A trigger-substrate model for smiling during an automated formative quiz: engagement is the substrate, not frustration

Harry J. Witchel*
Harry L. Claxton
Daisy C. Holmes
Thomas T. Ranji
Joe D. Chalkley
Carlos P. Santos
Carina E. I. Westling
Brighton and Sussex Medical School
and University of Sussex
Brighton, UK

Michel F. Valstar
School of Computer Science
University of Nottingham
Nottingham, UK

Matt Celuszak
Patrick Fagan†
Crowd Emotion Ltd.
London, UK
matt@crowdemotion.co.uk

ABSTRACT

Introduction: Automated tutoring systems aim to respond to the learner's cognitive state in order to maintain engagement. The end-user's state might be inferred by interactive timings, bodily movements or facial expressions. Problematic computerized stimuli are known to cause smiling during periods of frustration. **Methods:** Forty-four seated, healthy participants (age range 18-35, 18 male) used a handheld trackball to answer a computer-presented, formative, 3-way multiple choice geography quiz, with 9 questions, lasting a total of 175 seconds. Frontal facial videos (10 Hz) were collected with a webcam and processed for facial expressions by CrowdEmotion using a pattern recognition algorithm. Interactivity was recorded by a keystroke logger (Inputlog 5.2). Subjective responses were collected immediately after each quiz using a panel of visual analogue scales (VAS). **Results:** Smiling was five-fold enriched during the instantaneous feedback segments of the quiz, and this was correlated with VAS ratings for engagement but not with happiness or frustration. Nevertheless, smiling rate was significantly higher after wrong answers compared to correct ones, and frustration was correlated with the number of questions answered incorrectly. **Conclusion:** The apparent disconnect between the increased smiling during incorrect answers but the lack of correlation between VAS frustration and smiles suggests a trigger-substrate model where engagement is the permissive substrate, while the noises made by the quiz after wrong answers may be the trigger.

CCS CONCEPTS

• **HCI design and evaluation methods**;

*Correspondence: h.witchel@bsms.ac.uk
†Current address: IMS, Goldsmiths, University of London, UK

ECCE'18, September 5–7, 2018, Utrecht, Netherlands

ACM ISBN 978-1-4503-6449-2/18/09…$15.00
https://doi.org/10.1145/3232078.3232084

KEYWORDS

smile, online quiz, boredom, cognitive states, automated tutor

ACM Reference Format:
Harry J. Witchel, Harry L. Claxton, Daisy C. Holmes, Thomas T. Ranji, Joe D. Chalkley, Carlos P. Santos, Carina E. I. Westling, Michel F. Valstar, Matt Celuszak, and Patrick Fagan. 2018. A trigger-substrate model for smiling during an automated formative quiz: engagement is the substrate, not frustration. In *ECCE'18: Proceedings of the 36th European Conference on Cognitive Ergonomics (ECCE2018), September 5–7, 2018, Utrecht, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3232078.3232084

1 INTRODUCTION

Effective automated systems for tutoring should respond to the end user's cognitive states [1], which may be recognised non-intrusively by measuring facial expressions or postural movements [14]. Smile recognition can be quantified by human scoring with the Facial Action Coding System. This process has been revolutionized by applying automated pattern recognition to facial landmarks (e.g. the mouth, eyes and upper cheek) [13]. According to researchers such as Paul Ekman and his followers in the Facial Expression Programme, Duchenne smiles reflect inner states of cheerfulness or amusement, while Behavioural Ecology Theory suggests that all smiles, including Duchenne smiles, are tools used in social interactions; they claim that cheerfulness is neither necessary nor sufficient for smiling [12]; Crivelli et al. [4] have even suggested that being engaged in a social interaction is the strongest predictor for Duchenne smiles. Hoque et al. [6] showed that frustration can immediately lead to smiling by using as a triggering stimulus an apparent computer malfunction, leading to frustration; another interpretation, based on social intention, is that compliance is a trigger for smiling, as it reflects cooperation or submission to a dominant individual [9], even if the team mate or dominant individual is a computer [11].

Much has been made of the timing of Duchenne smiles, and that they are more rare than the Facial Expression Programme would predict [4], given the frequency of positive emotions. One model for understanding rare behaviour comes from the study of the electrical basis of arrhythmias, the Trigger-Substrate model [10]. Triggers are proximate causes that are short-lived, unpredictable, and idiosyncratic events, while substrates are permissive states that are medium-term, more predictable and measurable. A simple example from aviation would be that explosions require a spark to interact with available fuel. In human computer interaction, a substrate would be a mood, while a trigger could either be a computer event or an end user's passing thought. Our research question is: in our formative HCI quizzes, why is smiling more common just after the question ends than during the question or during the explanation, and does this reflect Non-Instrumental Movement Inhibition (NIMI) [14]? This is also an exploratory study that tests what other mental states besides NIMI leads to smiling.

2 METHODS

2.1 Participants and Subjective Measures

Forty-four healthy participants (age range 18-35, 18 male) were recruited from the university community. Ethical approval was granted via the local university ethics committee, and all participants provided informed consent according to the Declaration of Helsinki. Seated participants interacted with a computer while alone in a room according to our standard protocol [14]; their faces were filmed from a frontal aspect by a webcam (Logitech HD 720p) at 10 Hz. Each stimulus lasted approximately 175 seconds, and between stimuli participants were asked to rate the previous experience using a set of 10-cm Visual Analogue Scale (VAS) measures that had the anchors at 0 ("not at all") and 100 ("extremely"). Interactions were performed with a handheld trackball in order to minimize non-instrumental postural changes. All of the participants' interactions were tracked with millisecond accuracy by a keystroke logger (Inputlog 5.2 [8]).

2.2 Stimulus

In this study only one stimulus was considered: a geography quiz (GQ) that asked nine fairly difficult 3-way multiple-choice questions (average volunteer score 3.50 out of 9) in a fairly entertaining way (mean VAS interest = 71.14 out of 100). This quiz was unrelated to any course work or reward. Our analysis broke apart each of the nine questions into a question period and an answer period, and we refer to an entire question + answer as "an item" (see Figure 1). The question was accompanied by game-show music, and a countdown timer was visible as their time was running out. Although the participant could answer each question whenever they were ready to, each of the nine items lasted 19.5 seconds; after clicking on the proposed answer, the quiz immediately informed the participant of the correct answer (accompanied by a consonant musical sound when correct, and a comedy sound (a cow mooing) when the answer was incorrect) for three seconds ("the feedback"), followed by slowly revealing (over the course of 10 seconds) five interesting facts ("the explanation") relating to the question (until the 19.5 seconds were complete); the explanation was silent. Because the average time to answer each question was 7.63 seconds, during most questions participants had sufficient time to see at least some of the explanation.

2.3 Smile Recognition and Analysis

Scoring of each of the films for facial expressions was performed by CrowdEmotion (www.crowdemotion.co.uk). A training set of 40,000

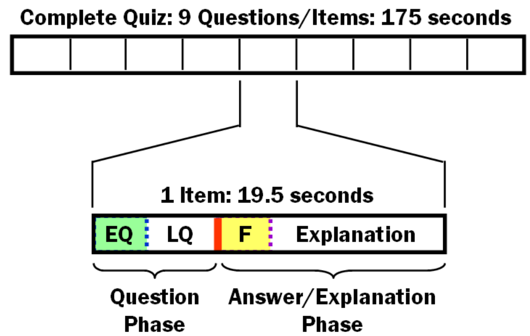

Figure 1: Structure of the geography quiz timing. The top half of the diagram represents the whole stimulus, and is followed by the subjective questionnaire. The bottom half shows the timing of each question. EQ = early question segment, LQ = late question segment, F = feedback (early answer) segment. Red line represents the moment when the participant selects an answer (clicks); this moment is selected by the participant, so the durations of the LQ and the explanation are variable.

ground truth face images were used to train a multiple-category Convolutional Neural Network (CNN) with the six Ekman facial expression categories: anger, disgust, fear, happiness, sadness, and surprise. The same Dynamic Deep Learning architecture was used as that presented by [7], but without the Long-Short Term Memory Neural Network on the end. Films were analysed as individual frames, and outputs are provided as percentages (or certainty) for a given emotion, allowing total certainty to be less than 100%. In this study only the smile-based ("happiness") time course data was used. csv files were read into Matlab, and individual smiles were identified using a local peak finding algorithm [3]; a smile was judged to be occurring if there was a local peak (delta = 0.2) above the nearest trough, or delta was set to 0.1 if the entire range was < 0.3. All peak detection initially reversed the time course (and restored the direction at the end) to guarantee that the initial peak was preceded by diminished smile certainty. A baseline was determined, and a peak start point was identified as 15% of the range between the baseline and the local peak. All statistics were non-parametric due to the floor and ceiling effects in VAS data; Matlab was used to calculate Spearman correlations and Sign Tests.

3 RESULTS

3.1 Smile Rate: Question vs. Answer

To verify whether participants smile more during the question phase than during the answer phase of each quiz item, mouse activity from Inputlog was used to divide the smile assessment time series into question and answer phases. Local peaks in the smile assessment curve were detected by using a peak detection algorithm (see methods). The initiation of each smile was recognized as the point at which the smile assessment curve first reached 15% of the range between the baseline and the local peak. Smile rates (smiles initiations recognized per second) were 80% higher during the answer periods (0.042 ± 0.006, $mean \pm s.e.m.$) than during the

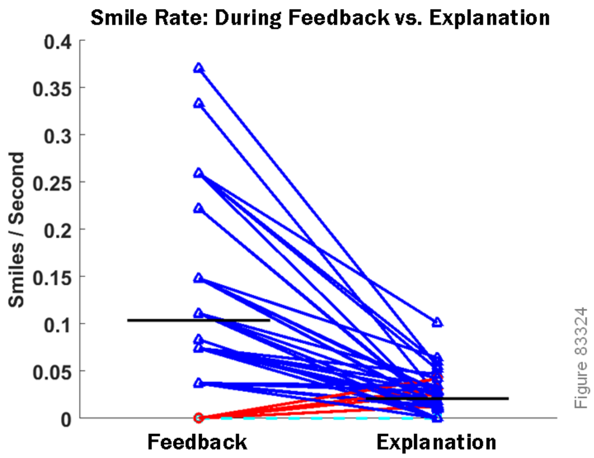

Figure 2: Comparison of smile rate during feedback (early answer) vs. during explanation (late answer), paired by participant. Black horizontal lines are mean values. For each participant: red line = smile rate for feedback < explanation, dark blue line = smile rate for feedback > explanation, dashed cyan line = smile rate for feedback is the same as for explanation.

question periods (0.023 ± 0.004, $p < 0.001$, $n = 44$, sign test). This supports our original hypothesis that NIMI inhibits smiling. We then explored the data to generate hypotheses, to determine whether a triggering phenomenon was responsible. The smile rates were calculated with the question period broken into two segments (the first three seconds, relating to the arrival of the new question, vs. the latter portion of the question), and with the answer/explanation periods divided into two segments (the first three seconds when the answer are shown (feedback), and the remainder where the explanation appears). The smile rate during the early part of the answer (0.10 ± 0.01) was four-fold larger than the rates for the other periods (early question 0.020 ± 0.005, late question 0.024 ± 0.005, and later answer/explanation 0.021 ± 0.003), and this difference was highly significant ($p < 0.00001$, sign test), while the other smile rates were not significantly different from each other ($p > 0.1$). Figure 2 shows a comparison by participant of smile rate during the early answer (feedback) vs. during the late answer (explanation) time segments. To understand whether the increased smile rate, apparently triggered during the transition from question to answer, was due to wrong answers that led to frustrated smiles [6], we compared the smile rate in the early answer (feedback) segments for questions that were answered correctly to the smile rate during feedback when items were answered incorrectly, grouped by participant. The smile rate immediately after answering incorrectly (0.12 ± 0.02) was significantly higher than immediately after answering correctly (0.08 ± 0.02, $p < 0.05$, sign test). We also found that smile rates during the feedback segment (but not during the question or during the explanation segment) was highly inversely correlated to the number of correct answers (rho = -0.412, $p < 0.01$). Thus, feedback from the quiz stating that the answer was wrong seems to trigger a higher smile rate than correct answers.

3.2 Effect of Multiple Wrong Answers

To verify that correct and wrong answers elicit the emotions expected, we ran Spearman's correlations between the number of correct/wrong answers for each participant vs. the VAS ratings they provided for different emotions. The VAS rating most correlated with correct/wrong answers was frustration, which was strongly inversely correlated with the number of correct answers (rho = -0.432, $p < 0.01$), and this was followed by an inverse correlation between correct answers and the statements "I wanted it to end earlier" (rho = -0.303, $p < 0.05$) and "I felt apathetic or detached" (rho = -0.301, $p < 0.05$). The number of correctly answered questions was significantly directly correlated with the statement "I wanted to see more" (rho = 0.329, $p < 0.05$) and there was a trend for correlation with "I cared about it" (rho = 0.267, $p < 0.10$) and "interested" (rho = 0.251, $p < 0.10$). Note that there was no significant correlation for "happiness". Together, this subjective data suggests that getting answers wrong on this quiz added to the participants' overall rating of frustration, while getting answers right was rewarding, and to a lesser extent interesting, and did not clearly contribute to happiness or positive affect. This fits with the idea that not all rewarding experiences are pleasurable or lead to positive affect [2].

3.3 Smile Rate vs. Cognitive States

To determine whether frustration may be triggering smiling, we calculated the correlations between the end-of-quiz VAS ratings of frustration (and other mental states) for each participant against the smile rate for each of the time segments (total quiz, questions only, early questions only, late question segment only, answer/explanations, early answer segments only, late answer/explanation segments only). For the full items, the only subjective ratings correlated with smile rate was engagement (rho = 0.38, $p < 0.01$) and challenging (rho = 0.61, $P < 0.01$). None of the other descriptors (bored, interested, happy, frustrated, wanted to see more, cared about it) were even close to significant.

To understand if these smile rate correlations are based on the answers being correct or on the segment of the item, we tested correlations in the eight conditions (correct vs. wrong, question vs. answer, early vs. late), which showed that engagement was correlated with smile rate during all segments except the early question segment, and this held completely true for questions answered incorrectly, but less so for questions answered correctly. During the late answer/explanation for incorrectly answered questions, challenging was directly correlated (rho = 0.677, $p < 0.01$) and bored was inversely correlated (rho = -0.321, $p < 0.05$). Also during incorrectly answered questions, during the late question segment, bored was inversely correlated (rho = -0.302, $p < 0.05$) and "I cared about it" was positively correlated (rho = 0.308, $p < 0.05$). Interestingly, during the early question segment, the only correlation with smile rate (only during incorrectly answered questions) was "I felt tired" (rho = 0.509, $p < 0.01$), as if the arrival of a new, impossible challenge elicited surrender or submission. During correctly answered questions, the only correlation was with challenging during the late answer/explanation.

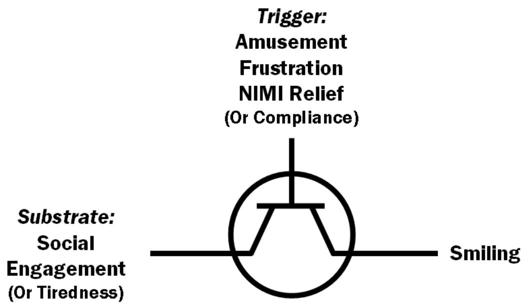

Figure 3: Schematic of a trigger-substrate model for controlling smiling during formative quizzes. The drive to engage or interact (e.g. socially) provided by the substrate is interrupted, controlled and directed by the triggering signals.

3.4 Time Spent Selecting the Answer

The time spent selecting an option before clicking (i.e. delay before answering) may relate to confusion, uncertainty, or effort. When comparing for each individual the time spent selecting correct answers ($mean = 7.3 \pm 0.3$ seconds) to the time spent selecting incorrect answers (7.9 ± 0.3), there was a trend toward significance ($P < 0.1$, Wilcoxon Signed Rank test). We determined whether the average time spent before selecting an option by each individual (7.57 seconds for the cohort) was correlated with any of the ratings for the VAS; only ratings for engagement were correlated (rho = 0.356, $P < 0.05$). For wrongly-answered items, the correlation was highly significant between time spent selecting and engagement (rho = 0.429, $p < 0.01$). For correctly answered items, there was a trend for correlation between time spent selecting and "tired" (rho = 0.372, $p < 0.1$). Taken together, time spent selecting a correct answer was modestly related to tiredness, and time spent selecting a wrong answer is significantly dependent on effort (i.e. engagement).

4 DISCUSSION

In this study we verified that, during these computerized quizzes, smiling was radically enhanced just after answering questions incorrectly, and that this behaviour was explained by self-ratings of engagement (and negatively with boredom), but it was not explained by ratings of happiness or frustration. A trigger-substrate model can explain our unpredicted results (Figure 3), which fit precisely with results from observations of winners of judo matches suggesting that smiling is linked to social engagement rather than happiness [4]. This supports Behavioural Ecology Theory [5].

The post-quiz VAS measurements that we make after our short quizzes (175 seconds) seem to register substrate moods rather than fleeting triggers. Our exploratory correlation analysis shows that engagement is a substrate for smiling throughout most of the quiz, feelings of challenge are a strong substrate after wrong answers, and tiredness can be a substrate, but only during the early question segment for triggers extending the entire stimulus; this may reflect compliance or effort. We have no conclusive evidence for the triggers for smiling in our quizzes, but we suspect that the quiz's noises that announce the correct/wrong answer (which could elicit amusement) may be one possibility. If frustration is a trigger, then those fleeting feelings of frustration do not penetrate into our post-test VAS self-assessment; however, the additive frustration from repeatedly getting questions wrong plainly does affect post-test self-assessment. Release from NIMI (or "relief") also may be a trigger, although this does not explain the increased smiling after wrong answers compared to correct answers.

We conclude that in this type of computerized formative quiz, participants are much more likely to smile during the initial answer – especially when wrong – than during the question or explanation segments, and that these smiles are a strong sign of engagement, challenge and learning motivation – the cognitive substrate. Our data suggest that in our quiz there is often a smiling trigger at the transition between the question and answer phases. To understand the trigger of smiling in our quizzes, future experiments are required. To disentangle the triggering relationship between the quiz noises, temporary frustration (or achievement emotions for correct answers), and NIMI, a larger experiment is required with versions of the quiz that a) provide no answer but have the noise upon answering, b) provide the answer but have no noise, and c) do not provide the answer or the noise.

ACKNOWLEDGMENTS

We gratefully acknowledge the BSMS IRP programme for funding, Chätrin Tolga and Terri Desmonds for administrative support, and Farrokh Bulsara for the original idea about body language.

REFERENCES

[1] I. Arroyo, B. P. Woolf, W. Burelson, K. Muldner, D. Rai, and M. Tai. 2014. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education* 24, 4 (2014), 387–426.
[2] K. C. Berridge. 1996. Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioural Reviews* 20 (1996), 1–25.
[3] E. Billauer. 2012. Peak detection using MATLAB. (2012). http://www.billauer.co.il/peakdet.html
[4] C. Crivelli, P. Carrera, and J. M. Fernández-Dols. 2015. Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evolution and Human Behavior* 36, 1 (2015), 52–58.
[5] A. J. Fridlund. 2017. The behavioral ecology view of facial displays: 25 years later. In *The Science of Facial Expression*, J.-M. Fernandez-Dols and J.A. Russell (Eds.). Oxford University Press, Oxford, 77–92.
[6] M. E. Hoque, D. J. McDuff, and R. W. Picard. 2012. Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Transactions on Affective Computing* 3, 3 (2012), 323–334.
[7] S. Jaiswal and M. Valstar. 2016. Deep learning the dynamic appearance and shape of facial action units. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 1–8.
[8] M. Leijten and L. Van Waes. 2013. Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication* 30, 3 (2013), 358–392. https://doi.org/10.1177/0741088313491692
[9] M. Mehu and R. I. Dunbar. 2008. Naturalistic observations of smiling and laughter in human group interactions. *Behaviour* 145, 12 (2008), 1747–1780.
[10] F. M. Merchant and A. A. Armoundas. 2012. Role of substrate and triggers in the genesis of cardiac alternans, from the myocyte to the whole heart: implications for therapy. *Circulation* 125, 3 (2012), 539–549.
[11] C. Nass, B. J. Fogg, and Y. Moon. 1996. Can computers be teammates? *International Journal of Human Computer Studies* 45 (1996), 669–678.
[12] B. Parkinson. 2005. Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review* 9, 4 (2005), 278–311.
[13] M. F. Valstar, I. Patras, and M. Pantic. 2005. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *IEEE Computer Science Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 76.
[14] H. J. Witchel, C. P. Santos, J. K. Ackah, C. E. Westling, and N. Chockalingam. 2016. Non-instrumental movement inhibition (NIMI) differentially suppresses head and thigh movements during screenic engagement: dependence on interaction. *Frontiers in Psychology* 7 (2016), 157.