point estimate* and its precision using an *interval estimate*.

Each point estimate is the single most credible value for the effect's magnitude given the data and our model. For example, if our point estimate of the AU 6 intensity effect in Model 1 was 0.2, then this would mean that each one standard deviation increase in AU 6 intensity would be associated with a 0.2 standard deviation increase in self-reported amusement. The median of each effect's posterior distribution was used as its point estimate (the median minimizes the posterior risk for the absolute-value loss function [43]).

Each interval estimate is the range of the most credible values for the effect's magnitude given the data, our model, the amount of variability in the random effect, and the amount of sampling error. The more uncertainty there is in the magnitude of an effect, the wider its interval estimate will be. For example, if our interval estimate for the same effect described earlier was from 0.1 to 0.3, we could be more certain in its size and sign than if its interval estimate was instead from −0.2 to 0.6. The 95 % highest density interval (HDI) of each effect's posterior distribution was used as its interval estimate (the 95 % HDI is the narrowest interval that contains 95 % of the posterior probability density [42]).

To provide information about the practical importance of each effect, we also used the region of practical equivalence (ROPE) technique [42]. We defined magnitudes within the range of −0.05 to 0.05 to be practically equivalent to zero<sup>1</sup> (i.e., too small to be meaningful in practice). We then calculated the proportion of each HDI that fell within this range. The smaller this proportion is, the more confident we can be that the effect is large enough to be practically important. If desired, we can also use the HDI+ROPE decision rule to make discrete interpretations of the results [42]. The rule is simple: If the entire HDI falls within the ROPE (1.0), accept the null hypothesis and consider the effect to be practically equivalent to zero. If the entire HDI falls outside the ROPE (0.0), reject the null hypothesis and consider the effect large enough to be practically important. Otherwise, withhold a decision.

### III. RESULTS

The AU intensity variables were highly correlated with one another ( $r = 0.64$ ), and self-reported amusement was weakly correlated with both AU6 intensity ( $r = 0.21$ ) and AU12 intensity ( $r = 0.28$ ). The results of Model 1 are summarized

TABLE II  
STANDARDIZED REGRESSION COEFFICIENTS FROM MODEL 2

Parameter	Estimate	95 % HDI	ROPE
Intercept	0.04	[ −0.06, 0.13 ]	.557
Sex (Male)	−0.07	[ −0.23, 0.07 ]	.342
Amusement	0.05	[ −0.02, 0.11 ]	.536
AU12 Intensity	0.63	[ 0.57, 0.70 ]	.000
Interaction Effect	−0.01	[ −0.08, 0.05 ]	.901

*Note.* The outcome variable in this model was AU6 intensity.

in Table I and depicted in Figure 3. The results of Model 2 are summarized in Table II and depicted in Figure 4. Complete results details are provided in the supplementary materials.

Model 1 found that, when holding either AU's intensity constant, the other AU's intensity was positively related to amusement. However, the effect of AU12 intensity was more than three times larger in magnitude than the effect of AU6 intensity. There was also some evidence of an interaction between the AU intensity variables. When AU12 intensity was low, the intensity of AU6 did not matter; but when AU12 intensity was high, higher AU6 intensity predicted more amusement. No reliable sex difference in amusement was apparent. When adopting the HDI+ROPE decision rule, only the AU12 intensity effect resulted in rejecting the null hypothesis. This model explained a little less than one-third of the variance in self-reported amusement,  $R^2 = 0.30$  [0.24, 0.35], which indicates that the intensities of AU6 and AU12 (as well as participant sex) constitute only a fraction of the information needed to successfully infer a participant's level of amusement.

Model 2 found that, when holding AU12 intensity constant, there was almost no relationship between amusement and AU6 intensity. In contrast, when holding amusement constant, AU12 intensity was strongly related to AU6 intensity. Indeed, the latter effect was over twelve times larger in magnitude than the former. No reliable interaction between amusement and AU12 intensity was apparent, and no reliable sex difference in AU6 was apparent. When adopting the HDI+ROPE decision rule, only the AU12 intensity effect resulted in rejecting the null hypothesis. This model explained around half of the variance in AU6 intensity,  $R^2 = 0.51$  [0.46, 0.55]. Two pieces of information can be gleaned from this. First, it was easier to predict AU6 intensity using the study variables than it was to infer amusement. And second, because it was the only meaningful predictor in the model, this means that AU12 intensity explained around half the variance in AU6 intensity.

To see what the effects of AU6 and amusement would have been without controlling for AU12 intensity, we re-ran both models without it. (The syntax and results for these models are provided in the supplemental materials.) When ignoring AU12 intensity, the effect of AU6 intensity in predicting amusement was 0.22 [0.15, 0.29], ROPE = .000. This means that controlling for AU12 intensity reduced this effect by 64 % and changed the HDI+ROPE decision from reject to withhold. Similarly, when ignoring AU12 intensity, the effect of amusement in predicting AU6 intensity was 0.24

<sup>1</sup>This range is conventionally used for standardized regression coefficients like ours because it corresponds to half of a “small” effect size [42, p. 277].

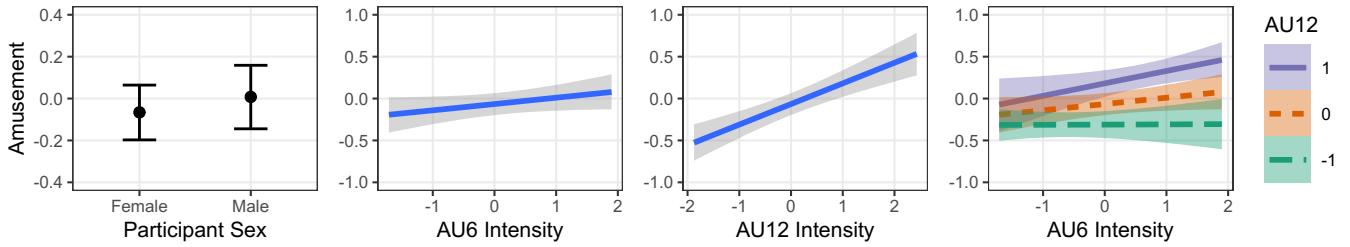


Fig. 3. Marginal Effects Plot from Model 1 with 95% HDI. This model predicted the amount of self-reported amusement associated with each smile.

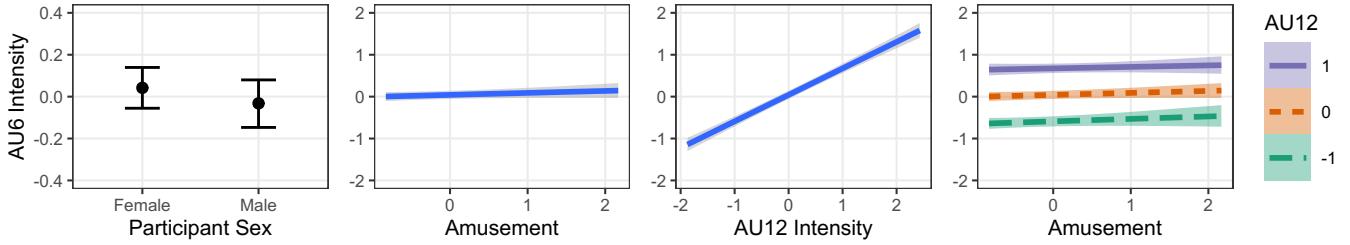


Fig. 4. Marginal Effects Plot from Model 2 with 95% HDI. This model predicted the intensity level of AU6 (the cheek raiser) in each smile.

[0.16, 0.31], ROPE = .000. This means that controlling for AU12 intensity reduced this effect by 79 % and changed the HDI+ROPE decision from reject to withhold.

#### IV. DISCUSSION

The Duchenne smile hypothesis [1]–[4] is that AU6 reveals the emotion underlying a smile and distinguishes felt smiles from false and negative emotion smiles. This hypothesis has become very popular in many research areas and is often used to justify the inference that a person is (or is not) feeling positive emotion based on the observation of a Duchenne (or non-Duchenne) smile. However, empirical support for this hypothesis has been called into question by researchers who have argued that AU6 may just be an artifact of high AU12 intensity. Thus, previously observed associations between AU6 and felt positive emotion may have been inflated by both variables' correlations with AU12 intensity [11], [28].

The current study directly tested this possibility in two different ways and found strong support for it both times. When trying to infer positive emotion from the appearance of a smile, the predictive power of AU6 was reduced considerably by controlling for AU12 intensity. Similarly, when trying to predict whether a smile would include AU6 (and to what degree), the predictive power of self-reported amusement was almost completely erased by controlling for AU12 intensity.

There are three main implications of these findings. First, future research on the Duchenne smile hypothesis should always measure and consider AU12 intensity. As shown in our study, failing to do so can seriously distort estimates of the relationship between AU6 and emotion (e.g., inflating their magnitude by up to 79 %). Second, researchers should refrain from inferring the presence or absence of positive emotion on the basis of Duchenne smiles alone. We found that such an

inference will often be incorrect, especially when the intensity of AU12 is low. Instead, future work should continue to explore other smile characteristics that have shown promise in revealing the meaning of smiles (e.g., symmetry and dynamics [27], [44], [45]). Given that the intensities of AU6 and AU12 only explained around 30 % of the variance in amusement, there is a lot of room for other smile characteristics to make contributions. Finally, researchers evaluating theories of emotion should consider the possibility that much of the evidence in favor of the Duchenne smile hypothesis may have been confounded by the unmeasured intensity of AU12 and therefore inflated [28]. Thus, theories that advance this hypothesis, such as discrete emotions theory [10], may have less evidentiary support than previously believed.

The reason for the strong association between AU6 and AU12 intensity is not completely understood, but it seems that an anatomical or psychomotor dependency between the facial muscles encourages AU6 and AU12 to co-occur [29]. Because the FACS manual [5] explicitly describes how to determine whether or not a strong AU12 includes AU6, we believe this association is real and not just a bias of the FACS coders.

In conclusion, the current study found that associations between felt positive emotion and Duchenne smiles are reduced when AU12 intensity is controlled for. These results replicate and extend previous work [11] by showing that this phenomenon generalizes beyond feigned smiles and applies to negative emotion smiles as well. We also quantified, for the first time, the extent of this reduction and found it to be substantial (i.e., 64 % when predicting amusement and 79 % when predicting AU6). This suggests that, rather than being a reliable indicator of positive emotion, AU6 may just be an artifact of intense smiles.

## SUPPLEMENTAL MATERIALS

Complete syntax and output for all analyses are provided through the Open Science Framework: <https://osf.io/t7e2d>

## REFERENCES

[1] P. Ekman and W. V. Friesen, "Felt, false, and miserable smiles," *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238–252, 1982.

[2] P. Ekman, R. J. Davidson, and W. V. Friesen, "The Duchenne smile: Emotional expression and brain physiology: II," *Journal of Personality and Social Psychology*, vol. 58, no. 2, pp. 342–353, 1990.

[3] M. G. Frank and P. Ekman, "Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles," *Humor - International Journal of Humor Research*, vol. 6, no. 1, 1993.

[4] M. G. Frank, P. Ekman, and W. V. Friesen, "Behavioral markers and recognizability of the smile of enjoyment," *Journal of Personality and Social Psychology*, vol. 64, no. 1, pp. 83–93, 1993.

[5] P. Ekman, W. V. Friesen, and J. Hager, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Salt Lake City, UT: Research Nexus, 2002.

[6] M. J. Bernstein, S. G. Young, C. M. Brown, D. F. Sacco, and H. M. Claypool, "Adaptive responses to social exclusion: Social rejection improves detection of real and fake smiles," *Psychological Science*, vol. 19, no. 10, pp. 981–983, 2008.

[7] R. Gadassi and N. Mor, "Confusing acceptance and mere politeness: Depression and sensitivity to Duchenne smiles," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 50, pp. 8–14, 2016.

[8] J. Hofmann, T. Platt, and W. Ruch, "Laughter and smiling in 16 positive emotions," *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 495–507, 2017.

[9] K. E. Haddad, S. Dupont, N. d'Alessandro, and T. Dutoit, "An HMM-based speech-smile synthesis system: An approach for amusement synthesis," in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, vol. 05, 2015, pp. 1–6.

[10] P. Ekman, "Are there basic emotions?" *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.

[11] E. G. Krumhuber and A. S. R. Manstead, "Can Duchenne smiles be feigned? New evidence on felt and false smiles," *Emotion*, vol. 9, no. 6, pp. 807–820, 2009.

[12] R. Soussignan and B. Schaal, "Forms and social signal value of smiles associated with pleasant and unpleasant sensory experience," *Ethology*, vol. 102, pp. 1020–1041, 1996.

[13] D. Matsumoto and B. Willingham, "Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals," *Journal of Personality and Social Psychology*, vol. 96, no. 1, pp. 1–10, 2009.

[14] P. Ekman, W. V. Friesen, and M. O'Sullivan, "Smiles when lying," *Journal of Personality and Social Psychology*, vol. 54, no. 3, pp. 414–414, 1988.

[15] E. Jakobs, A. S. R. Manstead, and A. H. Fischer, "Social motives, emotional feelings, and smiling," *Cognition & Emotion*, vol. 13, no. 4, pp. 321–345, 1999.

[16] U. Hess and P. Bourgeois, "You smile—I smile: Emotion expression in social interaction," *Biological Psychology*, vol. 84, no. 3, pp. 514–520, 2010.

[17] K. Schneider and I. Josephs, "The expressive and communicative functions of preschool children's smiles in an achievement-situation," *Journal of Nonverbal Behavior*, vol. 15, no. 3, pp. 185–198, 1991.

[18] C. R. Harris and N. Alvarado, "Facial expressions, smile types, and self-report during humour, tickle, and pain," *Cognition and Emotion*, vol. 19, no. 5, pp. 655–669, 2005.

[19] D. Keltner, "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame," *Journal of Personality and Social Psychology*, vol. 68, no. 3, pp. 441–441, 1995.

[20] V. Lee and G. Beattie, "The rhetorical organization of verbal and nonverbal behavior in emotion talk," *Semiotica*, vol. 120, pp. 39–92, 1998.

[21] D. Keltner and G. A. Bonanno, "A study of laughter and dissociation: Distinct correlates of laughter and smiling during bereavement," *Journal of Personality and Social Psychology*, vol. 73, no. 4, pp. 687–702, 1997.

[22] A. Papa and G. A. Bonanno, "Smiling in the face of adversity: The interpersonal and intrapersonal functions of smiling," *Emotion*, vol. 8, no. 1, pp. 1–12, 2008.

[23] J. M. Carroll and J. A. Russell, "Facial expressions in Hollywood's portrayal of emotion," *Journal of Personality and Social Psychology*, vol. 72, no. 1, pp. 164–176, 1997.

[24] S. D. Gunnery, J. A. Hall, and M. A. Ruben, "The deliberate Duchenne smile: Individual differences in expressive control," *Journal of Nonverbal Behavior*, vol. 37, no. 1, pp. 29–41, 2013.

[25] K. L. Schmidt, Z. Ambadar, J. F. Cohn, and L. I. Reed, "Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling," *Journal of Nonverbal Behavior*, vol. 30, no. 1, pp. 37–52, 2006.

[26] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American Journal of Physical Anthropology*, vol. 116, no. S33, pp. 3–24, 2001.

[27] K. L. Schmidt, S. Bhattacharya, and R. Denlinger, "Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 35–45, 2009.

[28] S. D. Gunnery and J. A. Hall, "The expression and perception of the Duchenne smile," in *The Social Psychology of Nonverbal Communication*, A. Kostic and D. Chadee, Eds. New York, NY: Palgrave Macmillan, 2015.

[29] P. Gosselin, M. Perron, and M. Beaupré, "The voluntary control of facial action units in adults," *Emotion*, vol. 10, no. 2, pp. 266–271, 2010.

[30] D. S. Messinger, T. D. Cassel, S. I. Acosta, Z. Ambadar, and J. F. Cohn, "Infant smiling dynamics and perceived positive emotion," *Journal of Nonverbal Behavior*, vol. 32, no. 3, pp. 133–155, 2008.

[31] J. DeCoster, A.-M. R. Iselin, and M. Gallucci, "A conceptual and empirical examination of justifications for dichotomization," *Psychological Methods*, vol. 14, no. 4, pp. 349–366, 2009.

[32] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. F. Cohn, Q. Ji, and L. Yin, "Multimodal spontaneous emotion corpus for human behavior analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3438–3446.

[33] K. L. Gwet, *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*, 4th ed. Gaithersburg, MD: Advanced Analytics, 2014.

[34] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: CRC Press, 2014, vol. 1542.

[35] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York, NY: CRC Press, 2016.

[36] P. D. Bliese and P. J. Hanges, "Being both too liberal and too conservative: The perils of treating grouped data as though they were independent," *Organizational Research Methods*, vol. 7, no. 4, pp. 400–417, 2004.

[37] P. J. Curran and D. J. Bauer, "Building path diagrams for multilevel models," *Psychological Methods*, vol. 12, no. 3, pp. 283–297, 2007.

[38] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[39] P.-C. Bürkner, "Brms: An R Package for Bayesian Multilevel Models Using Stan," *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, Aug. 2017.

[40] A. Gelman, D. Lee, and J. Guo, "Stan: A probabilistic programming language for Bayesian inference and optimization," *Journal of Educational and Behavioral Statistics*, vol. 40, no. 5, pp. 530–543, 2015.

[41] J. K. Kruschke and T. M. Liddell, "The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective," *Psychonomic Bulletin & Review*, vol. 25, no. 1, pp. 178–206, 2018.

[42] J. K. Kruschke, "Rejecting or accepting parameter values in Bayesian estimation," *Advances in Methods and Practices in Psychological Science*, vol. 1, no. 2, pp. 270–280, 2018.

[43] E. T. Jaynes, *Probability Theory: The Logic of Science*, 5th ed. Cambridge University Press, 2007.

[44] U. Hess and R. E. Kleck, "Differentiating emotion elicited and deliberate emotional facial expressions," *European Journal of Social Psychology*, vol. 20, no. 5, pp. 369–385, 1990.

[45] Z. Ambadar, J. F. Cohn, and L. I. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 17–34, 2009.