UNIVERSITÄT BERN

Institut für Marketing und Unternehmensführung

Consumer Behavior

How Unconscious Perception Influences Consumer Behavior

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

Sabrina Stöckli

Bern, Juni 2017

Die Fakultät hat diese Arbeit am 24.08.2017 auf Antrag der beiden Gutachter Prof. Dr. Claude Messner und Prof. Dr. Arnd Florack als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

Contents

Acknowledgements

I would like to thank Prof. Dr. Claude Messner for his encouragement, support and guidance. His enthusiasm and advice in understanding consumer behavior as well as his knowledge and intuition in experimental designs and operationalization were always inspiring. I am particularly thankful for all the method courses, conferences and equipment Claude granted me.

Further, I would like to express my gratitude towards Prof. Dr. Arnd Florack for kindly agreeing to review my doctoral thesis. Thank you for the valuable time and expertise.

A special thank you goes to all my colleagues from the University of Bern, namely Dr. Michael Dorn, Dr. Aline Stämpfli, Dr. Michael Schulte-Mecklenbeck, Dr. Adrian Brügger, Bettina Höchli, Stefan Borer, Stefan Liechti, Eva Niklaus, Geraldine Neeser, Michele Steffen, Eva Meyer, Daniela Lüdi, Evi Ackermann and Lukas Messerli. The discussions and cooperation within the Consumer Behavior Team were always stimulating, motivating and advanced my projects essentially.

Finally, I am grateful for my family and friends: Barbara, Niklaus, Raffael, Joël, Michelle, Martha, Kathrin, Dana, Tabea, Christine S., Aby, Nina, Lenna, Alex, Sharon, Sabrina, Martin, David, Christine G., Verica and Deborah. Their support, love, different perspectives and ideas were always inspiring and encouraging. I dedicate this thesis to them.

Bern, Juni 2017

Sabrina Stöckli

Introduction

This thesis consists of three papers addressing the role of unconscious perception in consumer behavior. Paper I and Paper II add to research on the effect of unconscious perception of environmental cues on food consumption. Paper III focuses on the methodological challenge to measure unconscious psychological processes. Together, these papers contribute to increasing research on how processes that occur outside of conscious awareness influence consumers' emotions, intentions and behavior (e.g., Chartrand, 2005; Dijksterhuis, Smith, van Baaren, & Wigboldus, 2005).

There are various factors in the environment that can activate processes, which automatically lead to changes in consumers' emotions, intentions and behavior. Given the practical relevance of effects of unconscious perception of environmental cues, it is not only important to replicate and test these effects in the field, but also to gain insight into the underlying processes. Paper I and Paper II find their origin in the "Giacometti effect". According to this effect, people consume less food when being exposed to unobtrusive thin sculptures by the artist Alberto Giacometti (Brunner & Siegrist, 2012). While Paper I tests the "Giacometti effect" in the field, Paper II examines the effect's underlying processes.

In order to understand how unconscious perception influences consumer behavior, it is vital to measure unconscious processes. Yet, validly measuring relevant dependent variables such as emotions, intentions and behavior is challenging. Consumers are often not willing or able to provide accurate self-reports, for example, due to social desirability or introspective inaccessibility (Fisher, 1993; Tourangeau & Yan, 2007). To measure variables that influence consumer behavior in an unintentional, resource-independent, unconscious or uncontrollable manner, it is often suggested to adopt indirect measurement procedures (see Gawronski & de Houwer, 2014; Wilson, Lindsey, & Schooler, 2000). Indirect measurement procedures are particularly suitable when measuring emotions (see Paulhus & Reid, 1991; Robinson & Clore, 2002; Scherer, 2005; Watson, 2000). This becomes evident, for instance, when attempting to better understand the "face of need effect". According to this effect, consumers give more to charity when the child on the perceived charity ad is sad-faced versus neutral-faced or happy-faced (Small & Verrochi, 2009). Paper III finds its origin in the claim that examining the "face of need effect" requires a valid measurement of unconscious emotional processes (see Stöckli & Stämpfli, 2016). Namely, Paper III focuses on the methodological challenge to validly measure emotions in consumers' faces.

Research on the influence of unconscious perception on consumer behavior is afflicted with diverse challenges. The three papers of this thesis address three of them. First, a certain effect of unconscious perception needs to be repeatedly evidenced, both in the laboratory and the field. Correspondingly, Paper I extends evidence on the "Giacometti effect" by initially testing the influence of subtle environmental cues on food consumption in the field. Second, underlying processes need to be examined. Accordingly, Paper II examines the priming processes that automatically drive the "Giacometti effect".[1] Third, relevant dependent variables need to be identified and validly measured. Initiated by the attempt to examine the "face of need effect", Paper III focuses on validating a measurement of consumers' emotions, namely automated facial expression analysis.

The Effect of Unobtrusive Environmental Cues on Consumers' Food Choices in the Field

In Paper I (i.e., Stöckli, Stämpfli, Messner, & Brunner, 2016), we conducted two field studies that showed the effect of subtle environmental cues in form of posters on the choices of

[1] Note that Paper I and Paper II were conducted as part of the National Research Program NRP 69 (Healthy Nutrition and Sustainable Food Production) and financially supported by the Swiss National Science Foundation (grant 12 No. 145189).

consumers regarding vending machine food. Thus it was examined for the first time whether the "Giacometti effect" occurs in a natural setting. Study 1 applied a one-factorial within-subjects design with four different poster conditions (nature vs. activity vs. fun fair vs. no poster). Food choices were retrospectively measured on the basis of sales numbers. Results showed that consumers exposed to a nature poster compared to a fun fair poster or no poster were more likely to buy healthy (vs. unhealthy) snacks. Consumers were also more likely to buy healthy snacks when exposed to an activity poster than when exposed to a fun fair poster. Study 2 replicated this effect, however, the nature poster was replaced by a poster of the thin, human-like Giacometti sculptures. Results revealed that consumers were more likely to buy healthy (vs. unhealthy) snacks when exposed to a Giacometti poster than when exposed to a fun fair poster or to no poster. Hence, study 2 supported the "Giacometti effect" found in study 1. Overall, Paper I replicated the "Giacometti effect" for the first time in the field and thus underpins the real-world relevance of environmental cues in helping consumers to automatically make healthy food choices.

Examining Underlying Priming Processes of the Effect of Environmental Cues

In Paper II (i.e., Stämpfli, Stöckli, & Brunner, 2017), we contributed to the understanding of the influence of environmental cues on food consumption by examining the underlying processes of the "Giacometti effect". We conducted two laboratory studies to test whether the "Giacometti effect" is driven by health-related or by weight-related priming processes. Therefore, study 1 examined the Giacometti cue's effect on healthy and unhealthy food consumption. Results showed that the Giacometti cue reduced food intake independent of how healthy the food was. This suggests that the "Giacometti effect" is driven by weight-related rather than by health-related processes. In addition, results showed that the "Giacometti effect" was moderated by restrained eating, that is, the tendency of chronically pursuing the goal of controlling one's body weight

(Herman & Mack, 1975; Stroebe, Mensink, Aarts, Schut, & Kruglanski, 2008). Specifically, restrained (vs. non-restrained) eaters consumed less after being exposed to the Giacometti cue. Study 2 used a fragmented word task with health-related and weight-related words to test whether health-related or weight-related priming processes drive the "Giacometti effect". Results showed that restrained (vs. non-restrained) eaters completed more weight-related words after being exposed to the Giacometti cue. Overall, Paper II contributed to the understanding of the underlying processes of the "Giacometti effect" by showing that the Giacometti sculptures are not a health-related, but a weight-related cue and thus are most effective for restrained (vs. non-restrained) eaters.

Validating Automated Facial Expression Analysis as Measure of Consumers' Emotions

In Paper III (i.e., Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2017a), we addressed the methodological challenge of validly measuring emotions in people's faces. This paper emerged from a project that aimed to shed light on the underlying processes of the "face of need effect". In detail, this project sought to test whether people overtake different emotional states (happiness, neutrality or sadness) expressed by faces in charity ads and thus also express these emotions by their own faces (see Stöckli & Stämpfli, 2016). Yet, measuring facial expressions turned out to be a major challenge. Conventional measurement instruments such as *Facial Electromyography* (fEMG) and *Facial Action Coding System* (FACS) have various drawbacks (e.g., labor and time intensity; see Ekman & Friesen, 1976; Hwang & Matsumoto, 2016). Although computer-based automated facial expression analysis constitutes a promising alternative, there is little scientific evidence on its validity (for exception see Lewinski, den Uyl, & Butler, 2014).

To close this gap within the literature, in Paper III, we tested the validity of iMotions' facial expression analysis (www.imotions.com), a software for automated facial expression analysis.

Specifically, we validated iMotions' AFFDEX and FACET, two independent software modules that are based on two distinct algorithms designed to analyze facial expressions. In study 1, database pictures of standardized emotional facial expressions were classified with both algorithms. Results showed a large variance in classification accuracy across emotions with a performance advantage for FACET over AFFDEX. In study 2, 110 respondents' facial responses were measured while being exposed to emotionally evocative database pictures. Again, results revealed that classification accuracy differs for distinct emotions, and that the classification accuracy is better for FACET than for AFFDEX. When contextualizing the classification accuracy of iMotions with the classification accuracy of conventional methods such as fEMG and human FACS coding, the results of these two studies suggest that automated facial expression analysis produces data with an acceptable degree of classification accuracy for prototypical facial expressions. Overall, Paper III showed that automated facial expression analysis allows the measurement of emotions in people's faces and that it is a potential alternative to conventional facial expression analysis methods such as fEMG. Given that consumer research and other research fields related to emotions have widely ignored automated facial expression analysis, we encourage researchers to advance this process measurement and therefore provide theoretical and technical guidance for the application of automated facial expression analysis (see Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2017b).

Conclusion

In conclusion, the three papers of this thesis addressed different challenges that are typical for research on the effect of unconscious perception on consumer behavior. In fact, this thesis has underpinned the need of replicating effects of unconscious perception on consumer behavior in both the laboratory and field (Paper I). Further, it has been demonstrated that a comprehensive understanding of effects of unconscious perception on consumer behavior requires examining underlying processes (Paper II). Therefore, identifying and validly measuring relevant dependent variables such as emotions expressed by consumers' faces is a core challenge (Paper III).

References

Brunner, T. A., & Siegrist, M. (2012). Reduced food intake after exposure to subtle weight-related cues. *Appetite*, *58*(3), 1109–1112.

Chartrand, T. L. (2005). The role of conscious awareness in consumer behavior. *Journal of Consumer Psychology*, *15*(3), 203–210.

Dijksterhuis, A., Smith, P. K., van Baaren, R. B., & Wigboldus, D. H. (2005). The unconscious consumer: Effects of environment on consumer behavior. *Journal of Consumer Psychology*, *15*(3), 193–202.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315.

Gawronski, B., & de Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York, NY: Cambridge University Press.

Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality*, *43*(4), 647–660.

Hwang, H. C., & Matsumoto, D. (2016). Facial expressions. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA handbook of nonverbal communication* (pp. 257–287). Washington, DC: American Psychological Association.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*(2), 307–317.

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*(1), 198–215.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of Marketing Research*, *46*(6), 777–787.

Stämpfli, A. E., Stöckli, S., & Brunner, T. A. (2017). A nudge in a healthier direction: How environmental cues help restrained eaters pursue their weight-control goal. *Appetite*, *110*, 94–102.

Stöckli S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017a). Facial expression analysis with AFFDEX and FACET: A validation study. Manuscript under revision at Behavior Research Methods.

Stöckli S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017b). Automated emotional facial expression analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide*. Chapter in preparation.

Stöckli, S., & Stämpfli, A. E. (2016, October). The role of mimicry in charity advertising. Poster session presented at North American Conference of the Association for Consumer Research, Berlin, Germany.

Stöckli, S., Stämpfli, A. E., Messner, C., & Brunner, T. A. (2016). An (un)healthy poster: When environmental cues affect consumers' food choices at vending machines. *Appetite*, *96*, 368–374.

Stroebe, W., Mensink, W., Aarts, H., Schut, H., & Kruglanski, A. W. (2008). Why dieters fail: Testing the goal conflict model of eating. *Journal of Experimental Social Psychology*, *44*(1), 26–36.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Watson, D. (2000). *Mood and temperament*. New York, NY: Guilford Press.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126.

Paper I

An (Un)healthy Poster:

When Environmental Cues Affect Consumers' Food Choices at Vending Machines

Stöckli, S., Stämpfli, A. E., Messner, C., & Brunner, T. A. (2016). An (un)healthy poster: When environmental cues affect consumers' food choices at vending machines. *Appetite, 96,* 368–374. doi:10.1016/j.appet.2015.09.034

Appetite 96 (2016) 368–374

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

An (un)healthy poster: When environmental cues affect consumers' food choices at vending machines

Sabrina Stöckli [a, *], Aline E. Stämpfli [a, b], Claude Messner [a], Thomas A. Brunner [b]

[a] University of Bern, Institute of Marketing and Management, Department Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland
[b] Bern University of Applied Sciences, School of Agricultural, Forest and Food Sciences HAFL, Food Science and Management, Länggasse 85, 3052 Zollikofen, Switzerland

A R T I C L E I N F O

Article history:
Received 9 April 2015
Received in revised form
26 September 2015
Accepted 28 September 2015
Available online 22 October 2015

Keywords:
Environmental cue
Healthy nutrition
Food choice
Priming
Mental concept

A B S T R A C T

Environmental cues can affect food decisions. There is growing evidence that environmental cues influence how much one consumes. This article demonstrates that environmental cues can similarly impact the healthiness of consumers' food choices. Two field studies examined this effect with consumers of vending machine foods who were exposed to different posters. In field study 1, consumers with a health-evoking nature poster compared to a pleasure-evoking fun fair poster or no poster in their visual sight were more likely to opt for healthy snacks. Consumers were also more likely to buy healthy snacks when primed by an activity poster than when exposed to the fun fair poster. In field study 2, this consumer pattern recurred with a poster of skinny Giacometti sculptures. Overall, the results extend the mainly laboratory-based evidence by demonstrating the health-relevant impact of environmental cues on food decisions in the field. Results are discussed in light of priming literature emphasizing the relevance of preexisting associations, mental concepts and goals.

Apple or chocolate? Water or lemonade? Food choices seem to be among the simplest decisions. However, as today's global obesity epidemic reveals, food decisions are complex and driven by many factors (Köster, 2009; World Health Organization, 2014). Food decisions depend on motives such as hunger, pleasure, sociability, weight control, and health (Renner, Sproesser, Strohbach, & Schupp, 2012) but also on our environment. Environmental cues can influence *how much* people eat (e.g., Brunner, 2010). A poster promoting a slim figure can result in a reduced amount of test snacks being eaten (Papies & Hamstra, 2010). More subtly, consumers eat less chocolate when there is a body-weight scale or a picture of skinny human-like Giacometti sculptures in the laboratory (Brunner, 2010; Brunner & Siegrist, 2012).

This paper is not focused on the influence of environmental cues on the amount of consumption but on the influence on consumption choice. The main question is whether environmental cues can influence the preference for healthy or unhealthy food. Two field studies with vending machines examine whether posters with a health-evoking but not directly food-related image lead to more sales of healthy over unhealthy foods compared to no poster or a poster with a hedonic-evoking motif.

1. When environmental cues determine food decisions

The homeostatic system is an internal signal structure that maintains a person's energy homeostasis. In contrast, the non-homeostatic system is driven by environmental cues. It is well known that the latter system particularly facilitates overindulgence and adds to today's increasing obesity rates (Berthoud, 2006; Hill & Peters, 1998; Seeley & Woods, 2003; Wadden, Brownell, & Foster, 2002).

Environmental cues can affect *how much* and *what* people eat or drink (see Wansink, 2004; for a review). A poster promoting a slim figure, or an experimenter stating that chocolate makes people happy but fat, decrease consumption volume (Brunner, 2010; Papies & Hamstra, 2010) and fruit odor brings people to prefer meals with fruits or vegetables (Gaillet, Sulmont-Rossé, Issanchou, Chabanet, & Chambaron, 2013). As shown by these examples, environmental cues influence people in various manifestations. Besides sensory cues such as odors (Gaillet et al., 2013), light (Areni & Kim, 1994), and music (North, Hargreaves, & McKendrick, 1999), normative cues such as ideal weight-reminders (Brunner, 2010; Papies & Hamstra, 2010), plate size (Van Ittersum & Wansink,

* Corresponding author.
E-mail addresses: Sabrina.Stoeckli@imu.unibe.ch (S. Stöckli), Aline.Staempfli@imu.unibe.ch (A.E. Stämpfli), Claude.Messner@imu.unibe.ch (C. Messner), thomas.brunner@bfh.ch (T.A. Brunner).

http://dx.doi.org/10.1016/j.appet.2015.09.034

2012), and eating companions (Brunner, 2010; Herman Koenig-Nobert, Peterson, & Polivy, 2005) also influence eating. For example, chocolate tasters tend to eat as much or as little as their companions do. Since our environment is full of cues, these can overpower each other. This becomes evident when the chocolate tasters of the previous illustration are additionally influenced by a body-weight scale and thus eat little regardless of their companion's food intake (Brunner, 2010). All in all, these examples imply that existing associations and concepts are considerable factors in determining which environmental cue is decisive in food choice situations (Bargh, 2006).

2. The role of priming, associations and concepts

Environmental cues can serve as primes, activate mental concepts and thereby influence decisions (see Bargh, 2006; for a review). As the example of reduced food intake after subtle exposure to Giacometti sculptures illustrates, applying the priming paradigm to the food consumption area has been proven to be effective (Brunner & Siegrist, 2012). Similar to Kay, Wheeler, Bargh, and Ross' (2004) activation of a competition concept by related objects (e.g., business suits), the Giacometti effect leads to the conclusion that body-weight or figure related cues activate corresponding concepts such as dieting. A similar mechanism occurs when the unconscious perception of fruit odor activates a fruit and vegetable concept and thus causes people to more frequently opt for meals with than without fruits or vegetables (Gaillet et al., 2013). Likewise, a diet recipe poster on a butcher's store door led customers with but not without a dieting goal to eat fewer test snacks, suggesting the activation of a motivational dieting concept (Papies & Hamstra, 2010). In line with priming research (see Bargh, 2006; for a review), these examples imply that environmental cues can activate various types of mental concepts such as traits, stereotypes, schemata or goals and thus induce a subsequent process or behavior (e.g., Bargh, 1990; Bargh & Gollwitzer, 1994; Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Custers & Aarts, 2005, 2007; Förster, Liberman, & Friedman, 2007; Sela & Shiv, 2009). Specific to the food consumption area, the idea of activating mental concepts by means of related environmental cues (see Bargh, 2006; for a review) implies that health-related or hedonic-related cues can increase the accessibility of healthy or hedonic dieting concepts. Notably, activating a certain diet-relevant concept requires a sufficiently strong associative cue-concept-link (Gaillet et al., 2013). Since people have both healthy and hedonic concepts, a person's association between an environmental cue and a health concept must be sufficiently strong in order to cause a healthy behavior.

In conclusion, two major generalizations can be drawn about the conditions under which environmental cues are effective in encouraging healthy food choices: First, a mental concept of healthy dieting should preexist. Second, the mental ties of these mental representations with the environmental cue must be sufficiently strong (cf. Gaillet et al., 2013).

Note that most of the existing findings stem from laboratory experiments (e.g., Brunner, 2010; Brunner & Siegrist, 2012; Harris, Bargh, & Brownell, 2009), but not from naturalistic settings (e.g., Papies & Hamstra, 2010). In order to learn more about the practical application of environmental cues, empirical attention should be given to the effectiveness of environmental cues in the field.

3. Research overview

The aim of the studies was to examine whether one's actual food choice can be influenced by environmental cues in a naturalistic setting. Contrary to previous studies, the present research focused on actual food choices (e.g., cookies vs. apples) rather than on quantitative decision aspects (e.g., one vs. two snacks). Whether specific snacks were presumed to be healthy or unhealthy was determined by people's perception rather than by actual nutritional value.

Two field studies tested the hypothesis that visual environmental cues with an associative link to (un)healthy food lead to choices of (un)healthy food alternatives. Both field studies were conducted using vending machines. The environmental cues were posters with a healthy or a hedonic motif, which were placed beside the vending machines. Thus, the field studies tested whether a poster with a healthy motif results in healthy snack choices whereas a poster with an unhealthy motif leads to unhealthy snack choices.

4. Field study 1

4.1. Method

4.1.1. Sample

A total of 634 snack purchases from vending machines at three locations of a European University of Applied Sciences were registered. Snack decisions were measured retrospectively by the amount of sales and not individually. Hence, no demographics for the sample were obtained. However, an independent survey with ten individuals at each of the three vending machine locations was conducted, giving a better notion of the sample. Data from these 30 individuals (25 females, $M_{age} = 29.60$, $SD_{age} = 9.77$) suggest that a campus sample can be assumed. Purchases of several non-food products (i.e., chewing gum or beverages) were excluded. Due to a daily altering assortment, sandwiches were also not considered for the analysis. The exclusion of these 106 purchases bypasses the ambiguity of classifying the products as healthy or unhealthy snacks. A final sample of 528 snack purchases remained for the analysis.

4.1.2. Design and procedure

The field study employed a one-factorial within-subjects design with four conditions. One of four options (a nature, activity, or fun fair poster or no poster) was placed next to the vending machine, with each option displayed for one fourth of the time. According to a Latin-square design, the four poster conditions were permuted and counterbalanced across the three test locations. In a period of one month (from April 28 to May 25, 2014), poster conditions were systematically combined with the vending machine at each test location in weekly intervals. Each poster-location combination was tested once.

Consumers were primed with a poster placed next to a vending machine. All testing occurred during the regular term time, that is, during a period with no exams or holidays and with a constant opening time. All snack purchases were registered in collaboration with the owner of the vending machines, a regional bakery. This was conducted in line with the usual and periodic vending machine restock. Specifically, the number of healthy and unhealthy snack purchases was subsequently measured by a daily inventory of snack sales for each condition and location.

4.1.3. Materials

4.1.3.1. Posters. In the treatment conditions, a nature poster, an activity poster, or a fun fair poster was placed in the consumers' visual line. No poster was placed for the control condition. All poster motifs were chosen by means of an exploratory pilot study. An accumulation of potential motifs was determined by unstructured single in-depth interviews with male and female students, asking what they associate with healthy and hedonic nutrition. All associations were collected in the form of key words. Frequently stated key words were discussed in expert groups. After that, two health-relevant and one hedonic motif were extracted. Specifically, nature and activity were considered to evoke associations with a healthy diet. In contrast, a fun fair was perceived as representative of hedonic needs.

The results of the pilot study served as the basis for the poster selection. Therefore the nature poster showed grassland, trees and a blue sky with clouds. In the activity condition there were running legs in sport shoes with asphalt in the background. In contrast, the fun fair poster showed two carousels with a summery blue sky in the background. All poster motifs were retrieved from an online database of free stock photos (http://www.fotocommunity.de/). Posters did not show text or food. The Web version of this paper offers an appendix with an example situational view of the test locations.

4.1.3.2. Snack choice. Consumers selected among a large variety of snacks (approximately 15 products). At each location, these were arranged and cooled in equally sized compartments on a ten-level carousel vending machine. The composition of the snack display was adapted for the field study. All vending machines offered a constant snack display. Since the snack offer was maintained in order to adhere to the usual naturalistic setting, individual vending machines offered a slightly different snack assortment. Likewise, snack displays in the individual vending machines and prices were held constant.

Each snack of the assortment was categorized as healthy or unhealthy. The categorization resulted from an independent pretest with 97 participants evaluating all snack alternatives from 1 (*very unhealthy*) to 7 (*very healthy*). Ratings were used to measure *snack choice*, that is, coding snack purchases as healthy ($M \geq 4$) or unhealthy ($M < 4$). The Web version of this paper offers an appendix with an overview of the snack display and perceived healthiness for all snacks.

4.2. Results and discussion

The trend of the results is in line with the hypothesis. Based on a 4 × 2 (poster [nature vs. activity vs. fun fair vs. control] × food [healthy vs. unhealthy]) contingency table, a chi-square test revealed an association between poster exposure and snack choice ($\chi^2(3, 528) = 10.45$, $p < .05$). While in the nature condition, 34% (or 46 out of 135) of the chosen snacks were healthy (vs. 66% unhealthy snacks), in the activity condition, 28% (or 33 out of 118) of the selected snacks were healthy (vs. 72% unhealthy). In contrast, the percentage of healthy snacks in the control condition was 22% (or 28 out of 128; vs. 78% unhealthy snacks) and in the fun fair condition 18% (or 27 out of 147; vs. 82% unhealthy snacks).

A post-hoc analysis with six separate 2 × 2 (poster × food [healthy vs. unhealthy]) comparisons revealed three significant relations, indicating poster-related differences in snack choices. As expected, when consumers were exposed to the nature poster rather than to the fun fair poster healthy snacks were chosen more frequently than unhealthy snacks ($\chi^2(1, 282) = 9.05$, $p < .01$). Furthermore, when consumers were exposed to the nature poster rather than to no poster healthy snacks were chosen more frequently than unhealthy snacks ($\chi^2(1, 263) = 4.84$, $p < .05$). Consistent with the hypothesis, when consumers were exposed to the activity poster rather than to the fun fair poster healthy snacks were selected marginally more frequently than unhealthy snacks ($\chi^2(1, 265) = 3.44$, $p = .06$). In line with the hypothesis, when consumers were exposed to either the nature poster or the activity poster no differences in the chosen amount of healthy or unhealthy snacks occurred ($\chi^2(1, 253) = 1.09$, $p = .30$). Likewise, when consumers were exposed to either the activity poster or to no poster no differences in the chosen amount of healthy or unhealthy snacks was observed ($\chi^2(1, 246) = 1.22$, $p = .27$). Finally, when consumers were exposed to the fun fair poster or to no poster no differences in the chosen amount of healthy or unhealthy snacks appeared ($\chi^2(1, 275) = 0.53$, $p = .47$).

Fig. 1 illustrates the frequencies of healthy and unhealthy snack choices in all poster conditions, accentuating those poster conditions that did not result in different influences on snack choice.

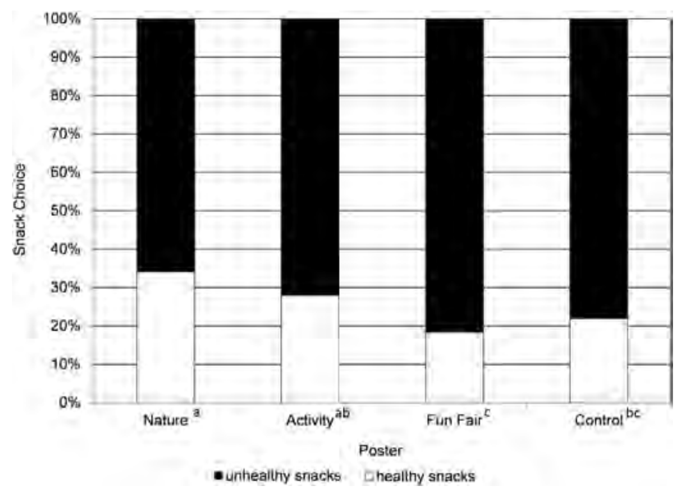

Fig. 1. Number of healthy versus unhealthy snack purchases in the four poster conditions of field study 1. Snack choices in the conditions marked with the same letter did not differ.

In conclusion, exposure to a poster affected the choice between healthy and unhealthy snacks. Importantly, the extent of this influence differed between the tested posters. Whereas individuals with a nature poster chose a healthy snack more often than individuals exposed to a fun fair or no poster, individuals in the activity condition were more likely to choose healthy snacks than those in the fun fair condition. No difference was found for consumers in the control condition compared with consumers primed by activity and fun fair posters. As assumed, consumers primed by the activity and nature poster did not differ in their snack choice. Overall, the trend of the results is in line with the hypothesis.

In order to improve generalizability and further demonstrate ecological validity, a replication study was designed. Thus, field study 2 aimed at replicating the previously observed effect, but with a different sample and an additional environmental cue.

While field study 1 was limited to the nature, activity and fun fair posters, field study 2 included an additional poster. Since the added poster showed the Giacometti sculptures, which are assumed to address a weight-related concept, consumers should theoretically respond by more frequently choosing healthy compared to unhealthy food. Thus, field study 2 examined whether the Giacometti effect is limited to quantitative food decisions as shown in research by Brunner and Siegrist (2012) or can be expanded to food choice. Furthermore, the validity of the impact of the Giacometti sculpture on food decisions was studied in a naturalistic setting.

5. Field study 2

5.1. Method

5.1.1. Sample

A total of 253 purchases from a vending machine at the National

Office of Public Health was registered. As in field study 1, the data gathering was conducted retrospectively by the amount of sales, and consumers were not individually registered and actively assigned to conditions. Thus, no demographics were available for the main study. However, an independent survey with 34 employees of the National Office of Public Health provided a rough estimate of the sample composition (21 females, $M_{age} = 42.53$, $SD_{age} = 10.34$). Similar to field study 1, one ambiguous purchasing option (soft drink produced from milk whey) was not included in the sample. Thus, a final sample of 252 snack purchases was included in the analysis.

5.1.2. Design and procedure

A similar design and procedure to that in field study 1 was used. That is, a one-factorial within-subjects design with four conditions was employed. One of four options (the Giacometti sculptures, activity or fun fair poster or no poster) was placed above the vending machine, with each option displayed for one fourth of the time. The poster conditions changed in weekly intervals for one month (from June 2 to June 30, 2014).

The primed consumers all chose from the provided food display under natural and constant environmental conditions. Their purchases were registered by means of the electronically controlled stock monitoring system of the vending machine operator. All transmitted data allowed the frequency of healthy and unhealthy purchases for each condition to be determined.

5.1.3. Materials

5.1.3.1. Posters. As in field study 1, the treatment conditions consisted of three posters, which were placed in consumers' visual line. No poster was used for the control condition. While the activity and fun fair posters remained identical to field study 1, the Giacometti motif was chosen on the strength of Brunner and Siegrist (2012) demonstration of its food-reducing influence.

5.1.3.2. Snack choice. Consumers selected among healthy and unhealthy foods. The arrangement and prices of all food were employed unmodified and held constant. To measure food choice as a dependent variable, each purchase was coded as either healthy or unhealthy. Based on the online pretest of field study 1, subcategories of healthy and unhealthy foods were extracted and used to either classify purchases as healthy (natural snacks and natural drinks) or unhealthy (chocolate, pastries, chips, and soft drinks). The Web version of this paper offers an appendix with an overview of all purchased food and its categorization.

5.2. Results and discussion

The results are in line with the hypothesis. As in field study 1, a chi-square test revealed a relation between poster exposure and food choice ($\chi^2(3, 252) = 16.94$, $p < .001$). The percentage of healthy snack purchases in the Giacometti condition was 58% (or 42 out of 73; vs. 42% unhealthy snacks), 44% (or 37 out of 85; vs. 56% unhealthy snacks) in the activity condition, 29% (or 19 out of 66; vs. 71% unhealthy snacks) in the fun fair condition and 21% (or 6 out of 28; vs. 79% unhealthy snacks) in the control condition.

Subsequently, a post-hoc analysis with separate 2 × 2 (poster × food [healthy vs. unhealthy]) comparisons revealed five significant relations, indicating poster-related differences in food choices. Specifically, consumers were found to be influenced in all health-related poster conditions. As expected, when consumers were exposed to the Giacometti poster rather than to the fun fair poster, healthy compared to unhealthy snacks were chosen more frequently ($\chi^2(1, 139) = 11.63$, $p < .001$). Likewise, when consumers were exposed to the Giacometti poster rather than to no poster, healthy snacks compared to unhealthy snacks were purchased more frequently ($\chi^2(1, 101) = 10.58$, $p = .001$). In line with the hypothesis, when consumers were exposed to the activity poster rather than to no poster, healthy snacks compared to unhealthy snacks were selected more frequently ($\chi^2(1, 113) = 4.6$, $p < .05$). Additionally, when consumers were exposed to the activity poster rather than to the fun fair poster, healthy snacks compared to unhealthy snacks were chosen marginally more frequently ($\chi^2(1, 151) = 3.46$, $p = .06$). When consumers were exposed to the Giacometti poster rather than to the activity poster healthy snacks compared to unhealthy snacks were chosen marginally more frequently ($\chi^2(1, 158) = 3.46$, $p = .08$).

Finally, when consumers were exposed to either the fun fair poster or to no poster, no difference in the amount of chosen healthy or unhealthy snacks occurred ($\chi^2(1, 94) = 0.54$, $p = .46$). The frequencies of healthy and unhealthy snack choices in all poster conditions are illustrated in Fig. 2.

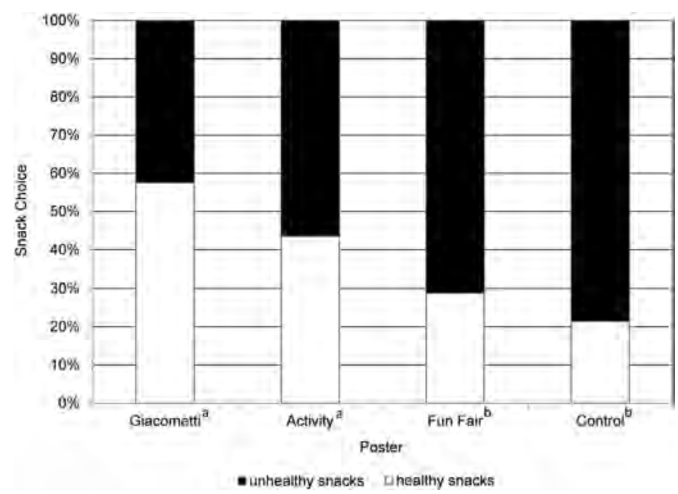

Fig. 2. Number of healthy versus unhealthy snack purchases in the four poster conditions of field study 2. Snack choices in the conditions marked with the same letter did not differ.

In conclusion, the results indicate that a poster can influence the choice between healthy and unhealthy foods and hence support the hypothesized effect. As in field study 1, the effects of poster images were substantial. That is, the three tested posters influenced food choices differently. Compared to the hedonic-related fun fair poster, the health-related Giacometti and activity posters were more likely to influence food choices.

6. General discussion

Environmental cues can affect food decisions (see Wansink, 2004; for a review). While laboratory research has repeatedly demonstrated the food-reducing or food-increasing influence of environmental cues (e.g., Brunner & Siegrist, 2012), less is known about environmental cues' impact on actual food decisions. The present studies asked whether environmental cues can direct consumers' real world food choices in favor of healthy or unhealthy snack alternatives. Overall, present findings are in line with priming research, revealing that environmental cues in the form of posters with content associated with (un)healthy diet influence food choices in favor of (un)healthy snacks. Specifically, while nature and activity posters increased preferences for healthy snacks, a fun fair poster increased preferences for unhealthy snacks (field study 1). The effect in favor of healthy food choices reoccurred for a poster with the Giacometti sculptures (field study 2). Moreover, findings contribute to previous research examining the impact of the environment on food decisions by indicating that environmental cues with only slight health or diet associations can be effective in influencing one's food decisions.

6.1. Growing evidence for environmental influences on food decisions

Past research has demonstrated that environmental cues such as Giacometti sculptures influence consumption volume decisions, that is, *how much* people eat (Brunner, 2010; Van Ittersum & Wansink, 2012). The present research adds to this by revealing that visual environmental cues are likewise able to influence food choices, that is, *what* people choose to eat. Thereby, evidence of the Giacometti effect was extended by showing that it appears not to be limited to reducing food intake but also occurs when choosing between healthy and unhealthy foods.

So far, the specific food consumption literature lacks a deep understanding of the effectiveness of environmental cues in shaping one's food decisions. This may be eliminated by an intensified consideration of more general findings of the priming research focusing on associations (e.g., Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000; Dijksterhuis & Smith, 2005), concepts (e.g., Bruner, 1957; Higgins, Rholes, & Jones, 1977), goals (e.g., Aarts, 2007; Bargh et al., 2001), and the level of awareness (e.g., Aarts, 2007; Chartrand, 2005).

6.2. The power of priming, associations, concepts and goals

The influence of environmental cues depends on associative links to a relevant concept (Aarts & Dijksterhuis, 2003). On one hand, the present findings imply an associative link between health and nature, activity, or Giacometti sculptures. On the other hand it supports the idea of an associative link between indulgence and a fun fair.

Regarding the stronger tendency to opt for healthy snacks in the nature compared to the activity condition as observed in field study 1, two explanations may be proposed. First, the nature compared to the activity poster might have been more likely to activate a healthy diet concept because of its stronger health-relevant associations. Second, it may be possible that both posters activated a diet concept, but that the nature concept more intensively fostered a healthy diet.

In respect to the importance of the activated associations, one can likewise argue that concepts other than a health or hedonic concept became activated. For instance, the Giacometti sculptures in field study 2 could have activated a body-weight concept (Brunner & Siegrist, 2012). However, it intuitively seems that a body-weight concept and health concept share some associations. It is also possible that the Giacometti sculptures primed other concepts such as culture or art. Nevertheless, this seems unlikely since there is no reason to expect an association between culture or art and eating behavior. Of course, this is speculative and future research should include the diversity of activatable associations.

Note that although the influence of environmental cues on food choices by priming seems plausible, the present study did not explicitly test this. Future research could apply lexical decision tasks (Wyer & Srull, 1989) to prove whether environmental cues effectively cause priming effects. By way of example, individuals exposed to nature, activity, or Giacometti posters are assumed to be faster in recognizing health-relevant words than individuals primed by a fun fair or no poster. Similarly, they should be more likely to complete word fragments with health-relevant instead of hedonic-relevant words (Kay et al., 2004; Tulving, Schacter, & Stark, 1982). This line of research could thus enable researchers to specify activated associations and to further investigate underlying processes.

While it is a general characteristic of mental concepts (e.g., traits, stereotypes, schemata and goals) to become activated by environmental cues, goals possess the specific characteristic of opening the doors for such priming effects by increasing the accessibility of environmental cues (Aarts, 2007; Aarts et al., 2005; Aarts, Dijksterhuis, & de Vries, 2001; Bargh et al., 2001; Bargh, 2006; Bruner, 1957; Custers & Aarts, 2010). In field study 2, health-related posters were more effective than a hedonic-related poster in influencing consumers' choices. Since data was collected in a health context, one could speculate that this is due to the health sensitivity of the sample. Indeed, National Office of Public Health employees are constantly concerned with matters of health. Assuming that the sample has a prevailing health goal, one could speculate that the accessibility of health-relevant cues is relatively likely. In fact, this would conform to the pattern showing that health-motivated people are relatively likely to respond to health-related cues (Fedoroff, Polivy, & Peter Herman, 2003; Herman et al., 2005). Here, the idea of selective attention is pivotal. According to this, one's active goal causes selective perception of goal-relevant environmental cues which make corresponding associations more likely to be activated (Aarts et al., 2001; Bruner, 1957).

Interestingly, when considering the influences of the posters in the two field studies, the activity poster appeared to be more effective in field study 2 than in field study 1. Specifically, in field study 2, differing consumer responses were found between the activity and control poster as well as between the activity and fun fair poster. Meanwhile, in field study 1, differing consumer responses only occurred between the activity and fun fair poster, but not between the activity and control poster. A speculative interpretation of this might be that the sample of field study 2 compared to the sample of field study 1 is more likely to pursue a health goal and thus to respond to a health-relevant cue. This study was not meant to analyze whether the different extent of the influence of the activity poster mirrors the interaction of environmental cues and consumers' goals. Still, this issue seems to be of practical relevance and further academic interest.

Considering that one might have both a health and a hedonic goal, one could speculate that people are influenced by environmental cues in terms of a compromise. For example, when primed by a nature poster, one may be more likely to choose the healthier chocolate cereal bar than the unhealthier pure chocolate bar and thus partially cope with both health and hedonic goals. Future research might incorporate this aspect in the definition of the dependent variable by measuring the effect on another level of extremeness. For example, participants could be faced with the choice between a relatively healthy cereal-chocolate bar and a pure, unhealthier chocolate bar instead of the usual healthy apple and the unhealthy chocolate bar.

6.3. Conscious versus unconscious influences of environmental cues

While both consciously and unconsciously perceived cues can trigger reactions (Bargh, 2006; Schacter, Chiu, & Ochsner, 1993), the influence of environmental cues is primarily thought to occur outside of one's awareness (Dijksterhuis et al., 2000). People are usually not consciously aware of the presence of environmental cues and almost never aware of the unconsciously activated processes or behavior patterns (Chartrand, 2005). For example, Giacometti-primed people do not consciously perceive that their food intake has been influenced (Brunner & Siegrist, 2012).

Importantly, the influence of environmental cues can vanish when cues are consciously perceived, but not when they are unconsciously perceived. For instance, when seeing a photo with a bottle of mineral water in the background, people are less likely to subsequently choose a bottle of the same brand when remembering the bottle compared to when not remembering it (Ferraro, Bettman, & Chartrand, 2009). This relates to the fact that whereas consciously perceived cues are controllable, unconsciously perceived cues are not (Chartrand, 2005; Daza, Ortells, & Noguera, 2007). Therefore, the influence of environmental cues on food decisions seems to be more effective when underlying processes are implicit, unconscious, and automatic rather than planned, conscious, and rational (Brunner & Siegrist, 2012; Ferraro et al., 2009). It is meaningful that the manipulation in the present studies was done in an unobtrusive rather than in a clearly unconscious way. Hence, it is unclear to what degree the cue's impact occurred outside of conscious awareness. The present results must, therefore, be interpreted with caution regarding underlying processes of the found effects. Future field research could test to what level of consciousness subtle environmental cues are perceived and processed and whether the effectiveness is higher when occurring on an unconscious rather than on a conscious level. Since unconscious influence requires nearly no cognitive resources, researchers could compare people's response to the exposure of cues with and without additional cognitive load (e.g., remembering numbers). Additional cognitive load reduces one's cognitive resources, and thus the ability to consciously perceive environmental cues declines. If the influence of environmental cues is conscious and thus cognitively costly, then the attenuated cognitive resources should lead to a reduced or no effect (McFerran, Dahl, Fitzsimons, & Morales, 2010). As the influence of environmental cues is assumed to be mainly unconscious, an influence irrespective of additional cognitive load seems more likely (McFerran et al., 2010; Uhlmann, Pizarro, & Bloom, 2008).

6.4. Practical relevance

The limited success of public awareness campaigns and health warnings to prevent obesity requires additional measures to damp strong affective impulses such as temptation and self-control conflicts (Downs, Loewenstein, & Wisdom, 2009; Fedoroff et al., 2003; Fishbach & Shah, 2006). As the present research proposes, more subtle approaches can be meaningful. That is, environmental cues can be applied as interventions for policy makers to shape people's behavior toward a healthier diet. Clearly, one may question the marginal benefit of choosing yogurt instead of a Twix bar. Indeed, it seems implausible that somewhat more than 100 calories can improve society's health and lower healthcare costs. Nevertheless, an accumulation of effective applications of environmental cues can help people to shed their extra calories and moreover to obtain the required intake of healthy nutrients over a long time.

Health professionals are therefore encouraged to complement current measures with subtle, less cognitive strategies. This can include structuring public places such as stores, restaurants, or schools. In doing so, literature suggests that monitoring and directing people's health-related and diet-relevant associations, concepts and goals are essential and an effective way of support a healthy diet. Future research should provide a more comprehensive understanding of the interaction of associations, concepts, goals and the level of awareness of subtle environmental cues in the food consumption area and thus help to implement more efficient health measures.

7. Conclusion

The present research indicates that environmental cues influence food choices in a naturalistic setting. While posters with an associative link to health lead to an increased choice of healthy food, environmental cues with an associative hedonic link increase the choice of unhealthy food. Overall, these findings offer a basis for improving society's food-related health.

Acknowledgments

This research was financially supported as part of the National Research Program NRP69—Healthy Nutrition and Sustainable Food Production, by Grant 12 No. 145189 from the Swiss National Science Foundation. We are grateful to Eva Meyer. She helped to conduct field study 1 for her Master's theses, which was submitted to the University of Bern.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.appet.2015.09.034.

References

Aarts, H. (2007). Health and goal-directed behavior: the nonconscious regulation and motivation of goals and their pursuit. *Health Psychology Review, 1*(1), 53–82. http://dx.doi.org/10.1080/17437190701485852.

Aarts, H., Chartrand, T. L., Custers, R., Danner, U., Dik, G., Jefferis, V. E., et al. (2005). Social stereotypes and automatic goal pursuit. *Social Cognition, 23*(6), 465–490. http://dx.doi.org/10.1521/soco.2005.23.6.465.

Aarts, H., & Dijksterhuis, A. (2003). The silence of the library: environment, situational norm, and social behavior. *Journal of Personality and Social Psychology, 84*(1), 18. http://dx.doi.org/10.1037/0022-3514.84.1.18.

Aarts, H., Dijksterhuis, A., & de Vries, P. (2001). On the psychology of drinking: being thirsty and perceptually ready. *British Journal of Psychology, 92*(4), 631–642. http://dx.doi.org/10.1348/000712601162383.

Areni, C. S., & Kim, D. (1994). The influence of in-store lighting on consumers' examination of merchandise in a wine store. *International Journal of Research in Marketing, 11*(2), 117–125. http://dx.doi.org/10.1016/0167-8116(94)90023-X.

Bargh, J. A. (1990). Goal and intent: goal-directed thought and behavior are often unintentional. *Psychological Inquiry, 1*(3), 248–251. http://dx.doi.org/10.1207/s15327965pli0103_14.

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36*(2), 147–168. http://dx.doi.org/10.1002/ejsp.336.

Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: automatic and strategic contingencies between situations and behavior. *Nebraska Symposium on Motivation, 41*, 71–124.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*(6), 1014. http://dx.doi.org/10.1037/0022-3514.81.6.1014.

Berthoud, H.-R. (2006). Homeostatic and non-homeostatic pathways involved in the control of food intake and energy balance. *Obesity, 14*(S8), 197S–200S. http://dx.doi.org/10.1038/oby.2006.308.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*(2), 123. http://dx.doi.org/10.1037/h0043805.

Brunner, T. A. (2010). How weight-related cues affect food intake in a modeling situation. *Appetite, 55*(3), 507–511. http://dx.doi.org/10.1016/j.appet.2010.08.018.

Brunner, T. A., & Siegrist, M. (2012). Reduced food intake after exposure to subtle weight-related cues. *Appetite, 58*(3), 1109–1112. http://dx.doi.org/10.1016/j.appet.2012.03.010.

Chartrand, T. L. (2005). The role of conscious awareness in consumer behavior. *Journal of Consumer Psychology, 15*(3), 203–210. http://dx.doi.org/10.1207/s15327663jcp1503_4.

Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: on the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89*(2), 129. http://dx.doi.org/10.1037/0022-3514.89.2.129.

Custers, R., & Aarts, H. (2007). In search of the nonconscious sources of goal pursuit: accessibility and positive affective valence of the goal state. *Journal of Experimental Social Psychology, 43*(2), 312–318. http://dx.doi.org/10.1016/j.jesp.2006.02.005.

Custers, R., & Aarts, H. (2010). The unconscious will: how the pursuit of goals operates outside of conscious awareness. *Science, 329*(5987), 47–50. http://dx.doi.org/10.1126/science.1188595.

Daza, M. T., Ortells, J. J., & Noguera, C. (2007). Negative semantic priming from consciously vs. unconsciously perceived single words. *Psicologica: International Journal of Methodology and Experimental Psychology, 28*(2), 105–127.

Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental*

Social Psychology, 36(5), 531–544. http://dx.doi.org/10.1006/jesp.2000.1427.
Dijksterhuis, A., & Smith, P. K. (2005). What do we do unconsciously? And how? *Journal of Consumer Psychology (Lawrence Erlbaum Associates), 15*(3), 225–229. http://dx.doi.org/10.1207/s15327663jcp1503_7.
Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for promoting healthier food choices. *The American Economic Review*, 159–164. http://dx.doi.org/10.1257/aer.99.2.159.
Fedoroff, I., Polivy, J., & Peter Herman, C. (2003). The specificity of restrained versus unrestrained eaters' responses to food cues: general desire to eat, or craving for the cued food? *Appetite, 41*(1), 7–13. http://dx.doi.org/10.1016/S0195-6663(03)00026-6.
Ferraro, R., Bettman, J. R., & Chartrand, T. L. (2009). The power of strangers: the effect of incidental consumer brand encounters on brand choice. *Journal of Consumer Research, 35*(5), 729–741. http://dx.doi.org/10.1086/592944.
Fishbach, A., & Shah, J. Y. (2006). Self-control in action: Implicit dispositions toward goals and away from temptations. *Journal of Personality and Social Psychology, 90*(5), 820. http://dx.doi.org/10.1037/0022-3514.90.5.820.
Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: a systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11*(3), 211–233. http://dx.doi.org/10.1177/1088868307303029.
Gaillet, M., Sulmont-Rossé, C., Issanchou, S., Chabanet, C., & Chambaron, S. (2013). Priming effects of an olfactory food cue on subsequent food-related behaviour. *Food Quality and Preference, 30*(2), 274–281. http://dx.doi.org/10.1016/j.foodqual.2013.06.008.
Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health Psychology, 28*(4), 404–413. http://dx.doi.org/10.1037/a0014399.
Herman, C. P., Koenig-Nobert, S., Peterson, J. B., & Polivy, J. (2005). Matching effects on eating: do individual differences make a difference? *Appetite, 45*(2), 108–109. http://dx.doi.org/10.1016/j.appet.2005.03.013.
Higgins, T. E., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141–154. http://dx.doi.org/10.1016/S0022-1031(77. http://dx.doi.org/10.1016/S0022-1031(77)80007-3.
Hill, J. O., & Peters, J. C. (1998). Environmental contributions to the obesity epidemic. *Science, 280*(5368), 1371–1374. http://dx.doi.org/10.1126/science.280.5368.1371.
Kay, A. C., Wheeler, S. C., Bargh, J. A., & Ross, L. (2004). Material priming: the influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes, 95*(1), 83–96. http://dx.doi.org/10.1016/j.obhdp.2004.06.003.
Köster, E. P. (2009). Diversity in the determinants of food choice: a psychological perspective. *Food Quality and Preference, 20*(2), 70–82. http://dx.doi.org/10.1016/j.foodqual.2007.11.002.
McFerran, B., Dahl, D. W., Fitzsimons, G. J., & Morales, A. C. (2010). I'll have what she's having: effects of social influence and body type on the food choices of others. *Journal of Consumer Research, 36*(6), 915–929. http://dx.doi.org/10.1086/644611.
North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology, 84*(2), 271. http://dx.doi.org/10.1037/0021-9010.84.2.271.
Papies, E. K., & Hamstra, P. (2010). Goal priming and eating behavior: enhancing self-regulation by environmental cues. *Health Psychology, 29*(4), 384. http://dx.doi.org/10.1037/a0019877.
Renner, B., Sproesser, G., Strohbach, S., & Schupp, H. T. (2012). Why we eat what we eat. The Eating Motivation Survey (TEMS). *Appetite, 59*(1), 117–128. http://dx.doi.org/10.1016/j.appet.2012.04.004.
Schacter, D. L., Chiu, C. Y. P., & Ochsner, K. N. (1993). Implicit memory: a selective review. *Annual Review of Neuroscience, 16*(1), 159–182. http://dx.doi.org/10.1146/annurev.ne.16.030193.001111.
Seeley, R. J., & Woods, S. C. (2003). Monitoring of stored and available fuel by the CNS: implications for obesity. *Nature Reviews Neuroscience, 4*(11), 901–909. http://dx.doi.org/10.1038/nrn1245.
Sela, A., & Shiv, B. (2009). Unraveling priming: when does the same prime activate a goal versus a trait? *Journal of Consumer Research, 36*(3), 418–433. http://dx.doi.org/10.1086/598612.
Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(4), 336. http://dx.doi.org/10.1037/0278-7393.8.4.336.
Uhlmann, E. L., Pizarro, D. A., & Bloom, P. (2008). Varieties of social cognition. *Journal for the Theory of Social Behaviour, 38*(3), 293–322. http://dx.doi.org/10.1111/j.1468-5914.2008.00372.x.
Van Ittersum, K., & Wansink, B. (2012). Plate size and color suggestibility: the Delboeuf Illusion's bias on serving and eating behavior. *Journal of Consumer Research, 39*(2), 215–228. http://dx.doi.org/10.1086/662615.
Wadden, T. A., Brownell, K. D., & Foster, G. D. (2002). Obesity: responding to the global epidemic. *Journal of Consulting and Clinical Psychology, 70*(3), 510. http://dx.doi.org/10.1037/0022-006X.70.3.510.
Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition, 24*, 455–479. http://dx.doi.org/10.1146/annurev.nutr.24.012003.132140.
World Health Organization. Controlling the global obesity epidemic. Accessed 08.12.14, Available at: http://www.who.int/nutrition/topics/obesity/en/.
Wyer, R. S., Jr., & Srull, T. K. (1989). *Memory and cognition in its social context* Accessed 8 December 2014, Available at: http://psycnet.apa.org/psycinfo/1989-98597-000.

Paper II

A Nudge in a Healthier Direction:

How Environmental Cues Help Restrained Eaters Pursue Their Weight-Control Goal

Stämpfli, A. E., Stöckli, S., & Brunner, T. A. (2017). A nudge in a healthier direction: How environmental cues help restrained eaters pursue their weight-control goal. *Appetite*, *110*, 94-102. doi: 10.1016/j.appet.2016.11.037

Appetite 110 (2017) 94–102

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

A nudge in a healthier direction: How environmental cues help restrained eaters pursue their weight-control goal

Aline E. Stämpfli [a, b, *], Sabrina Stöckli [a], Thomas A. Brunner [b]

[a] University of Bern, Institute of Marketing and Management, Department of Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland
[b] Bern University of Applied Sciences, School of Agricultural, Forest, and Food Sciences HAFL, Food Science and Management, Länggasse 85, 3052, Zollikofen, Switzerland

ARTICLE INFO

Article history:
Received 5 July 2016
Received in revised form
25 November 2016
Accepted 28 November 2016
Available online 30 November 2016

Keywords:
Environmental cue
Restrained eating
Weight-control goal
Dieting
Nudging

ABSTRACT

Losing weight is a goal for many people, but it is hard to pursue. However, dieting cues in the environment hold promise for improving individuals' eating behavior. For example, exposure to thin, human-like sculptures by the artist Alberto Giacometti has been found to promote healthy snack choices at a vending machine. Whether health- or weight-related processes drive such effects has not yet been determined. However, a detailed understanding of the content-related drivers of environmental cues' effects provides the first indications regarding a cue's possible use. Therefore, two laboratory studies were conducted. They examined the Giacometti sculptures' effects on unhealthy and healthy food intake (Study 1) and on the completion of weight- and health-related fragmented words (Study 2). Study 1 indicated that the sculptures are weight-related by showing that they reduced food intake independent of food healthiness. Furthermore, the "Giacometti effect" was moderated by restrained eating. Restrained eaters, who are known for their weight-control goal, ate less after having been exposed to the thin sculptures. The results of Study 2 pointed in the same direction. Restrained eaters completed more weight-related words after being exposed to the sculptures. Overall, these studies suggest that the thin sculptures are primarily weight-related cues and particularly helpful for restrained eaters. Environmental weight-control cues such as the Giacometti sculptures could act as a counterforce to our obesogenic environment and help restrained eaters pursue their weight-control goal. In this way, they could nudge food decisions in a healthier direction.

We eat what we eat particularly because we like it (Renner, Sproesser, Strohbach, & Schupp, 2012). In our "obesogenic" environment, with its abundance of tasty, high-calorie food, our goal of eating enjoyment gets constantly activated. As a result, we eat too much energy-dense food (Berthoud, 2006; Papies, 2016; Papies, Potjes, Keesman, Schwinghammer, & van Koningsbruggen, 2014; Renner et al., 2012). This contributes to the global obesity epidemic (World Health Organization, 2016). However, just as the obesogenic environment fosters unhealthy eating, the environment can foster healthy eating. For example, dieting cues in a restaurant menu stimulate the choice of low-calorie dishes (Papies & Veling, 2013). Such environmental cues are thought to activate weight-control or health goals (Papies, 2016).

Environmental cues that have repeatedly been found to foster healthy eating are the thin, human-like sculptures by the artist Alberto Giacometti. Exposure to these sculptures made healthy snack choices at a vending machine more likely (Stöckli, Stämpfli, Messner, & Brunner, 2016) and reduced the intake of unhealthy, high-calorie chocolate and chips (Brunner & Siegrist, 2012; Stämpfli & Brunner, 2016). However, it is uncertain which goal primarily drives this "Giacometti effect," as both a health and a weight-control goal are conceivable drivers. This ambiguity reflects the state of the literature on environmental cues. Despite manifold empirical evidence on the effects of environmental cues, the understanding of the specific semantic content activated by a cue is often not revealed (Bargh, 2006; e.g., Papies & Veling, 2013). A detailed understanding of the semantic content activated by a cue and thus driving a cue's effects would be a first indication regarding a cue's possible purpose. Therefore, the goal of the present research was to identify the semantic content that is activated by the Giacometti cue.

* Corresponding author. University of Bern, Institute of Marketing and Management, Department of Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland.

E-mail addresses: aline.staempfli@imu.unibe.ch (A.E. Stämpfli), sabrina.stoeckli@imu.unibe.ch (S. Stöckli), thomas.brunner@bfh.ch (T.A. Brunner).

http://dx.doi.org/10.1016/j.appet.2016.11.037

1. How environmental cues influence behavior

When cues in the environment influence eating behavior, they act as primes. Normally, individuals are not aware of being primed (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Chartrand, 2005). Primes unconsciously and temporarily activate semantically associated mental content that is then more likely integrated into ongoing mental processes and, more likely, influences behavior (Bargh, 2006; Bargh et al., 2001; Janiszewski & Wyer, 2014; Jones & Estes, 2012).

Goals are a specific type of mental content that can be activated (Aarts, 2007; Janiszewski & Wyer, 2014). Due to their motivating capacity (Custers & Aarts, 2005), goals are important drivers for priming effects (Aarts, 2007). For example, individuals with the goal of visiting a library spoke more quietly after being exposed to a picture of a library, compared to when they only saw the picture but did not have the goal in mind (Aarts & Dijksterhuis, 2003). Thus, regarding the Giacometti cue, it is important to determine not only whether the thin sculptures are primarily associated with weight or health, but also whether individuals have a weight-control or health goal in mind.

As mental content is embedded in an associative network, the activation of mental content spreads to associated contents (Aarts, 2007; Janiszewski & Wyer, 2014; Jones & Estes, 2012). In this way, activated weight-related content could activate health-related content. However, in the specific case of goals, it is difficult to predict how weight-control and health goals would interact with each other. On the one hand, they could facilitate each other when they serve as means to each other's attainment. On the other hand, they could inhibit each other when they are perceived as substitutive for an overarching purpose (Shah, Friedman, & Kruglanski, 2002).

When environmental cues are applied for public policy purposes—to improve public health, for example—priming is termed "nudging." Nudging means guiding people toward the interest of society as well as toward self-interested behavior by arranging the decision-making context (Thaler & Sunstein, 2009). Thus, the important role of personal goals for priming effects fits with the notion of nudging.

2. Environmental dieting cues particularly affect restrained eaters

Given the obesity epidemic (World Health Organization, 2016) and the societal ideal of thinness (van de Veer, van Herpen, & van Trijp, 2015), dieting is a goal for many people. Individuals with a chronic goal of weight control are referred to as "restrained eaters" (Herman & Mack, 1975; Stroebe, Mensink, Aarts, Schut, & Kruglanski, 2008). Although restrained eating has been conceptualized as an eating behavior independent of individuals' weight (Herman & Mack, 1975; Herman & Polivy, 1980; van Strien, Breteler, & Ouwens, 2002), restrained eating has repeatedly been found to correlate positively with body mass index (Snoek, Engels, van Strien, & Otten, 2013; van Koningsbruggen, Stroebe, & Aarts, 2011).

Paradoxically, restrained eating does not predict weight loss, but rather weight gain (Lowe, Doshi, Katterman, & Feig, 2013). This can be attributed to our obesogenic environment (Papies et al., 2014) in combination with the goal of eating enjoyment, by which restrained eaters are characterized as well (Stroebe et al., 2008; Stroebe, van Koningsbruggen, Papies, & Aarts, 2012). The fragile balance between restrained eaters' conflicting goals of weight control and eating enjoyment (Stroebe et al., 2008, 2012) makes them particularly sensitive to food-related cues (Fedoroff, Polivy, & Herman, 1997, 2003; Hofmann, van Koningsbruggen, Stroebe, Ramanathan, & Aarts, 2010; Papies, Stroebe, & Aarts, 2008; Soetens, Roets, & Raes, 2014), but, promisingly, also to dieting-related cues in the environment (Anschutz, van Strien, & Engels, 2008; Harris, Bargh, & Brownell, 2009; Papies & Hamstra, 2010; Papies & Veling, 2013; Versluis & Papies, 2016). Thus, the influence of the Giacometti sculptures on restrained eaters can provide important insight into whether the cue's effect is driven by a weight-related goal.

3. The present research: thin, human-like sculptures as an environmental health or weight-related cue

To examine whether the Giacometti effect is driven primarily by weight- or health-related mental content, Study 1 analyzes the sculptures' effects on the consumption volume of unhealthy and healthy foods by applying a between-subjects design. If the sculptures are primarily weight-related, it is hypothesized that being exposed to them will lead to participants' reduced food intake independent of food healthiness. This is because the goal of weight control, and thus calorie reduction, should drive the effect. In this case, no interaction is expected between the cue and food healthiness, but a main effect of the cue on food intake is expected. If the cue is primarily health-related, a health goal should drive the effect. It is hypothesized that in this case, exposure to the sculptures will inhibit the intake of unhealthy foods, but will facilitate the intake of healthy foods, as these are thought to improve one's health. This is because individuals in our sample should be aware of the prevailing insufficient intake of fruits and vegetables, due, for example, to the nationally-known health campaign "5 a day" (Cancer League Switzerland, 2016). They may also know the negative health consequences related to the insufficient intake of fruits and vegetables, such as heart diseases (World Health Organization, 2002, 2004). Thus, an interaction between the cue and food healthiness is expected if the cue is primarily health-related.

Study 2 directly examines the activation of weight- or health-related mental content by means of a word completion task. While the cue's effect on the completion of weight-related words should be facilitated by a weight-control goal, the cue's effect on the completion of health-related words should be facilitated by a health goal. In addition, the correlations of weight- and health-related word completions in the cue and the no-cue conditions are compared to discern the interplay of the potentially activated weight-control or health goals.

4. Study 1: The influence of thin sculptures on unhealthy and healthy food intake

4.1. Method

4.1.1. Participants

Members of a sensory consumer panel and employees and students of a university were invited personally or via e-mail for a food tasting on campus. The tasting objects were not disclosed to ensure that weight-control or health goals did not influence the registrations. Potential participants could choose an appointment on one of seven days between 8:00 a.m. and 18:00 p.m. No appointments were made between 12:00 and 14:00 p.m. in order to circumvent lunchtime influences. Individuals who had participated in a previous study using the Giacometti cue were excluded.

One hundred and thirty-three individuals participated in the study. As they were accustomed, the members of the consumer panel received a compensation of 25 Swiss Francs and the employees and students received a compensation of 10 Swiss Francs. The data of 133 participants were collected. The data of 114 participants were used for the analyses ($M_{age} = 31.72$ years, $SD_{age} = 14.11$; 61.95% female). Eighteen participants were excluded from the analyses because they stated that they had heard of the study before and therefore had an idea about the study's purpose. One participant was excluded because of a missing value for this question.

4.1.2. Design

A 2 (no cue vs. cue) × 2 (unhealthy vs. healthy food) between-subjects design was applied to examine the cue's influence on consumption volume.

4.1.3. Materials and measures

4.1.3.1. Cue. In the cue conditions, the Giacometti cue was applied as a screensaver. The screensaver showed an extract of a photograph depicting three thin figures from Giacometti's sculpture *Piazza*[1], moving in front of a black background (Brunner & Siegrist, 2012; Stämpfli & Brunner, 2016).

4.1.3.2. Food. Each participant was given either 20 chocolates in the unhealthy conditions ($M_{weight} = 45.21$ g, $SD_{weight} = 1.32$) or 20 blueberries in the healthy conditions ($M_{weight} = 39.02$ g, $SD_{weight} = 4.68$). The chocolates consisted of milk chocolate with a crunchy core. Care was taken to ensure that the blueberries were similar in size to the chocolates.

4.1.3.3. Measures. The dependent variable of this study, *consumption volume*, was captured by weighing the blueberries or chocolates in a small plastic bowl before and after the tasting and calculating the weight difference. To measure whether participants had a weight-control goal, *Restrained eating* ($\alpha = .71$) was captured with the German version (Dinkel, Berth, Exner, Rief, & Balck, 2005) of the Concern for Dieting subscale of the Revised Restraint Scale (Herman & Polivy, 1980). Comprising six items, this subscale has proven to capture restrained eating better than the entire restraint scale (van Strien et al., 2002). Example items are “How often are you dieting?”; “Do you give too much time and thought to food?”; and “Do you have feelings of guilt after overeating?” These were captured on 7-point Likert scales (1 = I do not agree at all; 7 = I entirely agree). For the purpose of ensuring that we created healthy and unhealthy conditions, the question “In your opinion, how healthy was the product which you have tasted?” was asked at the end of the study, using a 7-point Likert scale (1 = very unhealthy; 7 = very healthy).

To assess participants' *suspicion about the study purpose*, they were asked: “Have you heard about this study and therefore have an idea what the purpose of the study is?” To rate the foods and to answer further questions, participants completed a computer-based questionnaire generated with E-Prime, version 2.0.10.353 (E-Prime 2 Professional).

4.1.4. Procedure

In the cue conditions, participants entered the experimental room while the screensaver with thin, human-like sculptures by the artist Alberto Giacometti, running on the experimenter's laptop computer, was projected on a screen. Participants in the no-cue conditions entered the experimental room when the experimenter's laptop computer was closed. This way, the projection screen was lit in blue.

The experimental room was a computer room with tiers and a high desk in front. The computers used for the data collection were separated by partitions to build cubicles. First, participants were asked to come to the front tier to receive oral instructions from the experimenter. No partitions were installed in this first tier to ensure that all participants could see the screen. The direct exposure to the screen during the instructions took about 30 s. Afterward, participants chose a seat and the experimenter or a study assistant served the food samples for the tasting. Either blueberries or chocolates were served for each group. Then, participants had 5 min to taste and rate the blueberries or chocolates. They were instructed to eat as much as they wanted. After the food samples were distributed, the experimenter switched off the projector. After the tasting, participants completed the questionnaire.

[1] This sculpture can be found using Google's image search for “Giacometti Piazza”.

4.2. Results

4.2.1. Manipulation check

The creation of healthy and unhealthy conditions with blueberries or chocolates was successful. Participants rated the food samples to be healthier when they tasted blueberries ($M = 5.77$, $SD = 1.33$) than when they tasted chocolates ($M = 2.74$, $SD = 1.25$), $t(111) = 12.48$, $p < .001$, $d = 2.35$.

4.2.2. Unspecific “Giacometti effect”

With a two-factor ANOVA, the cue's effect on participants' consumption volume of unhealthy and healthy foods was examined. The analysis revealed that the projected Giacometti screensaver influenced how much food participants ate, $F(1, 110) = 3.96$, $p < .05$, $\eta^2 = .03$. The participants who had been exposed to the projected Giacometti screensaver ate less ($M = 17.83$ g, $SD = 9.68$) than the participants who had been exposed to the neutral blue projection screen ($M = 21.82$ g, $SD = 10.81$), $t(112) = 2.08$, $p = .04$, $d = 0.39$; see Fig. 1.

Importantly, food healthiness did not influence the Giacometti effect, $F(1, 110) = 0.20$, $p = .66$, $\eta^2 = .00$. Regarding the type of food, the ANOVA revealed a main effect of food healthiness, $F(1, 110) = 11.58$, $p < .001$, $\eta^2 = .09$. Participants ate more of the healthy blueberries ($M = 22.82$ g, $SD = 10.43$) than they did of the unhealthy chocolates ($M = 16.45$ g, $SD = 9.29$), $t(112) = 3.45$, $p < .001$, $d = 0.65$.

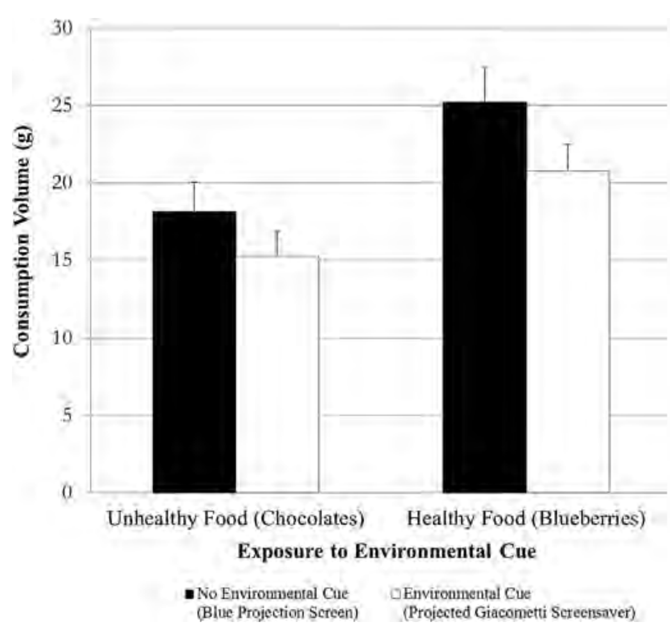

Fig. 1. Mean consumption volume of chocolates and blueberries (in grams) for the four conditions (no cue/cue × unhealthy/healthy food). Participants exposed to a projected screensaver with thin Giacometti sculptures consumed less food than participants exposed to a neutral projection screen. Food healthiness did not alter this effect (error bars represent standard errors).

4.2.3. The influence of the Giacometti sculptures depends on restrained eating

Because food healthiness did not influence the Giacometti effect, food healthiness was omitted in the following analyses. The role that restrained eating plays in the Giacometti effect was analyzed using an ANCOVA that included the cue, restrained eating, and their interaction as independent variables. This analysis revealed that the cue's effect depended on restrained eating, $F(1, 109) = 7.25$, $p = .01$, $\eta^2 = .06$; main effect of the cue, $F(1, 109) = 3.42$, $p = .07$, $\eta^2 = .03$, main effect of restrained eating, $F(1, 109) = 0.16$, $p = .69$, $\eta^2 = .00$. The Johnson-Neyman technique (Hayes, 2013) specified that the projected Giacometti screensaver influenced participants with a restrained eating score upwards of 3.15 on the 7-point scaled moderator variable restrained eating (with a significance level of $\alpha = .05$); see Fig. 2.

4.2.4. Analyses without exclusions

Because of the large number of excluded participants (19), all analyses were repeated without the exclusion of any participants at all. These analyses revealed a marginally significant Giacometti effect, $F(1, 129) = 3.76$, $p = .05$, $\eta^2 = .03$. The participants who had been exposed to the projected Giacometti screensaver ate by tendency less ($M = 17.86$ g, $SD = 9.41$) than the participants who had been exposed to the neutral blue projection screen ($M = 21.10$ g, $SD = 10.58$), $t(131) = 1.87$, $p = .06$, $d = 0.32$. Food healthiness and the cue did not interact, $F(1, 129) = 0.04$, $p = .85$, $\eta^2 = .00$. The main effect of food healthiness remained, $F(1, 129) = 16.12$, $p < .001$, $\eta^2 = .11$. Participants ate more of the healthy blueberries ($M = 22.65$ g, $SD = 10.00$) than they did of the unhealthy chocolates ($M = 16.04$ g, $SD = 9.03$), $t(131) = 4.00$, $p < .001$, $d = 0.69$. Importantly, restrained eating still moderated the Giacometti effect, $F(1, 128) = 7.25$, $p = .01$, $\eta^2 = .05$; main effect of the cue, $F(1, 128) = 3.61$, $p = .06$, $\eta^2 = .03$, main effect of restrained eating, $F(1, 128) = 0.19$, $p = .66$, $\eta^2 = .00$. The projected Giacometti screensaver influenced participants with a restrained eating score upwards of 3.16 (see Hayes, 2013).

4.3. Discussion

The fact that the Giacometti cue's effect was independent of food healthiness in the analyses with and without participant exclusions reveals that the cue is weight-related rather than health-related. While the cue's effect was only marginally significant in the analyses without exclusions, the Giacometti effect was found for restrained eaters in the analyses with and without participant exclusions. This indicates that the Giacometti effect is driven by a weight-control goal.

5. Study 2: The influence of thin sculptures on the completion of weight- and health-related fragmented words

To further analyze the mental content assumed to be activated by the Giacometti cue, Study 2 examined the content-related associations with the cue by means of a word completion task. In addition, the influence of weight- and health-related goals on the cue's effect on word completions was examined by analyzing the influence of restrained eating and general health interest.

5.1. Method

5.1.1. Participants

Participants from a campus other than the one where the first study was conducted were recruited in the university building. They were asked to take part in a study in exchange for a

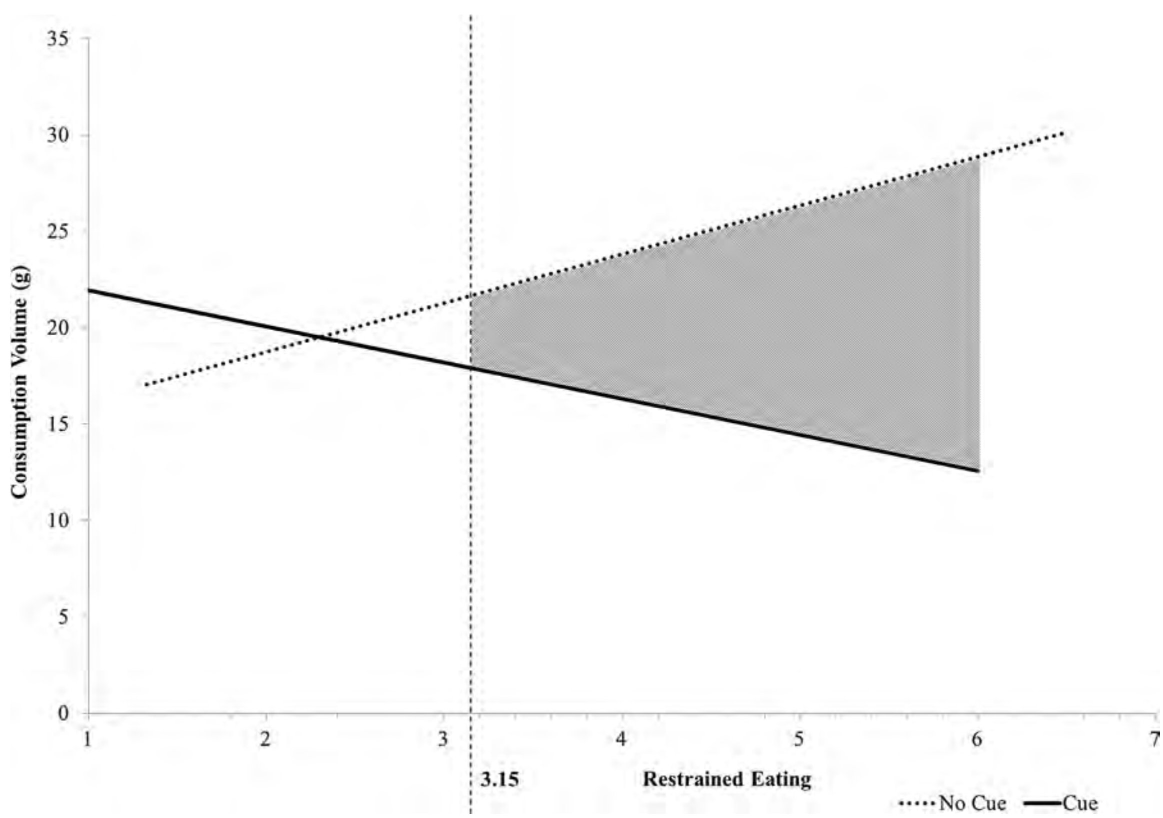

Fig. 2. Chocolate or blueberry consumption as a function of restrained eating and exposure to an environmental cue. Exposure to a projected screensaver with Giacometti sculptures reduced the food intake of participants with a restrained eating score upwards of 3.15.

compensation of 10 Swiss Francs. In accordance with the place of recruitment, the sample consisted almost entirely of students.

Seventy-one individuals took part in the study. The data of 61 participants were used for the analyses (M_{age} = 23.53 years, SD_{age} = 5.07; 63.93% female). One participant was excluded from the analyses because he was assumed to have seen the chocolates directly before the word completion task, and three participants were excluded because their German was insufficient. Another five were excluded because of a breakdown of their computer-based questionnaire. During the restart, these participants could have seen the video file named *Giacometti*. One more participant was excluded because he aborted his participation and therefore did not answer the question regarding whether he had heard of the study before and thus had a suspicion about the study's purpose.

5.1.2. Design

In this study, a one-factorial (no cue vs. cue) between-subjects design was applied to examine the cue's influence on how participants completed fragmented words in a word completion task.

5.1.3. Materials and measures

Fragmented words were created and pretested for their relatedness with weight or health (overview Appendix Table 1). Examples of the weight-related words are *slim*, *dieting*, and *fat*. Examples of the health-related words are *apple*, *balanced*, and *fit*. The dependent variable *mentioning weight* was the number of weight-related words mentioned in the word completion task. The dependent variable *mentioning health* was the number of health-related words mentioned in the word completion task. Very few of the words created by study participants had not been pretested. They were coded as weight- or health-related or neutral. They were also considered if they did not exactly match the gaps given in the fragmented words, since the associations with the sculptures were the focus of interest, not the correct completion of the fragmented words. The coding was done by two independent coders. In cases in which the coding results differed, the two coders reached agreement through discussion.

To examine the role of a weight-control goal in the Giacometti effect, *Restrained eating* (α = .63) was measured. As in Study 1, restrained eating was captured with the Concern for Dieting subscale (Dinkel et al., 2005). To operationalize a health-related goal, *General health interest* (α = .85) was captured. One item was not applied in the data collection because, by relating to cholesterol, it was considered too specific. Example items include the following: "The healthiness of food has little impact on my food choices"; "I always follow a healthy and balanced diet"; and "It is important for me that my daily diet contains a lot of vitamins and minerals" (Roininen, Lähteenmäki, & Tuorila, 1999). Both scales were captured on 7-point Likert scales (1 = I do not agree at all; 7 = I entirely agree). The question used to assess participants' *suspicion about the study purpose* was as follows: "Have you heard about this study and therefore have an idea of what the purpose of the study is?" The computer-based questionnaire was generated with E-Prime, version 2.0.10.353 (E-Prime 2 Professional).

5.1.4. Procedure

The Giacometti cue (see Study 1) was presented as a screensaver directly on participants' computers before they started the computer-based questionnaire. In the no-cue condition, the computers showed a static, white screen.

Participants chose a seat in a cubicle, and the experimenter explained the word completion task. While they were being seated

Table 1

A Pretest Reveals the Words Used to Measure Weight- and Health-Related Words Mentioned, the Dependent Variables of Study 2. N = 117

		"Please indicate how strongly you associate the word *x* with the categories weight and health." (0 = not at all; 5 = strongly)			"Now you have to tie yourself down: With which of those categories (weight, health, neither) do you associate the following words the most?"			
Words envisaged		**M_{weight} (SD)**	**M_{health} (SD)**	***p***	**Weight**	**Health**	**Neither**	***p***
Weight-related words[a]	eating[b]	4.32 (1.11)	4.23 (1.12)	.31	53	63	1	.40
	slim	4.50 (1.02)	3.54 (1.19)	< .001	101	13	3	< .001
	belly	3.83 (1.19)	2.91 (1.52)	< .001	87	23	7	< .001
	fasting	3.83 (1.27)	3.43 (1.43)	.01	62	47	8	.18
	light	3.64 (1.40)	2.55 (1.50)	< .001	92	13	12	< .001
	skinny	4.13 (1.29)	3.38 (1.51)	< .001	95	19	3	< .001
	kilo	4.71 (0.83)	2.67 (1.52)	< .001	112	1	4	< .001
	sugar	3.97 (1.30)	3.91 (1.28)	.57	60	49	8	.34
	dieting (losing weight)	4.70 (0.80)	3.42 (1.32)	< .001	108	7	2	< .001
	fat	4.74 (0.67)	3.91 (1.33)	< .001	108	8	1	< .001
Health-related words	movement	4.19 (1.13)	4.73 (0.54)	< .001	12	104	1	< .001
	strong	2.75 (1.44)	3.19 (1.35)	< .01	33	64	20	< .01
	orange	1.74 (1.53)	3.27 (1.55)	< .001	10	88	19	< .001
	lively	2.08 (1.66)	3.79 (1.38)	< .001	2	105	10	< .001
	apple	2.38 (1.55)	3.85 (1.32)	< .001	8	98	11	< .001
	active	3.74 (1.34)	4.36 (0.85)	< .001	10	103	4	< .001
	fruits	3.15 (1.34)	4.51 (0.82)	< .001	8	105	4	< .001
	balanced	3.44 (1.36)	4.10 (1.17)	< .001	12	103	2	< .001
	fit	3.79 (1.20)	4.50 (0.81)	< .001	14	102	1	< .001
	well	2.78 (1.70)	3.83 (1.35)	< .001	8	102	7	< .001

Notes:

[a] The words are translated from German except *light*, which is also used in German.

[b] Because *eating* was, in contrast to our expectation, rather assigned to health, it was dropped for the measurement of weight-related words in Study 2.

and receiving oral instructions from the experimenter, participants were exposed to the screensavers for about 30 s. Then, they received the instruction to start the computer-based questionnaire by pressing a certain key on their keyboards. Participants first dealt with the word completion task. The fragmented words were displayed for 30 s in the same randomly ordered sequential selection for each participant. During this time, participants had time to enter the word that first came to mind. After the word completion task, participants completed the questionnaire by answering questions, including the items on restrained eating and general health interest.

5.2. Results

5.2.1. The Giacometti sculptures increased the weight-related word completion of restrained eaters

One-factor ANOVAs revealed no effect of the Giacometti screensaver on the amount of weight-related, $F(1, 59) < 0.01$, $p = .99$, $\eta^2 = .00$, or health-related words mentioned, $F(1, 59) = 0.71$, $p = .40$, $\eta^2 = .01$. However, including restrained eating in an ANCOVA with mentioning weight as the dependent variable revealed an interaction of the screensaver with restrained eating, $F(1, 57) = 5.64$, $p = .02$, $\eta^2 = .09$; main effect of the screensaver, $F(1, 57) = 4.99$, $p = .03$, $\eta^2 = .08$, main effect of restrained eating, $F(1, 57) = 0.58$, $p = .45$, $\eta^2 = .01$. The Johnson-Neyman technique (Hayes, 2013) revealed that the Giacometti screensaver increased the creation of weight-related words in restrained eaters (in participants with a restrained eating score upwards of 4.03; with a significance level of $\alpha = .05$); see Fig. 3. In contrast, restrained eaters did not mention more health-related words after being exposed to the screensaver, compared to individuals low in restrained eating, $F(1, 57) = 0.07$, $p = .79$, $\eta^2 = .00$; main effect of the screensaver, $F(1, 57) < 0.01$, $p = .96$, $\eta^2 = .00$, main effect of restrained eating, $F(1, 57) = 0.57$, $p = .45$, $\eta^2 = .01$.

An ANCOVA including the cue, general health interest, and their interaction as independent variables and the number of health-related words mentioned as dependent variable revealed no interaction of the screensaver with general health interest, $F(1, 57) = 0.51$, $p = .48$, $\eta^2 = .01$; main effect of the screensaver, $F(1, 57) = 0.26$, $p = .61$, $\eta^2 = .00$, main effect of general health interest, $F(1, 57) = 1.75$, $p = .19$, $\eta^2 = .03$.

5.2.2. Correlations of mentioned weight- and health-related words

In order to explore the effect of the activated weight-related content on health-related content, we examined the possible changes in the correlation of mental weight- and health-related content as a consequence of the cue exposure. Bivariate correlation analyses were conducted. They revealed that the participants' mentioning of weight- and health-related words did not correlate, both in participants exposed to the neutral screensaver, $r_{Spearman}(29) = .33$, $p = .08$, and in participants exposed to the Giacometti screensaver, $r_{Spearman}(32) = .11$, $p = .54$. In addition, the association of weight- and health-related words mentioned, measured with the difference of health-mentioning and weight-mentioning, did not differ between the neutral condition ($M = 1.90$, $SD = 1.70$) and the cue condition ($M = 2.28$, $SD = 2.05$), $t(59) = 0.79$, $p = .43$, $d = 0.20$. An ANCOVA analyzing the effects of the cue, restrained eating, and the interaction of cue and restrained eating on the difference of weight- and health-related words mentioned revealed that the association of weight- and health-related words mentioned between the cue and the no-cue condition did not depend on restrained eating, $F(1, 57) = 1.76$, $p = .19$, $\eta^2 = .03$; main effect of the screensaver, $F(1, 57) = 2.08$, $p = .15$, $\eta^2 = .02$, main effect of restrained eating, $F(1, 57) = 1.49$, $p = .23$, $\eta^2 = .03$. These results indicate that weight- and health-related content did not correlate in our sample and that this did not change with either cue exposure or cue exposure and restrained eating.

5.2.3. Analyses without exclusions

No significant results were found when all of the analyses were conducted without any exclusion of participants. One-factor ANOVAs revealed no effect of the screensaver on the number of weight-related, $F(1, 69) = 0.26$, $p = .62$, $\eta 2 = .00$, or health-related words mentioned, $F(1, 69) = 0.52$, $p = .48$, $\eta 2 = .01$. Analyzing the data with an ANCOVA that included restrained eating did not yield any relationships. There was no interaction of the screensaver with restrained eating, $F(1, 66) = 0.08$, $p = .78$, $\eta 2 = .00$, a main effect of the screensaver, $F(1, 66) = 0.01$, $p = .93$, $\eta 2 = .00$, or a main effect of restrained eating, $F(1, 66) = 0.86$, $p = .36$, $\eta 2 = .01$. An ANCOVA including the cue, general health interest, and their interaction as independent variables and the number of health-related words mentioned as dependent variable revealed no interaction of the screensaver with general health interest, $F(1, 66) = .48$, $p = .49$, $\eta^2 = .01$; main effect of the screensaver, $F(1, 66) = .26$, $p = .61$, $\eta^2 = .00$, main effect of general health interest, $F(1, 66) = 1.99$, $p = .16$, $\eta^2 = .03$.

In addition, no indications of a difference in the association of weight- and health-related words mentioned as a consequence of the cue exposure or the cue exposure and restrained eating were found for the sample without participant exclusions.

5.3. Discussion

The results of the analyses with participant exclusions in Study 2 are in line with the results of Study 1. With restrained eaters' increased mentioning of weight-related words after they were exposed to the thin Giacometti sculptures, Study 2 indicates that the Giacometti cue is weight-related and that the Giacometti effect is driven by a weight-control goal. However, because the activation of mentioning weight-related words by the Giacometti cue for restrained eaters could not be found in the sample without participant exclusions, no firm conclusion should be drawn from these results. In contrast, calculated with and without participant exclusions, the cue had no influence on the mentioning of health-related words, even in individuals with a relatively high general health interest. The results of the correlation analyses indicate that weight- and health-related content did not correlate in our sample.

6. General discussion

The present paper aimed to shed light on the content-related processes underlying priming effects with a distinct environmental cue—thin, human-like sculptures by the artist Alberto Giacometti. This is because content-related cognitive processes mostly have been neglected in existing priming studies using environmental cues (Bargh, 2006). In our studies, the Giacometti sculptures were found to be a weight-related environmental cue that can help restrained eaters in facilitating their dieting by reducing their consumption volume.

6.1. Priming weight is not priming health

A detailed understanding of the specific mental content activated by an environmental cue provides the first indications with respect to a cue's possible use. Such understanding also indicates which individuals could be addressed with a distinct cue—i.e., individuals who have a goal that the cue can activate.

However, when an environmental cue, such as the Giacometti sculptures, has an effect on weight-related content, it is conceivable that health-related content is also activated (Janiszewski & Wyer,

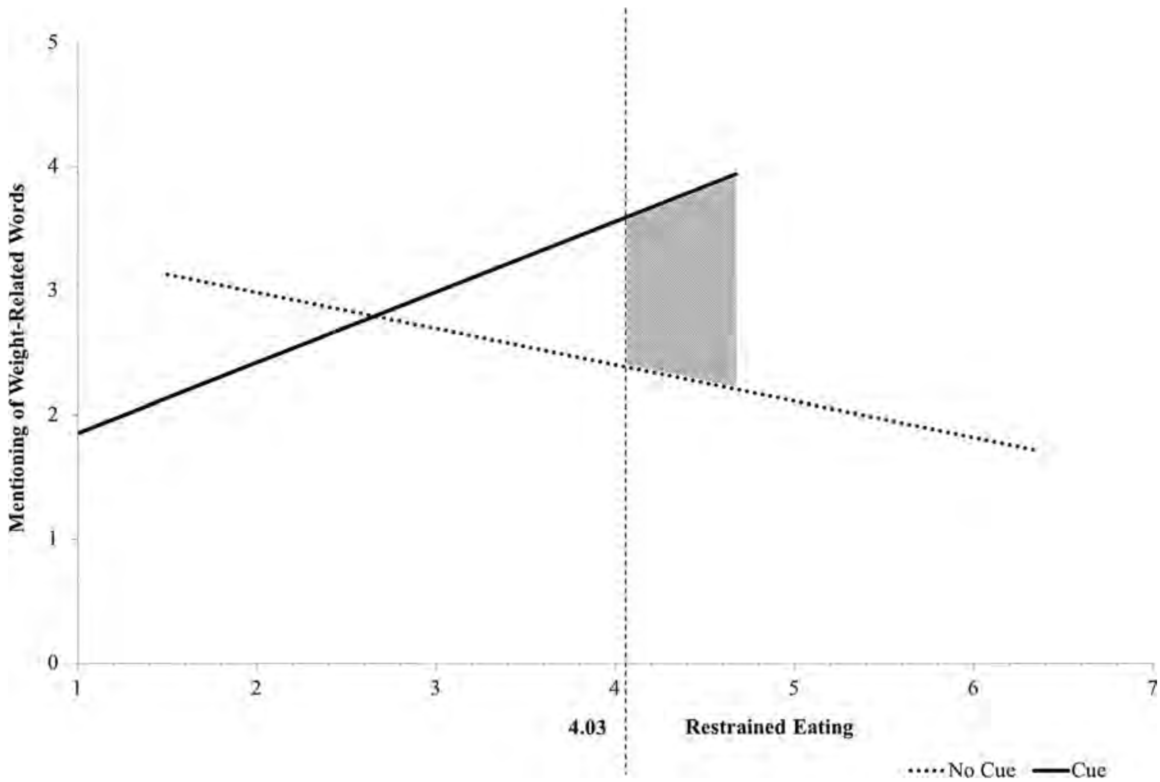

Fig. 3. Mentioning of weight-related words as a function of restrained eating and exposure to an environmental cue. A screensaver with thin, human-like sculptures increased the mentioning of weight-related words in a word completion task in restrained eaters (upwards of a restrained eating score of 4.03).

2014). This is because weight and health are commonly thought to be semantically related. This could not be shown in our sample, however, and health-related content was not activated by the Giacometti sculptures. One potential reason why a relationship could not be determined between weight- and health-related content may be the young age of the participants in Study 2. Health problems caused by weight may not yet be manifest in youth. Nonetheless, being overweight has negative health consequences (World Health Organization, 2002).

6.2. Implications

In regard to the prevailing epidemic of overweight and obesity (World Health Organization, 2016), environmental weight-control cues could play a pivotal role. Primary weight-related cues can be seen as counterparts to the abundance of food and food-related cues in our obesogenic environment (Papies et al., 2014).

An example of applied environmental cues for the promotion of health is the deterrent pictures on cigarette packages, which show the physical consequences of smoking (European Union, 2014). However, studies have found these deterrent pictures to be ineffective (Glock & Kneer, 2009). This indicates that obese figures and the "fear of fat" (Anschutz, Engels, Becker, & van Strien, 2009) would be less effective drivers against eating high-calorie food than cues such as the thin Giacometti sculptures. These sculptures can be seen as motivators, as they emphasize the positive consequences of eating less high-calorie food in order to get closer to the ideal of a thin figure (van de Veer et al., 2015). Results from neural research substantiate that motivation works better than deterrence in the domain of eating. Besides homeostatic regulation, eating is assumed to be controlled by a neural network, which is supposed to consist of a reward pathway and a control pathway (Chen, Papies, & Barsalou, 2016). Interestingly, thinking about the long-term benefits of not eating has been found to increase activity in the inhibitory neural pathway and to reduce activity in the reward pathway more than thinking about the long-term costs of eating (Yokum & Stice, 2013). Evidence from research on reactions to thin and round figures further indicates that thin figures may have more influence on reducing calorie intake than obese figures. For example, dieters ate less when their server was thin than when she was overweight (McFerran, Dahl, Fitzsimons, & Morales, 2010).

With regard to the specific body forms of the Giacometti sculptures, it must be acknowledged that human bodies with figures similar to these sculptures would be seriously underweight. Thus, they would be perceived as less attractive and thus less motivating than figures corresponding to the lower ranges of normal body mass indices (Tovée, Edmonds, & Vuong, 2012; Tovée, Furnham, & Swami, 2007; Weeden & Sabini, 2005). When using human models as environmental cues, using healthier-looking human figures could thus work better than skinny human figures. Supporting evidence for this demonstrates that female television viewers ate less unhealthy food when they watched average-sized or slightly oversized models than they did when exposed to thin models (Anschutz et al., 2009). However, when compared to using human models as environmental cues, the Giacometti sculptures seem to have the advantage of being more generally applicable. Social comparison processes due to characteristics such as clothing or age should be prevented when using artistically simplified human sculptures (Corcoran, Crusius, & Mussweiler, 2011).

6.3. Limitations

Besides the conceivable application of environmental cues for public policy purposes, the question arises whether dieting cues could be used intentionally by individuals for losing weight. If applied intentionally, a cue could be processed more controlled than when used as a subtle prime. There is evidence that intention could even support a cue's influence, as primes can affect behavior through both automatic and controlled processes (Payne, Brown-Iannuzzi, & Loersch, 2016). Because losing weight is a long-term process, another question that arises is what would happen if cues are applied repeatedly. To our knowledge, there is very little evidence revealing the effects of repeatedly exposing individuals to an environmental weight- or health-related cue (Klesse, Goukens, Geyskens, & de Ruyter, 2012). A constant reactivation of goals and a habituation to the cue with a decreasing effect of the cue (Rankin et al., 2009) are both conceivable.

With a long-term application of environmental weight-control cues, the unintended effects of exposing people to the thin ideal become more important and have to be taken into consideration. Examples of unintended effects are negative affect, increased body dissatisfaction, and disordered eating patterns for vulnerable groups of people, such as vulnerable adolescents (Stice, Spangler, & Agras, 2001) or unsuccessful restrained eaters (Schaumberg, Anderson, Anderson, Reilly, & Gorrell, 2016).

7. Conclusion

In sum, the present research indicates that exposure to thin, human-like sculptures by the artist Alberto Giacometti reduces food intake in restrained eaters and thus that the Giacometti effect is driven by a weight-control goal. Given that restrained eaters are often unsuccessful in dieting, partly because of the obesogenic environment with its abundance of food and food-related cues (Lowe et al., 2013; Papies et al., 2014; Stroebe et al., 2008), weight-control cues in the environment can be seen as helpful counterparts. By helping restrained eaters to pursue their weight-control goal, environmental weight-control cues could act as daily nudges in a healthier direction (Hill, Wyatt, Reed, & Peters, 2003).

Role of funding sources

This research was financially supported by the Swiss National Science Foundation as part of the National Research Program NRP 69— Healthy Nutrition and Sustainable Food Production (grant 12 No. 145189). The study design was partly (Study 1) included in the funding proposal. The collection, analysis, and interpretation of the data, in addition to all work contributing to the present manuscript and journal selected for submission, were done independently of the funding sponsor.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Appendix

Pretest Fragmented Words

To assess whether the Giacometti screensaver activates health- or weight-related mental content, 40 fragmented words were created: 20 that could be completed either by a weight-related or a neutral word and 20 that could be completed either by a health-related or a neutral word. Efforts were made to ensure that the words were not too difficult to complete. A qualitative pretest with 14 individuals (57% female) ensured this and also assessed how often the fragmented words were completed by a weight- or health-related (i.e., semantic category-related) word instead of a neutral word. Ten fragmented words were chosen per semantic category. They were completed with a category-related word between 7% and 57%. With this choice, a sufficient variance was expected in semantic category-related and neutral word completions per fragmented word. The expected weight-related words included, e.g., *slim*, *dieting*, and *fat*. The expected health-related words included, e.g., *apple*, *balanced*, and *fit*.

A second independent pretest was conducted to examine whether the words conceived to represent the weight- and health-related semantic categories can be assigned distinctly to weight or health. One hundred and forty-eight individuals participated in an online questionnaire. The link to this questionnaire was posted in online market places of university websites. The data of everyone who completed the questionnaire (117 participants) were analyzed. In a first step, the participants had to rate how strongly they associated the envisaged weight- and health-related words with both weight and health (0 = not at all; 5 = strongly). They associated all semantic category words with the expected semantic category. All words had a mean rating of higher than 3, which was, with two exceptions, higher than the mean of the competing semantic category. The words *eating* and *sugar*, which were created to represent the weight category, were not associated significantly more strongly with weight than with health (see Table 1). In a second step of this pretest, participants had to decide the category with which they associated each word the most: weight, health, or neither of these categories. Binomial tests revealed that all assignments were made as expected except for the words *eating*, *sugar*, and *fasting*, which were envisaged to represent the weight category. While *sugar* ($p = .34$) and *fasting* ($p = .18$) were rather assigned to weight, *eating* was assigned, in contrast to our expectation, rather to health ($p = .40$; see Table 1). As a consequence, the word *eating* was dropped.

References

Aarts, H. (2007). Health and goal-directed behavior: The nonconscious regulation and motivation of goals and their pursuit. *Health Psychology Review, 1*(1), 53–82. http://dx.doi.org/10.1080/17437190701485852.
Aarts, H., & Dijksterhuis, A. (2003). The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology, 84*(1), 18–28. http://dx.doi.org/10.1037/0022-3514.84.1.18.
Anschutz, D. J., Engels, R. C., Becker, E. S., & van Strien, T. (2009). The effects of TV commercials using less thin models on young women's mood, body image and actual food intake. *Body Image, 6*(4), 270–276. http://dx.doi.org/10.1016/j.bodyim.2009.07.007.
Anschutz, D. J., van Strien, T., & Engels, R. C. (2008). Exposure to slim images in mass media: Television commercials as reminders of restriction in restrained eaters. *Health Psychology, 27*(4), 401–408. http://dx.doi.org/10.1037/0278-6133.27.4.401.
Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36*(2), 147–168. http://dx.doi.org/10.1002/ejsp.336.
Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*(6), 1014–1027. http://dx.doi.org/10.1037//0022-3514.81.6.1014.
Berthoud, H.-R. (2006). Homeostatic and non-homeostatic pathways involved in the control of food intake and energy balance. *Obesity, 14*(5), 197S–200S. http://dx.doi.org/10.1038/oby.2006.308.
Brunner, T. A., & Siegrist, M. (2012). Reduced food intake after exposure to subtle weight-related cues. *Appetite, 58*(3), 1109–1112. http://dx.doi.org/10.1016/j.appet.2012.03.010.
Cancer League Switzerland (2016). *5 a day*. Retrieved from https://www.5amtag.ch/en/.
Chartrand, T. L. (2005). The role of conscious awareness in consumer behavior. *Journal of Consumer Psychology, 15*(3), 203–210. http://dx.doi.org/10.1207/s15327663jcp1503_4.
Chen, J., Papies, E. K., & Barsalou, L. W. (2016). A core eating network and its modulations underlie diverse eating phenomena. *Brain and Cognition, 110*, 20–42. http://dx.doi.org/10.1016/j.bandc.2016.04.004.
Corcoran, K., Crusius, J., & Mussweiler, T. (2011). Social comparison: Motives, standards, and mechanisms. In D. Chadee (Ed.), *Theories in social psychology* (pp. 119–139). Oxford, UK: Wiley-Blackwell.
Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89*(2), 129–142. http://dx.doi.org/10.1037/0022-3514.89.2.129.
Dinkel, A., Berth, H., Exner, C., Rief, W., & Balck, F. (2005). Deutsche Adaptation der Restraint Scale zur Erfassung gezügelten Essverhaltens. *Diagnostica, 51*(2),

67–74. http://dx.doi.org/10.1026/0012-1924.51.2.67.
E-Prime 2 Professional (version 2.0.10.353) [Computer software], Sharpsburg, PA. *Psychology software tools*.
European Union. (2014). *Directive 2014/40/EU of the European parliament and of the council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the member states concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC*. Retrieved from http://ec.europa.eu/health/tobacco/docs/dir_201440_en.pdf.
Fedoroff, I., Polivy, J., & Herman, C. P. (1997). The effect of pre-exposure to food cues on the eating behavior of restrained and unrestrained eaters. *Appetite, 28*(1), 33–47. http://dx.doi.org/10.1006/appe.1996.0057.
Fedoroff, I., Polivy, J., & Herman, C. P. (2003). The specificity of restrained versus unrestrained eaters' responses to food cues: General desire to eat, or craving for the cued food? *Appetite, 41*(1), 7–13. http://dx.doi.org/10.1016/S0195-6663(03)00026-6.
Glock, S., & Kneer, J. (2009). Are deterrent pictures effective? The impact of warning labels on cognitive dissonance in smokers. *Applied Psychology: Health and Well-Being, 1*(3), 356–373. http://dx.doi.org/10.1111/j.1758-0854.2009.01019.x.
Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health Psychology, 28*(4), 404–413. http://dx.doi.org/10.1037/a0014399.
Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality, 43*(4), 647–660. http://dx.doi.org/10.1111/j.1467-6494.1975.tb00727.x.
Herman, C. P., & Polivy, J. (1980). Restrained eating. In A. J. Stunkard (Ed.), *Obesity* (pp. 208–225). Philadelphia/London/Toronto: W. B. Saunders Company.
Hill, J. O., Wyatt, H. R., Reed, G. W., & Peters, J. C. (2003). Obesity and the environment: Where do we go from here? *Science, 299*(5608), 853–855. http://dx.doi.org/10.1126/science.1079857.
Hofmann, W., van Koningsbruggen, G. M., Stroebe, W., Ramanathan, S., & Aarts, H. (2010). As pleasure unfolds: Hedonic responses to tempting food. *Psychological Science, 21*(12), 1863–1870. http://dx.doi.org/10.1177/0956797610389186.
Janiszewski, C., & Wyer, R. S. (2014). Content and process priming: A review. *Journal of Consumer Psychology, 24*(1), 96–118. http://dx.doi.org/10.1016/j.jcps.2013.05.006.
Jones, L. L., & Estes, Z. (2012). Lexical priming: Associative, semantic and thematic influences on word recognition. In J. S. Adelman (Ed.), *Visual word recognition* (vol. 2). Hove, UK: Psychology Press.
Klesse, A. K., Goukens, C., Geyskens, K., & de Ruyter, K. (2012). Repeated exposure to the thin ideal and implications for the self: Two weight loss program studies. *International Journal of Research in Marketing, 29*(4), 355–362. http://dx.doi.org/10.1016/j.ijresmar.2012.06.003.
van Koningsbruggen, G. M., Stroebe, W., & Aarts, H. (2011). Through the eyes of dieters: Biased size perception of food following tempting food primes. *Journal of Experimental Social Psychology, 47*(2), 293–299. http://dx.doi.org/10.1016/j.jesp.2010.10.012.
Lowe, M. R., Doshi, S. D., Katterman, S. N., & Feig, E. H. (2013). Dieting and restrained eating as prospective predictors of weight gain. *Frontiers in Psychology, 4*(577), 1–7. http://dx.doi.org/10.3389/fpsyg.2013.00577.
McFerran, B., Dahl, D. W., Fitzsimons, G. J., & Morales, A. C. (2010). Might an overweight waitress make you eat more? How the body type of others is sufficient to alter our food consumption. *Journal of Consumer Psychology, 20*(2), 146–151. http://dx.doi.org/10.1016/j.jcps.2010.03.006.
Papies, E. K. (2016). Goal priming as a situated intervention tool. *Current Opinion in Psychology, 12*, 12–16. http://dx.doi.org/10.1016/j.copsyc.2016.04.008.
Papies, E. K., & Hamstra, P. (2010). Goal priming and eating behavior: Enhancing self-regulation by environmental cues. *Health Psychology, 29*(4), 384–388. http://dx.doi.org/10.1037/a0019877.
Papies, E. K., Potjes, I., Keesman, M., Schwinghammer, S., & van Koningsbruggen, G. M. (2014). Using health primes to reduce unhealthy snack purchases among overweight consumers in a grocery store. *International Journal of Obesity, 38*(4), 597–602. http://dx.doi.org/10.1038/ijo.2013.136.
Papies, E. K., Stroebe, W., & Aarts, H. (2008). The allure of forbidden food: On the role of attention in self-regulation. *Journal of Experimental Social Psychology, 44*(5), 1283–1292. http://dx.doi.org/10.1016/j.jesp.2008.04.008.
Papies, E. K., & Veling, H. (2013). Healthy dining. Subtle diet reminders at the point of purchase increase low-calorie food choices among both chronic and current dieters. *Appetite, 61*(1), 1–7. http://dx.doi.org/10.1016/j.appet.2012.10.025.
Payne, B. K., Brown-Iannuzzi, J. L., & Loersch, C. (2016). Replicable effects of primes on human behavior. *Journal of Experimental Psychology. General, 145*(10), 1269–1279. http://dx.doi.org/10.1037/xge0000201.
Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., … Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory, 92*(2), 135–138. http://dx.doi.org/10.1016/j.nlm.2008.09.012.
Renner, B., Sproesser, G., Strohbach, S., & Schupp, H. T. (2012). Why we eat what we eat. The Eating Motivation Survey (TEMS). *Appetite, 59*(1), 117–128. http://dx.doi.org/10.1016/j.appet.2012.04.004.
Roininen, K., Lähteenmäki, L., & Tuorila, H. (1999). Quantification of consumer attitudes to health and hedonic characteristics of foods. *Appetite, 33*(1), 71–88. http://dx.doi.org/10.1006/appe.1999.0232.
Schaumberg, K., Anderson, D. A., Anderson, L. M., Reilly, E. E., & Gorrell, S. (2016). Dietary restraint: what's the harm? A review of the relationship between dietary restraint, weight trajectory and the development of eating pathology. *Clinical Obesity, 6*(2), 89–100. http://dx.doi.org/10.1111/cob.12134.
Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*(6), 1261–1280. http://dx.doi.org/10.1037//0022-3514.83.6.1261.
Snoek, H. M., Engels, R. C., van Strien, T., & Otten, R. (2013). Emotional, external and restrained eating behaviour and BMI trajectories in adolescence. *Appetite, 67*, 81–87. http://dx.doi.org/10.1016/j.appet.2013.03.014.
Soetens, B., Roets, A., & Raes, F. (2014). "Food for memory": Pictorial food-related memory bias and the role of thought suppression in high and low restrained eaters. *The Psychological Record, 64*(1), 105–114. http://dx.doi.org/10.1007/s40732-014-0016-0.
Stämpfli, A. E., & Brunner, T. A. (2016). The art of dieting: Exposure to thin sculptures effortlessly reduces the intake of unhealthy food in motivated eaters. *Food Quality and Preference, 50*, 88–93. http://dx.doi.org/10.1016/j.foodqual.2016.01.012.
Stice, E., Spangler, D., & Agras, W. S. (2001). Exposure to media-portrayed thin-ideal images adversely affects vulnerable girls: A longitudinal experiment. *Journal of Social and Clinical Psychology, 20*(3), 270–288. http://dx.doi.org/10.1521/jscp.20.3.270.22309.
Stöckli, S., Stämpfli, A. E., Messner, C., & Brunner, T. A. (2016). An (un)healthy poster: When environmental cues affect consumers' food choices at vending machines. *Appetite, 96*, 368–374. http://dx.doi.org/10.1016/j.appet.2015.09.034.
van Strien, T., Breteler, M. H., & Ouwens, M. A. (2002). Restraint Scale, its sub-scales concern for dieting and weight fluctuation. *Personality and Individual Differences, 33*(5), 791–802. http://dx.doi.org/10.1016/S0191-8869(01)00192-1.
Stroebe, W., van Koningsbruggen, G. M., Papies, E. K., & Aarts, H. (2012). Why most dieters fail but some succeed: A goal conflict model of eating behavior. *Psychological Review, 120*(1), 110–138. http://dx.doi.org/10.1037/a0030849.
Stroebe, W., Mensink, W., Aarts, H., Schut, H., & Kruglanski, A. W. (2008). Why dieters fail: Testing the goal conflict model of eating. *Journal of Experimental Social Psychology, 44*(1), 26–36. http://dx.doi.org/10.1016/j.jesp.2007.01.005.
Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness (revised and expanded ed.)*. London, England: Penguin Books.
Tovée, M. J., Edmonds, L., & Vuong, Q. C. (2012). Categorical perception of human female physical attractiveness and health. *Evolution and Human Behavior, 33*(2), 85–93. http://dx.doi.org/10.1016/j.evolhumbehav.2011.05.008.
Tovée, M. J., Furnham, A., & Swami, V. (2007). *Healthy body equals beautiful body? Changing perceptions of health and attractiveness with shifting socioeconomic status*. Basingstoke, Hampshire, England: Palgrave Macmillan.
van de Veer, E., van Herpen, E., & van Trijp, H. C. (2015). How do I look? Focusing attention on the outside body reduces responsiveness to internal signals in food intake. *Journal of Experimental Social Psychology, 56*, 207–213. http://dx.doi.org/10.1016/j.jesp.2014.10.003.
Versluis, I., & Papies, E. K. (2016). Eating less from bigger packs: Preventing the pack size effect with diet primes. *Appetite, 100*, 70–79. http://dx.doi.org/10.1016/j.appet.2016.02.011.
Weeden, J., & Sabini, J. (2005). Physical attractiveness and health in Western societies: A review. *Psychological Bulletin, 131*(5), 635–653. http://dx.doi.org/10.1037/0033-2909.131.5.635.
World Health Organization. (2002). *The world health report*. Retrieved from http://www.who.int/whr/2002/en/whr02_en.pdf?ua=1.
World Health Organization. (2004). *Global strategy on diet, physical activity and health*. Retrieved from http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf.
World Health Organization. (2016). *Nutrition: Controlling the global obesity epidemic*. Retrieved from http://www.who.int/nutrition/topics/obesity/en/.
Yokum, S., & Stice, E. (2013). Cognitive regulation of food craving: Effects of three cognitive reappraisal strategies on neural response to palatable foods. *International Journal of Obesity, 37*(12), 1565–1570. http://dx.doi.org/10.1038/ijo.2013.39.

Paper III

Facial Expression Analysis with AFFDEX and FACET: A Validation Study

Stöckli S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017). Facial expression analysis with AFFDEX and FACET: A validation study. Manuscript under revision at Behavior Research Methods.

Facial Expression Analysis with AFFDEX and FACET: A Validation Study

Sabrina Stöckli[a]

Michael Schulte-Mecklenbeck[a, b]

Stefan Borer[a]

Andrea C. Samson[c,d]

a Institute of Marketing and Management, University of Bern, Bern, Switzerland

b Max Planck Institute for Human Development, Berlin, Germany

c Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

d Department of Psychiatry and Behavioral Science, Stanford University School of Medicine, Stanford, CA, USA

Author Note

We thank Adem Halimi and Elena von Wyttenbach for their help with the data collection.

Correspondence about this article should be addressed to: Sabrina Stöckli, University of Bern, Institute of Marketing and Management, Department of Consumer Behavior, Engehaldenstrasse 4, 3012 Bern, Switzerland, Email: Sabrina.Stoeckli@imu.unibe.ch

Abstract

The goal of this study was to validate AFFDEX and FACET, two software-based algorithms designed to analyze emotional facial expressions. In Study 1, pictures of standardized emotional facial expressions from three databases, *Warsaw Set of Emotional Facial Expression Pictures* (WSEFEP), *Amsterdam Dynamic Facial Expression Set* (ADFES) and *Radboud Faces Database* (RaFD), were classified with both modules. Results show a large variance in accuracy across emotions and databases, with a performance advantage for FACET over AFFDEX. In Study 2, 110 respondents' facial responses were measured while the respondents were exposed to emotionally evocative pictures from the *International Affective Picture System* (IAPS), *Geneva Affective Picture Database* (GAPED) and *Radboud Faces Database* (RaFD). Accuracy again differed for distinct emotions, and FACET performed better. Overall, iMotions can achieve acceptable accuracy for standardized pictures of prototypical (vs. natural) facial expressions, but only weak accuracy for more natural facial expressions. We discuss potential sources for limited validity and suggest research directions in the broader context of emotion research.

Keywords: emotion classification; facial expression; FACS; AFFDEX; FACET

Facial Expression Analysis with AFFDEX and FACET: A Validation Study

The de facto standard for measuring emotional facial expressions is the *Facial Action Coding System* (FACS; Ekman & Friesen, 1976). This anatomy-based system allows human coders to evaluate emotions based on 46 observable Action Units (AU), facial movements that account for facial expressions and in turn for the expression of emotions (Ekman & Friesen, 1976). FACS coding requires certified coders who are trained for up to 100 hours (e.g. at workshops by the Paul Ekman Group LLC). On top of this time-intensive training, the coding process itself is also time- and labor-intensive. Video recordings of participants' faces are often recorded with a resolution of 24 frames per second, meaning that for each second of recording the coder has to produce 24 ratings of the 46 AUs. So for one participant with only one minute of video, 1440 individual ratings are necessary. Assuming that a coder could rate one picture per second, this would add up to approximately 24 minutes of work for one minute of video data (cf. Ekman & Oster, 1979).

Automated facial expression analysis is a promising tool that may overcome the limitations of human-based FACS coding. Automated detection of discrete facial emotions has progressed significantly in the last three decades. This progress is largely due to rapid developments in computer science which have made automated facial expression analysis more valid, reliable and accessible (Swinton & El Kaliouby, 2012; Valstar, Jiang, Mehu, Pantic, & Scherer, 2011; Lewinski, den Uyl, & Butler, 2014).

One commercial tool for automated facial expression analysis is part of a software suite by iMotions (www.imotions.com). iMotions's biometric research platform can be used for various types of academic and business-related research and offers automated facial expression analysis in combination with EEG, GSR, EMG, ECG, eye tracking and surveys. The automated facial expression analysis part allows the user to record videos with a laptop camera, mobile phone camera or standalone webcams. To analyze these videos, iMotions detects changes in key face features (i.e., facial landmarks such as brows, eyes and lips) and generates raw, unfiltered data representing the basic emotions of the recorded face. Researchers can choose between two different modules to classify emotions of facial expressions: the FACET module, based on the FACET algorithm (formerly the Computer Expression Recognition Toolbox (CERT) algorithm; Littlewort et al., 2011) and the AFFDEX module, based on the AFFDEX algorithm by Affectiva Inc. (El Kaliouby & Robinson, 2005; McDuff, El Kaliouby, Kassam, & Picard, 2010). These algorithms detect facial landmarks and apply a set of rules based on psychological theories and statistical procedures to classify emotions. FACET and AFFDEX differ in their statistical procedures (e.g., they use different facial landmarks) as well as in the facial databases that are used to train the machine-learning procedures (iMotions, 2016).

In contrast to the increasing interest in automated facial expression analysis, there seems to be little to no interest in validating these measures, and there are hardly any publications on the topic in peer-reviewed journals. (Note that there are several conference presentations on this topic; e.g., Baltrusaitis, Robinson, & Morency, 2016; Taggart, Dressler, Kumar, Khan, & Coppola, n.d.) FaceReader, a software marketed by Noldus (www.noldus.com), is the only tool we are aware of with available published evaluation work (den Uyl & van Kuilenburg, 2005; Lewinski, den Uyl, & Butler, 2014; van Kuilenburg, Wiering, & den Uyl, 2005). As there is no such validation for iMotions's AFFDEX and FACET modules, the present research fills this gap by validating and comparing their performance.

The Theory of Facial Expression Analysis

Facial expressions reveal much about our emotional state (Ekman & Friesen, 1982; Ekman, 1992a; Ekman & Oster, 1979). Early research on facial expressions has focused on the six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) that are universally recognized (e.g., Ekman, 1992a; Ekman et al., 1987). More recent research has emphasized specific cultural and contextual aspects of the recognition of displayed emotions (e.g., Aviezer, Trope, & Todorov, 2012; Barrett, Mesquita, & Gendron, 2011; Elfenbein & Ambady, 2002) and facial expression of compound emotions; i.e., combinations of single components of basic emotions which create new and distinct emotional expressions (e.g., Du, Tao, & Martinez, 2014).

One important milestone in facial expression research was the introduction of FACS (Ekman & Friesen, 1976). Within this system, 46 facial AUs represent distinct movements displayed on the face, and emerge by activating one or a combination of facial muscles. FACS provides a coding schema for AU activity and intensity (Ekman et al., 1987; Ekman & Friesen, 1976). FACS coding, in turn, allows inferences about the six basic emotions, because the combination of certain AUs are associated with certain emotions. For instance, activating AU 4 (i.e., brow lowerer; *corrugator supercilii*) leads to a lowering of the eyebrows. This movement typically occurs when expressing emotions such as anger, disgust or sadness (Du et al., 2014; Ekman & Friesen, 2003).

Measuring Facial Expressions

In addition to human observation and coding of facial expressions (e.g., by means of FACS), there are two automated methods of measuring emotions by means of facial expressions (cf. Cohn & Sayette, 2010; iMotions, 2016; Wolf, 2015): *facial electromyography activity* and computer-based video classification algorithms (e.g., AFFEDEX, FACET, or FaceReader).

Facial Electromyography

Facial electromyography activity (fEMG) directly measures electrical changes in facial muscles and thus can record even subtle facial muscle activities. However, fEMG requires special biosensors placed on the face, is sensitive to motion artifacts and can be intrusive (i.e., the electrodes can raise awareness about the measurement, see Schulte-Mecklenbeck, et al., 2017 for a discussion on intrusiveness of process tracing methods). Further, the number of muscles fEMG can capture is limited by how many electrodes can be attached. Further disadvantages of fEMG are crosstalk signals that originate from surrounding muscles and impede the analysis of specific muscles, and that the direction of a specific muscle activity cannot be detected. Other than with human FACS coding or automated facial expression analysis, it is therefore often not possible to clearly classify a distinct emotion. For instance, fEMG is not able to identify which emotion is shown when it captures brows moving, because this muscle activity is associated with various different emotions including anger and sadness (Huang, Chen, & Chung, 2004; iMotions, 2016; Stets & Turner, 2014; Wolf, 2015).

Automated Facial Expression Analysis

In the last decade, most advancements in the area of automated facial expression analysis were on detecting distinct basic emotions (and their intensity) and specific facial muscle activities (El Kaliouby & Robinson, 2005; Lewinski et al., 2014; Valstar et al., 2011; Zeng, Pantic, Roisman, & Huang, 2009; for a review see Calvo et al. 2014).

CERT (precursor of FACET; Littlewort et al., 2011) and Noldus's FaceReader (den Uyl & van Kuilenburg, 2005) are the first software tools that were developed to automatically classify static (i.e., still picture) and dynamic (i.e., video) facial expressions. Since then, the market for automated facial expression analysis has changed rapidly. Currently, there are three major software tools for automated AU identification and emotion classification: Noldus's FaceReader (den Uyl & van Kuilenburg, 2005), iMotions's AFFDEX module (El Kaliouby & Robinson, 2005; Zeng et al., 2009) and iMotions's FACET module (Littlewort et al., 2011).

It is important to note that there is an ongoing debate about appropriate theory integration of automated emotion classification. These tools generate probability(-like) measures for basic emotions and are trained with databases of prototypical facial expressions of basic emotions. However, the underlying paradigm of six distinct basic emotion categories (Ekman, 1992a, 1992b) is rather simplistic (Du, Tao, & Martinez, 2014) because real life facial expressions often reflect compound (vs. distinct) emotions. People rarely show spontaneous facial expressions that are prototypical (Naab & Russel, 2007; Scherer & Ellgring, 2007) and often experience and express emotional states that cannot be assigned to only one basic emotion (Scherer, Wranik, Sagsue, Tran & Scherer, 2004). Furthermore, people do not only express emotions with their faces. Facial expressions can also be reflected through other means, e.g., social interactions. Moreover, the possibility of regulating emotional feeling states by altering outward facial expressions (ranging from hiding or suppressing facial expressions to portraying facial expressions which don't reflect inner feeling states) adds to the complexity of the picture (Gross, 2002). Thus, it is often not clear how the expressive facial patterns are related to underlying emotional processes (Barrett & Wager, 2006; Ekman, 1992b; Ortony & Turner, 1990; Russell, 2003; Scherer, 2005). This criticism implies that, though automated facial expression analysis recognizes basic emotional expression categories, it does not ultimately measure emotional states. Clearly, the fact that automated facial expression analysis relies on the assumption of emotional coherence (i.e., that there is coherence between emotion and facial expression (cf. Bonanno & Keltner, 2004; Reisenzein, Studtmann, & Horstmann, 2013)) limits the interpretation of data generated by automated facial expression analysis and questions the generalizability of automated emotion classification (Wolf, 2015). Some researchers argue that inference based on data generated by automated facial expression analysis should build upon emotion theories that go beyond the distinct-emotion perspective, adopt an appraisal perspective and allow more flexibility to consider different contexts. (For an extended overview of the proposition of a paradigm shift from categorical emotion recognition to an appraisal perspective, see Vallverdu, 2014.)

Measuring Emotions with iMotions's Facial Expression Analysis

Initially, iMotions implemented automated facial expression analysis by introducing the software module FACET. This module is based on the FACET algorithm (cf. Littlewort et al., 2011) developed by Emotient. In 2016, iMotions announced a switch to facial expression analysis technology from Affectiva, and introduced the AFFDEX module. (This switch was most likely connected to the acquisition of Emotient by Apple.) While new customers of iMotions are only able to purchase AFFDEX, existing customers are still able to apply FACET until 2020 (personal conversation with iMotions employee, 2016)[1].

Surprisingly, there is only a little evidence that automated facial expression analysis is as reliable as human FACS coding and fEMG (Lewinski et al., 2014; Littlewort et al., 2011; Terzis, Moridis, & Economides, 2010). A validation study of FaceReader (Version 6; den Uyl & van Kuilenburg, 2005; Lewinski et al., 2014) resulted in a classification accuracy of 88% of the emotions in the *Warsaw Set of Emotional Facial Expression Pictures* (WSEFEP) and of 89% of the *Amsterdam Dynamic Facial Expression Set* (ADFES), two publicly available datasets of validated facial expressions of emotions. In terms of specific emotions, FaceReader performs best for happiness (classification accuracy of 96% for WSEFEP and ADFES) and worst for anger (classification accuracy of 76% for WSEFEP and ADFES). FaceReader correctly recognizes 94% of neutral faces. Although Lewinski et al. (2014) provide a first estimation of the automated classification accuracy, we see room for

[1] For a detailed description of the technical background, the data generation and analytics of iMotions's facial expression analysis, see Appendix A.

improvement given that it is not clear what criteria the authors apply to classify a picture as correctly recognized (cf. Lewinski et al., 2014).[2]

There is currently no research available which validates and compares iMotions's AFFDEX and FACET modules. We aim to close that gap with this research.

Research Overview

We performed two studies to validate and compare the performance of iMotions's facial expression analysis modules AFFDEX and FACET (iMotions, 2016). In Study 1, we adapted a validation procedure based on that of Lewinski et al. (2014) by computing accuracy measures for recognizing facial expressions in images from three databases of normed facial expressions. In Study 2, we exposed participants to emotionally evocative pictures. We computed accuracy measures for the emotional content of the pictures and for participants' facial expressions.

Study 1

Method

Design and Procedure. We measured the accuracy of emotion classification of iMotions's AFFDEX and FACET using three publicly available databases of facial expression pictures: WSEFEP (Olszanowski, Pochwatko, Kukliński, Ścibor-Rylski, Lewinski, & Ohme, 2008), ADFES (van der Schalk, Hawk, Fischer, & Doosje, 2011) and RaFD (Langner et al., 2010). For each module, a subset of 600 pictures from the three databases was analyzed. The emotion classification was conducted in an automated manner using iMotions. Therefore, we generated a video (MP4 format) for all faces in all emotional states separately for WSEFEP, ADFES and RaFD pictures. In the video, every pictures was shown for 5 seconds. For the analysis we cut the first and last second of data and analyzed the

[2] In addition, the FaceReader validation has not been conducted on the whole WSEFEP database (i.e., only on 207 instead of 210 picture). Furthermore, the authors neither specify exclusion criteria in their paper nor did they provide such information upon request.

'middle' 3 seconds to avoid the spikes the AFFDEX algorithm generates in its classification, mostly at the beginning of a classification.

Materials.

The Amsterdam Dynamic Facial Expression Set (ADFES). This database consists of dynamic (video) and static (still picture) facial expressions of 22 face models (van der Schalk, Hawk, Fischer, & Doosje, 2011). We only included the 153 static pictures (JPEG format, 1024 x 768 pixels) of the emotions anger, contempt, disgust, fear, happiness, sadness and surprise. (ADFES does not provide a picture of face model F10 expressing surprise.)

The Warsaw Set of Emotional Facial Expression Pictures (WSEFEP). This database consists of 210 pictures (JPEG format, 1725 × 1168 pixels) of 30 face models (Olszanowski et al., 2015). We only included the pictures of the emotions anger, disgust, fear, happiness, sadness and surprise. For technical reasons, it was not possible to generate a video for face model MK. Thus, we used 174 WSEFEP pictures for this study.

The Radboud Faces Database (RaFD). This database consists of 536 pictures of 67 face models expressing basic emotions (Langner et al., 2010). We only included 273 pictures of 39 white adults that express the emotions anger, contempt, disgust, fear, happiness, sadness and surprise.

Setting and Apparatus. iMotions's AFFDEX and FACET modules (Version 6.2) were used to classify the pictures from the three databases. We ran iMotions on a Lenovo T450s with Windows 8.1. Standard settings as described in the iMotions manual (cf. iMotions, 2016) were used.

Preliminary to the data analysis, we corrected the obtained measurements for all pictures by comparing them to individual baseline values. These baseline values were generated on the basis of neutral facial expression pictures (WSEFEP, ADFES and RaFD all provide a picture of a neutral facial expression for each face model). By subtracting individual baseline values from the raw evidence values for all emotions, we accounted for the face models' natural variations in facial expressiveness. iMotions provides measures for all basic emotions anger, disgust, fear, happiness, sadness, surprise including contempt (cf. iMotions, 2016).

Results

Matching Scores for Basic Emotions. Replicating the analysis technique of Lewinski et al., (2014) we computed a *Matching Score* (MS), which represents an estimate of iMotions's accuracy at recognizing facial expressions of basic emotions. MS is defined as the percentage of pictures that iMotions classified correctly (cf. Lewinski et al., 2014; Nelson & Russell, 2013). A classification was recorded as 'correct' when the highest evidence value (out of all evidence values for all basic emotions) matched with the database's emotion label. Thus, a higher MS indicate a greater likelihood of correctly classifying the target emotion. We computed MS for AFFDEX and FACET separately for each emotion (see Table 1)[3].

Overall, AFFDEX correctly recognized 72% of the emotions across the database pictures. AFFDEX recognized 73% of the emotions in ADFES, 67% of the emotions in WSEFEP and 75% of the emotions in RaFD. In contrast, FACET correctly recognized 95% of the emotions across all the database pictures. FACET recognized 97% of the emotions in ADFES, 89% of the emotions in WSEFEP and 98% of the emotions in RaFD. While AFFDEX failed to detect a face at all in 1% of the pictures, FACET's analysis did not result in any detection failures.

As Table 1 reveals, the algorithms performed differently for different emotions. Both modules correctly recognized 100% of the *happy* expressions for all databases (WSEFEP, ADFES, RaFD). AFFDEX showed poor accuracy with the emotions *fear* and *anger.*

---

[3] Data and analysis code from both studies is available at: https://github.com/michaelschulte/FacialExpressionAnalysis

Distinctness Index for Emotion Classification. To provide evidence on how distinct the matching for emotions (i.e., the MS) is we also computed a *Distinctness Index* (DI). The DI describes how confident the classification is by comparing how close the probability of the first predicted emotion is to the probability of the second predicted emotion. The DI is defined as the distance from the evidence value of the classified emotion to the evidence value of the next-highest-scoring emotion. Thus, higher DIs indicate a better performance of iMotions's classification and differentiation abilities. We computed average DI separately for all correctly recognized pictures for all emotions for AFFDEX and FACET. We z-transformed the DI, creating a standardized version (sDI) to allow a direct comparison of AFFDEX and FACET. (For detailed information about the form of AFFDEX's and FACET's evidence values see Appendix B.)

Table 1 summarizes the sDI for iMotions's AFFDEX and FACET for all basic emotions and picture databases. Whereas AFFDEX had an overall sDI of 0.10, FACET had an overall sDI of 0.05. Relatively low sDI (across all databases) for AFFDEX were found for the emotions *anger* and *fear*[4]. Relatively low sDI for FACET were found for the emotions *sadness* and *fear*.

Table 1

Classification Accuracy for Seven Basic Emotions Separately for the iMotions Modules AFFDEX and FACET

Emotion	Database	Number (AFFDEX) (FACET)	Matched (AFFDEX) (FACET)	MS (AFFDEX) (FACET)	sDI (AFFDEX) (FACET)

[4] Note that for fear, we could only compute the DI for ADFES because we had an MS of 0% for WSEFEP and RaFD.

Anger	ADFES	22	8 22	0.36 1	-1.06 -0.15
	WSEFEP	29	13 23	0.45 0.79	-0.54 -0.37
	RaFD	39[a]	23 39	0.59 1	-1.31 -0.38
Disgust	ADFES	22	21 22	0.95 1	-0.24 0.32
	WSEFEP	29[b]	24 28	0.83 0.97	-0.02 -0.33
	RaFD	39	37 39	0.95 1	-0.72 0.20
Fear	ADFES	22[c]	1 20	0.05 0.91	-1.07 -0.74
	WSEFEP	29	0 18	0 0.62	- -0.67
	RaFD	39	0 37	0 0.95	- -0.39
Happiness	ADFES	22	22 22	1 1	0.12 1.42
	WSEFEP	29	29 29	1 1	0.60 2.07
	RaFD	39	39 39	1 1	0.13 1.95
Sadness	ADFES	22	22 21	1 0.95	0.18 -0.78
	WSEFEP	29	21 27	0.72 0.93	0.02 -0.93
	RaFD	39	34 38	0.87 0.97	0.62 -0.64

Surprise	ADFES	21[d]	19 21	0.90 1	0.58 -0.20
	WSEFEP	29	29 29	1 1	0.50 -0.02
	RaFD	39	38 39	0.97 1	0.32 0.27
Contempt	ADFES	22	19 21	0.86 0.95	0.32 -0.42
	WSEFEP	-	-	-	-
	RaFD	39	33 36	0.85 0.92	0.78 -0.26
Total	ADFES	153[e]	112 149	0.73 0.97	0.08 -0.06
	WSEFEP	174[f]	116 154	0.67 0.89	0.21 0.03
	RaFD	273[g]	204 267	0.75 0.98	0.04 0.12
Average		600[h]	432 570	0.72 0.95	0.10 0.05

Note. Number = number of classified database pictures; Matched = number of pictures that iMotions classified correctly with the database's emotion label (true positives). MS = Matching Score. sDI = standardized Distinctness Index. FACET successfully detected all faces in the pictures. For AFFDEX, facial detection was successful for [a]38 of 39 | [b]26 of 29 | [c]21 of 22 and [d]19 of 21 database pictures. Thus, the number of detected faces for AFFDEX was [e]150 in the ADFES, [f]171 in the WSEFEP, [g]272 in the RaFD database and [h]593 in total (out of 600).

Study 1 provides the first evidence regarding iMotions's accuracy in classifying emotions of prototypical facial expressions from a standardized facial expression database. Specifically, FACET generally outperforms AFFDEX; the modules differ for the employed picture databases and distinct emotions. Given these results, we cannot make any inferences about iMotions's accuracy for natural (vs. prototypical) and dynamic (vs. static) emotional facial expressions.

In order to validate iMotions in a more natural setting with more subtle, dynamic facial expressions, Study 2 employed a validation procedure using human participants, with more natural facial expressions. Specifically, first iMotions's accuracy was examined when identifying participants' emotional facial expressions in response to emotional pictures. Second, iMotions's accuracy was examined when identifying emotional facial expressions in participants who were instructed to imitate pictures of facial expressions.

Study 2

Method

Participants. A total of 119 students of a Swiss University participated in this study. Only participants without facial artifacts (e.g., disruptive glasses, beard or scarves) were included. Data from 9 participants were excluded from the sample because of missing data, i.e., the software was not able to detect their face (due to technical problems, head movements and/or insufficient video quality). Specifically, participants were excluded from the sample when iMotions failed to generate sensor data for more than 10 percent of all displayed pictures. We considered iMotions to have failed in generating sensor data for a certain picture when it was not possible to detect a participants' face in more than 50 percent of all measurements. For every picture, sensor data from the first 177 frames of the 6 second, 30 Hz video recording was used (because iMotions did not record 180 frames for all pictures). The final sample consisted of 110 participants (63 female; $M_{Age} = 21.20$, $SD_{Age} = 5.20$). Three Amazon vouchers, worth 500 Swiss francs, were raffled among participants.

Design and Procedure. Participants signed a consent form declaring that they agreed to being filmed with a webcam. The study was part of a set of multiple unrelated studies and always ran first in the session. To ensure good data quality, the laboratory was evenly and clearly lit. Participants were seated in a chair in front of a screen and instructed to remain in a stable and straight position without their hands near their face. Subsequently, the experimenter asked participants to read the description of the study procedure and instructions on the screen.

In the first part of the study, participants were exposed to two blocks of emotionally evocative pictures (constant block order: IAPS, GAPED) and their facial expressions were recorded. Within these blocks, pictures were shown in random order. Each picture was presented for 6 seconds and was preceded by a neutral black slide with a white, centrally displayed fixation cross (3 s). The neutral slides provided baseline measurements for the classification.

In the second part of the study, participants were asked to imitate facial expressions for all pictures in the RaFD database for 6 seconds (i.e., as long as every picture was displayed). The RaFD pictures were separately displayed in a random order. Finally, participants were asked for demographics, were thanked and debriefed.

Materials.

Emotional Facial Responses to Emotionally Evocative Pictures. In order to capture iMotions's accuracy at detecting participants' emotional facial expressions in response to emotional pictures, we exposed participants to a subset of emotionally evocative pictures from the IAPS and GAPED database. Here, we rely on the assumption that there is coherence between the displayed pictures, participants' emotions, and their facial expressions.

The IAPS database consists of pictures (JPEG format, varying resolution) showing a wide range of emotional content, confirmed to be emotionally evocative (Lang, Bradley, & Cuthbert, 1999). Based on a valence assessment (ranging from *unpleasant* to *pleasant*), we chose four pictures. We chose the pictures with the most distinct (i.e., highest and lowest) valence. The specific picture numbers are: 1710, 1750 (highest valence showing puppies and bunnies); 9940, 9570 (lowest valence showing a hurt dog and an explosion[5]).

The GAPED database consists of pictures (JPEG format, 640 x 480 pixels) that include negative, neutral and positive emotional content (Dan-Glauser & Scherer, 2011). Based on a valence assessment (ranging from *very negative* to *very positive*), we chose two pictures, one with positive content (P067 showing a landscape; highest valence) and one with negative content (A075 showing a cow bleeding to death; lowest valence).

Imitation of Facial Expressions. As in Study 1, we used pictures from the RaFD database (Langner et al., 2010). We chose one female face model (female model number 01) looking frontal into the camera and showing the six basic emotions anger, disgust, fear, happiness, sadness and surprise. Participants were exposed to the six RaFD pictures and instructed to imitate the currently displayed facial expression.

Setting and Apparatus. We closely followed iMotions's recommendations for experimental setups. (For details see the definitive guide for facial expression analysis, https://imotions.com/guides/) The iMotions software (Version 6.2) ran on a Lenovo T450s with Windows 8.1 and an attached 24" (60 cm) BenQ XL2411Z screen to display the pictures. A Logitech C920 webcam recorded participants' faces. With this apparatus, sensor data (i.e., evidence values for basic emotions) was generated approximately every 32 ms for a total of 177 measurements (frames) for every picture.

Results

Emotional Facial Responses to Emotionally Evocative Pictures. We computed a MS (see Study 1 for details) to estimate the accuracy of classifying the valence of participants' responses to pictures with negative and positive emotionally evocative content.

[5] There were concerns about showing the lowest valence pictures (e.g., burn victims). Thus, less disturbing pictures with low valence were chosen.

Higher MS values indicate a greater likelihood of correct emotion classification. We computed MS separately for AFFDEX and FACET for the positive and negative picture set (IAPS, GAPED pictures; see Table 2).

Prior to computing the MS, we baseline-corrected evidence values for all basic emotions for participants' facial responses to all pictures. We subtracted the median of the baseline slide's evidence value for a certain emotion from all 177 frames of the pictures' evidence values for that emotion. Based on this, we identified the maximal evidence value for all emotions within the 177 measurements for every picture. Finally, these maximal evidence values were used to classify the valence of participants' facial responses as positive or negative. If a maximal evidence value was recorded for happiness, we labeled the facial response as positive. If a maximal evidence value was recorded for anger, contempt, disgust, fear or sadness, we labeled the facial response as negative (in accordance with the valence classification iMotions uses; iMotions, 2016). To compute the MS, we identified the number of detected participant faces and the number of correctly labeled facial responses for every picture. We coded participants' facial responses for a certain picture as 'correctly labeled' when the assigned valence label for the facial response matched the database's valence label.

As Table 2 reveals, AFFDEX classified 57% of all facial responses with the correct valence; it correctly classified 17% of facial responses to positive pictures and 97% of facial responses to negative ones. FACET classified 67% of all facial responses with the correct valence; it correctly classified 63% of facial responses to positive pictures and 71% of facial responses to the negative pictures. Table 2 provides a detailed overview of how iMotions performed for all facial responses to positive and negative pictures. Note that for some negative pictures (IAPS 9940, GAPED A075), iMotions (both modules) failed to detect participant faces.

Overall, results suggest that FACET better recognizes the valence of facial expressions than AFFDEX. Further, the two modules differ in their accuracy of classifying negative and positive valenced video recordings of participants displaying emotional expressions. This valence-specific difference is more pronounced for AFFDEX than for FACET.

Table 2

Classification Accuracy of Valence for Seven Basic Emotions Separately for iMotions's AFFDEX and FACET

Valence	Picture	Matched (AFFDEX) (FACET)	picturewise MS (AFFDEX) (FACET)	valencewise MS (AFFDEX) (FACET)	Overall MS (AFFDEX) (FACET)
Positive	IAPS 1710	29 77	0.26 0.70		
	IAPS 1750	20 62	0.18 0.56	0.17 0.63	
	GAPED P067	8 65	0.07 0.59		0.57 0.67
Negative	IAPS 9940	106[a] 79	0.97 0.72		
	IAPS 9570	106 74	0.96 0.67	0.97 0.71	
	GAPED A075	105[a] 80	0.96 0.73		

Note. Matched = number of participant faces that match the picture's valence (true positives). MS = Matching Score. [a]Face detection failed for one participant due to head movements.

Imitation of Facial Expressions. We computed the MS for estimating iMotions's accuracy when classifying emotions displayed on participants' faces when they imitate the basic emotions displayed in the RaFD pictures. MS is defined as the percentage of participants' imitations that iMotions matched with the correct emotion. We computed MS separately for AFFDEX and FACET for each RaFD picture (see Table 3). We applied the same baseline correction procedure as described above for the valence task.

Table 3 reveals that AFFDEX classified 55% and FACET 63% of all facial imitations with the correct emotion. MS differed considerably across emotions. While both modules were relatively accurate in recognizing posed facial expressions of *happiness* (AFFDEX: 91%; FACET: 98%), they performed poorly for posed facial expressions of *fear* (AFFDEX: 1%; FACET: 10%).

To provide evidence on how distinct these MSs are, we computed standardized DI (sDI) following the procedure described in Study 1. Table 3 provides DI for all RaFD pictures and both AFFDEX and FACET. Whereas AFFDEX had an overall DI of 0.02, FACET had an overall DI of 0.35. For AFFDEX, the lowest DI related to *fear* and the largest DI to *contempt*. For FACET, the lowest DI also related to *fear* and the largest DI to *happiness*.

Table 3

Classification Accuracy for Instructed Facial Expressions for Basic Emotions Separately for iMotions's AFFDEX and FACET

RaFD picture	Matched (AFFDEX) (FACET)	MS (AFFDEX) (FACET)	sDI (AFFDEX) (FACET)
Anger	54 86	0.49 0.78	-0.10 0.10

Contempt	75 47	0.68 0.43	0.87 -0.13
Disgust	87 90	0.79 0.82	-0.31 0.48
Fear	1[a] 11	0.01 0.1	-0.95 -0.85
Happiness	100 108	0.91 0.98	-0.52 1.52
Sadness	39[a] 44	0.35 0.40	-0.32 -0.66
Surprise	67 97	0.61 0.88	0.62 -0.01
Average	423 483	0.55 0.63	0.02 0.35

Note. Matched = Number of participant faces where the detected emotion matched the picture's emotion as reported in the dataset (female RaFD model number 01). MS = Matching Score. sDI = standardized Distinctness Index. [a]Face detection failed for one participant due to technical problems, head movements and/or insufficient video quality.

Study 2 provides the first evidence regarding iMotions's accuracy in classifying the emotions in natural and dynamic emotional facial expressions within a laboratory setting. Compared to iMotions's accuracy for classifying standardized, prototypical facial expression pictures (Study 1), Study 2 reveals reduced accuracy for people's natural facial responses to diverse emotionally evocative pictures. The accuracy of iMotions differs for distinct emotions (and valence), and is higher for FACET than for AFFDEX.

General Discussion

This research validates iMotions's facial expression analysis modules AFFDEX and FACET as software-based tools for emotion classification. When identifying prototypical facial expressions from three picture databases (Study 1), we find overall accuracy of 72% for AFFDEX and 95% for FACET. When using participants instead of prototypical pictures, accuracy drops for the valence of people's facial responses to diverse emotionally evocative pictures (57% for AFFDEX, 67% for FACET; Study 2). Taken together, iMotions's performance is better for recognizing prototypical static versus more natural dynamic facial expressions, and shows different results for distinct emotions (and valence). Overall, FACET outperforms AFFDEX on nearly all measures.

Validation and Comparison of iMotions (AFFDEX and FACET)

This research contributes by independently measuring and comparing the performance of iMotions's AFFDEX and FACET modules, and making the results publicly available for a broad audience. In general, there is support for the idea that automated facial expression analysis is technically feasible (e.g., Baltrusaitis et al., 2016; Bartlett, Hager, Ekman, & Sejnowski, 1999; Lien, Kanade, Cohn, & Li, 1998; Littlewort, Bartlett, Fasel, Susskind, & Movellan, 2006; Meiselman, 2016; Vallverdu, 2014). Moreover, it is evident that automated facial expression analysis (e.g., Noldus's FaceReader) can produce valid data for prototypical facial expressions that are recorded under standardized conditions (Lewinski et al., 2014; Littlewort et al., 2006; Valstar et al., 2011). This research shows how far these general findings are true for iMotions.

The present findings support the skepticism that the current automated facial expression analysis is not yet mature enough for operational use (Meiselman, 2016) by revealing that, while iMotions's automated facial expression analysis can produce data with an acceptable degree of accuracy for prototypical facial expressions, it is less accurate for subtle, more natural facial expressions.

Accuracy measures for AFFDEX and FACET show that iMotions can provide data as valid as that produced by human judges. Human performance in recognizing emotions in prototypical facial expressions in database pictures is often situated between 60% and 80% and normally does not attain 90% accuracy (Nelson & Russell, 2013). Human judges are usually better at selecting the correct emotion label for *happy* than for other emotional facial expressions. When discriminating between non-happy expressions (i.e., anger, disgust, fear, sadness, surprise), judges' accuracy in recognizing emotions is particularly weak for fearful faces (Calvo et al., 2014; Nelson & Russell, 2013). Testing iMotions's accuracy (on similar pictures of prototypical emotions; Study 1) reveals comparable performance to human judges. One can also compare the performance of human judges and iMotions for identical sets of facial expressions. For the WSEFEP and ADFES databases, human judges have a performance of 85% (cf. Lewinski et al., 2014; Olszanowski et al., 2015; van der Schalk et al., 2011). The performances of the AFFDEX and FACET modules are 70% and 93% respectively (Study 1)[6]. While AFFDEX's accuracy is in the middle of the range of the accuracy of human judges (i.e., 60–80%), FACET's accuracy seems to outperform human judges. Moreover, results show that, like human judges, iMotions's accuracy differs for distinct emotions and performs particularly well (poorly) for happy (fearful) faces.

A comparison of iMotions's automated facial expression analysis modules with Noldus's FaceReader leads to similar inferences. Lewinski et al., (2014) found FaceReader to correctly classify 88% of emotions in the WSEFEP and ADFES pictures. According to the results of Study 1, iMotions's AFFDEX shows lower performance than Noldus's FaceReader (70% vs. 88%); however, FACET outperforms Noldus's FaceReader (93% vs. 88%). One implication of this result may be that based on the larger number of facial landmarks used (6 versus 34) FACET outperforms AFFDEX. It is important to consider that the present comparison of the performance of Noldus's FaceReader and iMotions's AFFDEX and FACET could be biased because producers do not consistently disclose databases used in the algorithm's training set. If one facial expression analysis engine but not the others includes

[6] As per Lewinski et al. 2014, we computed unweighted MS for the AFFDEX and FACET module based on the MS for the ADFES and WSEFEP.

WSEFEP or ADFES in the machine learning process, then this will result in an overestimated relative accuracy.

Regarding a direct comparison of the validity of automated facial expression analysis with human FACS coders, two problems arise. First, automated facial expression analysis is based on FACS and uses FACS classified pictures as training database. Second, FACS coders primarily describe AUs (i.e., anatomically independent facial muscle movements) and do not directly measure emotions. Looking into the literature reveals that many studies on FACS coder accuracy focus on performance on certain AUs rather than on emotion classification (cf. Lewinski et al., 2014). Clearly, certain AU configurations are associated with certain basic emotions. Such predictions of emotions, however, involve comprehensive definitions of AU configurations and consistent decisions on which (variants of prototypical) AU configurations account for a certain basic emotion. This makes direct comparisons unreliable.

Limitations of the Present Research

The standardized and controlled setting may impede generalizability of our results. Study 1 classifies prototypical, static facial expressions that are uncommon in real-life situations. Accuracy measures are thus likely to be inflated. Study 2 partially addresses this limitation by using more natural, dynamic facial expressions within a controlled laboratory setting. Still, real-life settings differ from laboratory settings in motion and uneven light and color.

We also build on the assumption that positive (negative) pictures elicit positive (negative) facial responses. This assumption, however, is controversial. Facial expressions occur for various reasons: they can be generated internally (e.g., by thoughts or memories), produced by external stimuli (e.g., photographs or films; Reiman et al., 1997) and be determined by social interaction and display rules (Vallverdu, 2014). Further, positive (negative) stimuli do not only produce positive (negative) facial expressions but also expressions that are reserved for negative (positive) emotions or a mix of diverse emotions (Aragón, Clark, Dyer, & Bargh, 2015; Fredrickson & Levenson, 1998). We thus cannot be sure whether our positive (negative) pictures were actually effective in eliciting the intended valence in participants' faces. This ties into the finding that iMotions's performance is better at recognizing negative versus positive facial expressions. It is important to refer to a bias that is introduced by iMotions valence classification: According to this classification, positive valence is recorded for happiness and negative valence for anger, contempt, disgust, fear and sadness (iMotions, 2016). Hence, simple probability (i.e., positive valence is only recorded for one emotion while negative valence is recorded for five emotions) calls into question the conclusion that iMotions's performance is better for negative versus positive facial expressions.

A second limitation of Study 2 is that we rely on the assumption that participants can imitate pictures of emotional facial expressions. In fact, we do not know how accurately participants imitated the displayed facial expressions. Results of Study 2 could therefore be confounded by limitations in participants' ability to imitate emotions accurately; we cannot rule out that iMotions would actually perform better.

Overall, these limitations substantiate the need to improve the application of automated facial expression analysis in real life settings. It is thus not surprising that affective computing researchers are currently addressing issues such as varying camera angles and changing head poses. Improvements are also needed in analyzing non-posed faces, the sensitivity of measuring subtle changes in facial expressions and the discrimination of more difficult expressions (i.e., compound emotions) and expression intensity (see, e.g., Facial Expression Recognition and Analysis challenge 2015 (www.sspnet.eu/fera2015/) and 2017 (www.sspnet.eu/fera2017/); McDuff, et al., 2010).

Other, more theoretical, limitations become apparent when interpreting the present results under consideration of the ongoing debate about an appropriate theory for automated facial expression analysis. Automated facial expression analysis tools typically generate probability(-like) measures for distinct basic emotions and are trained with databases of prototypical facial expressions. Thus, these tools are often successful with prototypical facial expressions (Lewinski et al., 2014; Vallverdu, 2014). This prototypical perspective, however, is problematic as it limits the generalizability of automated facial expression analysis. There are many types of facial expressions that vary in their distinctness and intensity, ranging from subtle to very intense (Ekman, Friesen & Ancoli, 1980; Hess, Banse, & Kappas, 1995). In the present research, we did not distinguish between measuring prototypical versus natural facial expressions; i.e., Study 1 and Study 2 were not designed to compare iMotions's accuracy. Nevertheless, it seems unsurprising that the present research found higher accuracy when classifying posed, intense facial expressions (Study 1) rather than subtle, more natural facial expressions (Study 2).

Implications for Researchers and Practitioners

There are various approaches to measuring emotions, from verbal ratings to nonverbal indicators. The advantages of automated facial expression analysis are low time and labor costs, simplicity and the potential for less intrusive measurements (cf. iMotions, 2016; Meiselman, 2016). Thus, valid automated facial expression analysis offers opportunities in diverse fields of emotion research, not only for academics but also for practitioners such as marketers or IT providers. The commercial application of such tools, for example in smartphones, media and advertisement testing, or even the design of avatars, has recently become pronounced (cf. iMotions, 2016; Lee, Sang Choi, Lee, & Park, 2012). In the future, academics could use such tools to efficiently validate new databases of prototypical basic emotional expressions.

In view of this need for valid facial expression analysis tools, it would be advantageous if providers of automated facial expression analysis would not only improve the validity of their products further, but also provide transparent and complete product information that complies with scientific requirements. For instance, development and algorithmic details should be clear and sufficiently documented; the databases on which the algorithms are trained should be specified; and details on the generation and interpretation of sensor data, as well as on the validity of this data, should be available.

We encourage researchers to define and apply standard methods to validate and compare automated facial expression analysis tools. The present accuracy measures, for instance, could be used to (re-)validate (updated) automated facial expression analysis tools in a standardized manner. To a certain extent, these accuracy measures could also serve to compare automated facial expression analysis with other measurement methods.

Conclusion

Two validation studies reveal that iMotions has the potential to measure basic emotions expressed by faces. iMotions performs better for prototypical versus natural facial expressions, and shows different results depending on the studied emotion. iMotions's FACET module outperforms the AFFDEX module.

References

Aragón, O. R., Clark, M. S., Dyer, R. L., & Bargh, J. A. (2015). Dimorphous expressions of positive emotion: Displays of both care and aggression in response to cute stimuli. *Psychological Science*, *26*(3), 1–15.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality and Social Psychology*, *103*(1), 20–37.

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016). OpenFace: An open source facial behavior analysis toolkit. Proceedings from *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1–10). IEEE.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.

Barrett L. F., & Wager T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science, 15,* 79–83.

Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology*, *36*(02), 253–263.

Bonanno, G., & Keltner, D. (2004). Brief Report: The coherence of emotion systems: Comparing “on-line” measures of appraisal and facial expressions, and self-report. *Cognition and Emotion*, *18*(3), 431–444.

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., & Nummenmaa, L. (2014). Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*(4), 549–567.

Cohn, J. F., & Sayette, M. A. (2010). Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. *Behavior Research Methods*, *42*(4), 1079–1086.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477.

Den Uyl, M. J., & van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. In *Proceedings of Measuring Behavior 2005* (pp. 589–590).

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, *111*(15), 1454–1462.

Ekman, P. (1992a). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Ekman, P. (1992b). Are there basic emotions? *Psychological Review, 99*(3), 550–553.

Ekman, P. (1992c). Facial expressions of emotion: New findings, new questions. *Psychological Science*, *3*(1), 34–38.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75.

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

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, *39*(6), 1125–1134.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Cambridge, MA: Malor Books.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K. … Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712–717.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, *30*(1), 527–554.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*(2), 203–235.

El Kaliouby, R., & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In B. Kisačanin, V. Pavlović & T. S. Huang (Eds), *Real-time vision for human-computer interaction* (pp. 181–200). New York, NY: Springer.

Fredrickson, B. L., & Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition & Emotion*, *12*(2), 191–220.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291.

Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology*, *69*(2), 280–288.

Huang, C.-N., Chen, C.-H., & Chung, H.-Y. (2004). The review of applications and measurements in facial electromyography. *Journal of Medical and Biological Engineering*, *25*(1), 15–20.

iMotions (2016). *Facial Expression Analysis: The definitive guide*. Retrieved from https://imotions.com/facialexpression-guide-ebook

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida. Retrieved from http://www2.unifesp.br/dpsicobio/Nova_versao_pagina_psicobio/adap/instructions.pdf

Lee, H., Choi, Y. S., Lee, S., & Park, I. P. (2012). Towards unobtrusive emotion recognition for affective social communication. In *Consumer Communications and Networking Conference (CCNC), 2012 IEEE* (pp. 260-264). IEEE.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236.

Lien, J. J., Kanade, T., Cohn, J. F., & Li, C.-C. (1998). Automated facial expression recognition based on FACS action units. Proceedings from *Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 390–395). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/670980/

Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., & Movellan, J. (2006). Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, *24*(6), 615–625.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). Proceedings from *2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011),* (pp. 298–305). IEEE.

McDuff, D., El Kaliouby, R., Kassam, K., & Picard, R. (2010). Affect valence inference from facial action unit spectrograms. Proceedings from *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition — Workshops* (pp. 17–24). IEEE.

Meiselman, H. L. (2016). *Emotion measurement*. Cambridge, UK: Woodhead.

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, 7, 736–744.

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*(1), 8–15.

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*, *5,* 1–8.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological Review*, *97*(3), 315–331.

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., … Chen, K. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *American Journal of Psychiatry*, *154*(7), 918–925.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, *5*(1), 16–23. https://doi.org/10.1177/1754073912457228

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, 7, 113–130.

Scherer, K. R., Wranik, T., Sangsue, J., Tran, V., & Scherer, U. (2004). Emotions in everyday life: Probability of occurrence, risk factors, appraisal and reaction patterns. *Social Science Information, 43*, 499–570.

Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social Science Information*, *44*(4), 695–729.

Stets, J. E., & Turner, J. H. (2014). *Handbook of the Sociology of Emotions* (Vol. 2). Heidelberg, Germany: Springer.

Swinton, R., & El Kaliouby, R. (2012). *Measuring emotions through a mobile device across borders, ages, genders and more*. ESOMAR. Retrieved from http://www.affectiva.com/wp-content/uploads/2014/09/Measuring_Emotion_Through_Mobile_Esomar.pdf

Taggart, R. W., Dressler, M., Kumar, P., Khan, S., & Coppola, J. F. (n.d.). Determining emotions via facial expression analysis software. Retrieved from http://csis.pace.edu/~ctappert/srd2016/2016PDF/c2.pdf

Terzis, V., Moridis, C. N., & Economides, A. A. (2010). Measuring instant emotions during a self-assessment test: The use of FaceReader. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research* (p. 18). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1931362

Vallverdu, J. (2014). *Handbook of research on synthesizing human emotion in intelligent systems and robotics*. Hershey, PA: IGI Global.

Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., & Scherer, K. (2011). The first facial expression recognition and analysis challenge. Proceedings from *2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011),* (pp. 921–926). IEEE.

Van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920.

Van Kuilenburg, H., Wiering, M., & den Uyl, M. (2005). A model based method for automatic facial expression recognition. In *European Conference on Machine Learning* (pp. 194–205). Porto, Portugal.

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience*, *17*(4), 457–462.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58.

Appendix A

Technical Background of the FACET and AFFDEX Module

Software-based facial expression analysis is based on algorithms that apply a set of distinct rules based on psychological theories and statistical procedures. In general, these algorithms work in three steps: (1) detecting faces, (2) detecting facial landmarks and (3) classifying emotions.

In step 1, a face in a picture or video is detected (e.g., by applying the Viola–Jones cascaded classifier algorithm) (iMotions, 2016).

In step 2, facial landmarks (e.g., eye and mouth corners and nose tip) are identified (e.g., by an active appearance model; Cootes & Taylor, 2004). While the FACET module uses six facial landmarks, the AFFDEX module uses 34 facial landmarks. Based on these landmarks, an artificial reticular face model that represents a simplified version of the actual face is developed. The face model adopts the position, size, and scale of the actual face and is adjusted instantly when the face moves.

In step 3, a classification algorithm is used to translate the landmarks into values for emotional states. Specifically, the classification expresses how likely it is that the target face shows a certain basic emotion (or doesn't, if the result is negative). This classification procedure is based on machine-learning methods. That is, the classification tool is trained on a facial database. Since all facial expression analysis software tools are trained on different facial databases, it is possible that they would provide slightly different results for a certain face (iMotions, 2016).

Appendix B

Data Generation and Analytics

The translation from facial landmarks into values is based on a statistical procedure that compares the expression on the actual face with the expressions of the faces stored in the database. The value that iMotions assigns to every basic emotion is the evidence value.

The evidence value for the FACET module is similar to a z-score. That is, positive (negative) values indicate evidence that a certain emotion is expressed (absent). While values larger than three indicate strong evidence for the presence of a certain emotion, negative values smaller than three indicate strong evidence for the absence of an emotion. An evidence value of zero indicates that there is no evidence either way. (For more details, see iMotions help center, 2016).

The evidence value returned by the AFFDEX module represents probabilities from zero to one. Accordingly, a value of zero indicates no evidence and a value of one the highest evidence that a certain emotion is fully expressed.

With these characteristics, evidence values are sufficient for descriptive and inferential statistics.

When it comes to the interpretation of iMotions sensor data, one should differentiate between the raw evidence values and the baseline-corrected evidence values.

The raw evidence values represent the classification for a certain emotion of a face compared to the facial expressions in the global database. With this, an individual's facial expression can be directly compared with countless other facial expressions. This is useful when one is interested in aggregating data of several individuals or comparing data from two or more groups of individuals (iMotions help center, 2016).

The baseline-corrected evidence values anticipate that people differ in their "neutral" expression. Baseline-corrected raw evidence values reflect the relative expression of an individual independent of the database. In some cases, it is useful to consider an individual's baseline as this allows you to detect changes in emotional expression relative to the individual's neutral expression. This helps to avoid misestimating the extent of a certain emotion when using raw evidence values (iMotions help center, 2016).

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 19.06.2017

Ort, Datum

Sabrina Stöckli