Journal of Articles in Support of the Null Hypothesis
Vol. 16, No. 1
 1539-8714
www.jasnh.com

Applying an implicit approach to research on the uncanny feeling

Javier Villacampa[1]
Gordon P.D. Ingram[2]
Guido Corradi[3]
Antonio Olivera-La Rosa[1,4]
[1] Human Evolution and Cognition Group, associated group to IFISC (University of the Balearic Islands – CSIC)
Carr. de Valldemossa, 07122, Palma de Mallorca, Spain
[2] Department of Psychology, Universidad de los Andes
Cra. 1 #18a-12, Bogotá, Colombia
[3] Departament of Psychology, Faculty of Health and Education, Universidad Camilo José Cela
[4] Department of Psychology and Social Sciences, Universidad Católica Luis Amigó
Tv. 51a #6790, Medellín, Colombia

Contradictory findings with regard to the nonlinear relation between human likeness and affective reactions have characterized psychological research on the uncanny valley hypothesis (Mori 1970/2005). In the present study we explored the phenomenology of the uncanny feeling (UF) by assessing implicit associations between uncanny stimuli (by android faces) and two emotional responses previously associated with the uncanny: fear and disgust. Further, we tested whether perception of uncanny stimuli would facilitate cognitions of deviant ("sick") morality and mental illness, as suggested by previous literature. Across five Single-Target Implicit Association Tests we found support only for a slight association of the UF with moral disgust (relative to fear). We found no evidence of an implicit link between the UF and fear or general disgust, nor did the UF implicitly facilitate cognitions of psychopathy.

Keywords: Cognitive psychology, emotion, uncanny feeling, fear, disgust, implicit measures

Corresponding author:
Antonio Olivera-La Rosa, Email: antonio.oliverade@amigo.edu.co / acensulay@yahoo.es
Address: Tv. 51a #6790, Medellín, Antioquia (Colombia)
Phone number: +57 (4) 4487666 Fax number: +57 (4)3849797

Introduction

The "uncanny valley" hypothesis (Mori 1970/2005) states that entities which look close to being human can produce negative feelings in an observer. Originally, this hypothesis was designed to describe how emotional responses vary with perceived human likeness in robots: the more human-like robots look, the more pleasant they are experienced, until a point is reached at which they start to elicit a distinctive negative emotional response (the uncanny feeling, UF). Despite a rise in interest in conducting empirical research on the phenomenon, contradictory findings have raised concerns among researchers about its scientific standing and even its theoretical plausibility (Burleigh, Schoenherr, & Lacroix, 2013; Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Wang, Lilienfeld, & Rochat, 2015). The present study addresses one central aspect underlying conflicting evidence in uncanny valley research: the absence of a detailed characterization of UF phenomenology.

The UF as an evolved mechanism

In the last few years, empirical research on the uncanny valley has centered on psychological explanations of the UF. In keeping with our intended focus on the particular phenomenology of the UF, we review only those hypotheses that focus on the emotional nature of the UF. Therefore, we do not review here the more "cognitive" hypotheses explaining the UF, such as the categorization difficulty hypothesis (Burleigh et al., 2013; Yamada, Kawabe, & Ihaya, 2013) and the perceptual mismatch hypothesis (Feng et al., 2018; MacDorman & Chattopadhyay, 2016; Seyama & Nagayama, 2007; for reviews of the most influential psychological explanations of the UF, see Kätsyri et al., 2015; Wang et al., 2015). Instead, we devote our attention to two evolved psychological mechanisms that may be involved in UF phenomenology through the generation of unpleasant emotions: pathogen avoidance and mortality salience.

Mori himself, in developing the uncanny valley hypothesis (1970/2005), had originally proposed that the UF was related to self-preservation. His intuition was later developed by Keysers, who proposed that the UF should be understood as a part of an evolved pathogen avoidance mechanism designed to prevent contact with infectious entities and rooted in the emotion of disgust (MacDorman & Ishiguro, 2006). According to this perspective, perceived imperfections in a human-like entity are interpreted as indicative of a heightened risk for transmissible diseases, therefore causing the UF. Although theoretically plausible, empirical evidence on the pathogen avoidance hypothesis is still lacking. To the best of our knowledge, only one study by MacDorman and Entezari (2015) indirectly addressed the role of disgust in the UF by showing that individual differences in disgust sensitivity predicted differential sensitivity to the uncanny valley.

In contrast, proponents of Terror Management Theory (Goldenberg et al., 2001) argue that certain stimuli seem particularly threatening to humans because they bring to mind our vulnerability to death. Therefore, it has been proposed that some humanoid replicas (such as certain dolls, clowns, mannequins, sex toys, wax figures, humanoid robots, etc.) may be uncanny because they resemble dead, "soulless" individuals who seem to have come alive, in turn triggering defensive psychological systems to handle the resulting anxiety about mortality (Ho, MacDorman, & Pramono, 2008; MacDorman & Ishiguro,

2006). However, while there is some evidence that uncanny androids remind people of the dead (MacDorman & Ishiguro, 2006), it is not clear that androids necessarily induce fear. To date, only Wang and Rochat (2017, Study 2) have tested the mortality salience hypothesis at the implicit level, using a visual looming task. They found that those faces that induced the UF also elicited shorter estimated time-to-contact, which could indicate that participants felt under threat. However, as the authors admitted, further rigorous attempts to test this theory are clearly needed.

The UF, a controversial emotional response

Indeed, research on the emotional components of the UF is still incipient, which may be a major cause of the controversial findings in the field. For example, a crucial problem in empirical approaches to the uncanny valley is the absence of a clear definition of the UF (the main dependent variable). Therefore, different researchers have interpreted the UF in divergent ways, in terms of "familiarity" (MacDorman & Ishiguro, 2006), "acceptability" (Hansson, 2005), "eeriness" (Mitchell et al., 2011), "valence" (Cheetham, Suter, & Jäncke, 2011), "likability" and "affinity" (Katsyri et al., 2015).

Interestingly, theoretical discussion of the UF can be traced long before the concept of the uncanny valley itself, to the early 20th century, when some authors regarded uncanniness as a kind of fear relating to things that are uncertain and unfamiliar (Jentsch, 1906/1997; Freud 1919/1964). More recently, Mangan (2015) understood the UF as a fringe experience, which involves appraisals of familiarity and wrongness, and feelings of threat and disgust. At the empirical level, Ho and colleagues (2008) showed that judgments of "creepy" and "eerie" are involved in the uncanny response to a greater degree than judgments of "strangeness". This is important because while eeriness and creepiness are principally associated with fear (but also with disgust, nervousness, and shock), strangeness seems to involve less of an emotional response. Similarly, Burleigh and colleagues' (2013) data suggested that eeriness was mainly linked to fear, but also to disgust and dislike. Finally, Wang and Rochat (2017) found that feelings of eeriness, disgust, unsettlingness, attractiveness, threateningness and likability were all involved in the UF, suggesting that the UF may implicate a variety of emotional responses.

Other studies have explored the emotional feeling of "creepiness" outside the context of the uncanny valley. McAndrew and Koehnke (2016) found evidence supporting the hypothesis that the "creepy" psychological reaction is related to anxiety caused by the presence of an ambiguous threat. These results were extended by Watt, Maitland, and Gallagher (2017), whose findings suggested that creepiness was associated with a sense of social ambiguity and that creepiness evaluations were most easily caused by facial features. Building on these insights, it has been proposed that the UF can be seen as "an unpleasant emotional response triggered by violations of humanness expectations, which is experienced as a disturbing, macabre feeling that something is ambiguously wrong ('at odds') with the humanness of the human-like entity" (Olivera-La Rosa, 2018, p.41).

Wrong Outside, Wrong Inside?

We believe that upcoming evidence on the phenomenology of the UF may offer new insights into the function of this emotional response. In this vein, research by Tinwell,

Nabi, and Charlton (2013) demonstrated that viewers' perception of characters as uncanny influenced their ascription of psychopathic traits, suggesting that certain facial expressions are linked both to perceptions of psychopathy and to the UF. Further, Olivera-La Rosa (2018) claimed that perceptions of uncanny physical features in a human-like entity influence its "moral status" by signaling that something is "not right" with the entity, and that it should therefore be avoided. Two major bodies of research support these hypotheses. First, there is a great deal of evidence that people draw multiple social inferences about a person from minimal facial cues (Todorov, Mendle-Siedlecki, & Dotsch, 2013; Todorov, Said, Engell, & Oosterhof, 2008). For instance, deciding whether somebody is trustworthy and cooperative is a highly automatic process that may rely on appraisals of face typicality (Sofer, Dotsch, Wigboldus, & Todorov, 2015). These data are consistent with Mathur and Reichling's (2016) results showing that trust-motivated behavior follows a UV-related pattern (i.e., robot faces tended to elicit less trust when they were somewhat human-like).

Second, it is widely supposed that, in daily life, intuitive/emotional processes play a crucial role in moral judgments (Haidt, 2001). In particular, people often use their "gut feeling" as a kind of embodied evaluation of social events (Bargh, Schwader, Hailey, Dyer, & Boothby, 2012; Eskine, Kacinik, & Prinz, 2011; Schwarz & Clore, 1983). For instance, several studies suggest that, depending on certain moderators, embodied disgust influences moral judgments (Laakasuo, Sundvall, & Drosinou, 2017; Schnall, 2016; Schnall, Haidt, Clore, & Jordan, 2015; van Dijke, van Houwelingen, De Cremer, & De Schutter, 2018; but see Johnson et al., 2016; Landy & Goodwin, 2015). Research on the automaticity of social judgments thus strongly suggests that affective responses can be used as diagnostic signals when making moral evaluations. The UF could well be another example of this.

The current study

In the present study we used implicit methods to investigate the phenomenology of the UF, which has been little studied despite the UF being the dependent variable in many recent empirical studies. Previous research on the UF has largely relied on explicit measures (e.g., self-report questionnaires) in order to assess participants' emotional reactions towards uncanny stimuli (MacDorman & Ishiguro, 2006; Seyama & Nagayama, 2007). This justifies the adoption of an implicit methodology to complement existing explicit data. Another important reason for using implicit measures is that they are indirect: that is, they allow us to measure a psychological construct without having to ask the participant for a self-report, which could be more vulnerable to social biases (Bohner, Siebler, González, Haye, & Schmidt, 2008). To the best of our knowledge, only two studies have applied an implicit methodology to assess the UF. In addition to the aforementioned study by Wang and Rochat (2017), Zlotowski and colleagues (2015) applied a brief implicit association test to test the effects of repeated interactions with a robot on perceived eeriness. Their implicit measure failed to replicate the finding obtained with explicit measures: while repeated interactions with the robot lowered the explicit ratings of eeriness, implicit results did not corroborate this effect. Therefore, a secondary objective of the present study is to assess the validity of an implicit methodology in researching the UF.

We believe that an implicit approach is also justified in the light of research on face perception. For instance, studies of first impressions suggest that initial perceptions develop dynamically after some minimal amount of relatively uniform information about a novel

target (Cone, Mann, & Ferguson, 2017). Consistent with this claim, some have argued that the UF should be understood as an example of automatic, stimulus-driven processing that occurs in the early stages of perception (MacDorman, Green, Ho, & Koch, 2009), which is consistent with current evidence showing that discriminating human facial expressions is a task for which the human visual system has developed specialized expertise (Hassin & Trope, 2000; Simpson, Varga, Frick, & Fragaszy, 2011; Willis & Todorov, 2006).

Hypotheses

Based on previous research suggesting that both fear and disgust are involved in the phenomenology of the UF (Burleigh et al., 2013; Ho et al., 2008), in this study we assessed implicit associations between uncanny stimuli (android faces) and the emotions of fear and disgust. By doing so, we aimed to contrast two hypotheses regarding the psychological explanation of the UF: the pathogen avoidance hypothesis (which suggests that the UF is driven by the emotion of disgust) and the mortality salience hypothesis (which suggests that it is driven by the emotion of fear).

Furthermore, we hypothesized that the perception of uncanny stimuli may facilitate judgments of deviant morality (i.e., moral disgust). According to Rozin and colleagues (2008), the reaction of moral disgust is limited to a subclass of particularly egregious moral offenses which reveal that an individual is lacking in normal human motives (people and behaviors that are seen as "sick" or "twisted"). Based on this claim, we theorized that the UF function may be similar to "moral disgust", in that both types of emotional response seem to be induced by encountering people who appear to lack normal human motivations (Hodson et al., 2014; Rozin et al., 2008). Yet there is a crucial difference: whereas moral disgust is a reaction to "twisted" *actions* (which are linked to appraisals of poor moral character; Uhlamn, Pizarro, & Diermeier, 2015), the UF may respond to "twisted" (i.e., weird or atypical) physical *features* (Olivera-La Rosa, 2018).

Based on these theoretical considerations, we designed our study in response to one overarching research question and three hypotheses.

RQ1. What are the implicit associations between uncanny stimuli (android faces), fear and disgust?
H1. Uncanny stimuli should be more strongly associated with "moral disgust" than with disgust or fear.
H2. Uncanny stimuli should be more strongly associated with "moral disgust" than with standard moral violations (e.g., stealing, cheating, or lying).
H3. Uncanny stimuli should be more strongly associated with mental illness than with physical disease.

Method

Participants

We recruited 176 participants (117 women, 59 men; mean age = 23.2, *SD* = 1.45). All participants were undergraduate students at one of two universities in Colombia, who

were invited to join the experiment as a part of their Psychology course credits. We recruited all of them via internal email or announcements in class, and they provided written consent in accordance with ethical procedures approved by the two universities. All participants had normal or corrected-to-normal vision and were over 18 years of age.

Material and Procedure

Materials and procedures were similar to those we had used in previous research on implicit processes (Olivera-La Rosa et al., 2017; Villacampa, Ingram, Martí-Vilar, & Olivera-La Rosa, 2018). We displayed the stimuli on a 20-inch screen (60Hz screen refresh rate) with a PC running OpenSesame v. 3.0.7 (Mathôt, Schreij, & Theeuwes, 2012) on Windows 8 (Microsoft Corporation). In order to assess the implicit associations of uncanny stimuli, we used the Single-Target IAT (ST-IAT; Karpinski & Steinman, 2006). In the ST-IAT, participants are asked to categorize each presented stimulus as quickly and accurately as possible. In a typical procedure, the experimenters assess the association of the target category (e.g., uncanny faces) towards each pole of a bipolar attribute category (e.g., "fear" vs. "disgust") through a series of categorization tests requiring prompt responses. The reasoning behind the ST-IAT is based on response interference or compatibility. If one has a stronger implicit association between uncanny faces and fear than between uncanny faces and disgust, it should be easier to classify uncanny stimuli and fear stimuli with a single key than to classify uncanny stimuli and disgust stimuli with the same key. The easiness of the task is evaluated through response latencies (i.e., reaction times; RTs): shorter latencies indicate easier stimulus/category assignment (i.e., less interference and more compatibility), suggesting a stronger implicit association (Bohner et al., 2008). Although the IAT it has been subject to some criticism recently (reviewed by Brownstein, Madva, & Gawronski, 2019)—notably concerning what implicit measures represent, especially in the context of "implicit bias"—it remains a robust and well-replicated method that has been shown to have important predictive capabilities across a wide range of domains (e.g., Greenwald, Banaji, & Nosek, 2015; Tello, Harika-Germaneau, Serra, Jaafari, & Chatard, 2019).

Participants completed five consecutive versions of the ST-IAT: (1) disgust versus fear; (2) "moral disgust" versus "ordinary" disgust; (3) "moral disgust" versus fear; (4) "moral disgust" versus standard moral violations; and (5) mental illness versus physical illness (Table 1). In the five ST-IATs, the target category ("uncanny faces") consisted of five female android faces obtained from a set of stimuli previously used in uncanny valley research (Mathur & Reichling, 2016). The latter authors measured participants' responses to a large, objectively chosen sample of real-world android faces using participants' explicit judgments of humanness (i.e., "how mechanical/human does this robot face look?") and likability (i.e., "estimate how friendly and enjoyable (versus creepy) it might be to interact with each face in an everyday situation") of each face. In order to select which android faces would represent the target category, we followed the criteria that all faces had to be rated below neutral on the likability scale and had to be scored above the midpoint of the humanness scale. We also considered it important that none of the selected faces should depict a canonical emotional expression (e.g., fear). On this basis, we selected faces 68, 71, 73, 77 and 79 (Appendix A; for a more detailed description of the stimuli see Mathur & Reichling, 2016). Furthermore, in order to control for cross-cultural differences between Mathur and Reichling (2016) U.S. sample and our Colombian sample, we recruited and additional 20 participants (12 women) for an explicit pretest of the preliminary selected

Table 1. *Category assignment and stimulus proportions across ST-IAT blocks for an exemplary participant*

				Number of stimuli						
Block	Task description	Left key concepts (z)	Right key concepts (m)	Disgust	Fear	MD	SMV	MI	PI	**UV**
1	Evaluative training trials	Disgust	Fear	10	10					
2	Initial block	Disgust + UV	Fear	10	15					**10**
3	Reversed block	Disgust	Fear + UV	15	10					**10**
4	Evaluative training trials	MD	Disgust	10		10				
5	Initial block	MD	Disgust + UV	10		15				**10**
6	Reversed block	MD + UV	Disgust	15		10				**10**
7	Evaluative training trials	Fear	MD		10	10				
8	Initial block	Fear + UV	MD		10	15				**10**
9	Reversed block	Fear	MD + UV		15	10				**10**
10	Evaluative training trials	MD	SMV			10	10			
11	Initial block	MD	SMV + UV			15	10			**10**
12	Reversed block	MD + UV	SMV			10	15			**10**
13	Evaluative training trials	MI	PI					10	10	
14	Initial block	MI + UV	PI					10	15	**10**
15	Reversed block	MI	PI + UV					15	10	**10**

MD: Moral Disgust
SMV: Standard Moral Violations
MI: Mental illness
PI: Physical illness
UV: Uncanny Valley

faces on the dimension of likability/eeriness (i.e., "estimate how friendly and enjoyable (versus creepy) it might be to interact with each face in an everyday situation"; see Mathur & Reichling, 2016). All faces were rated below neutral in this dimension.

Attribute categories consisted of five "ordinary" disgust words (e.g., *repugnante* [repugnant]), five fear words (e.g., *pavor* [dread]), five "moral disgust" words (e.g., *perversión* [perversion]), five ordinary moral violation words (e.g., *trampa* [cheating]), five mental illness words (e.g., *psicópata* [psychopath]) and five physical disease words (e.g., *fiebre* [fever]). All attribute items were selected from Emofinder (Fraga et al., 2018; see Appendix B). Following the same procedure as Bluemke and Friese (2008), each stimulus was presented at least twice, adding up to 35 trials per combined block. Target stimuli, coupled and uncoupled attribute stimuli occurred in a ratio of 10:10:15 trials (Table 1). The categorization task started with 20 trials for the training block, prior to the first combined block. This training block considered only two attribute categories (e.g., disgust and fear, but not uncanny faces) and the scores obtained were not explored in the analysis of the results. The order for the five versions of the ST-IAT was counterbalanced between participants, and the order of the item/category assignment trials was randomized within participants (Table 1).

Results

Based on past ST-IAT research (Bluemke & Friese, 2008; Olivera-La Rosa et al., 2017), we omitted participants who committed 30% or more errors (i.e., incorrect responses in the item/category assignment) in at least one of the five ST-IATs. Therefore, our final analysis was based on a sample size of 136 participants (91 women). Likewise, we recoded the trial latencies (RTs) that were below 300 ms (0.04% of the total trials of the task) or

above 3,000ms (2.9%) to the respective values (i.e., Rts below 300ms or above 3000ms were recoded to 300ms and 3000ms, respectively) and replaced the 4.63% trials that were errors by the block mean of correct latencies plus 600ms (for a detailed description of the rationale behind this procedure, see Greenwald, Nosek, & Banaji, 2003; Richetin et al., 2015).

ST-IAT effects were calculated on the basis of the attribute trials only, by computing the widely used D-measure algorithm (i.e., a computed score from attribute trials indicating an implicit association for one of the two attribute categories resulting from quicker response times if the target stimuli had to be sorted into the respective category; Greenwald et al., 2003). To explore ST-IAT effects we calculated five separate D-measures. These revealed that participants evidenced stronger implicit associations of uncanny stimuli (android faces) with moral disgust than with fear (D = -0.16). However, based on the conventional level of strength adopted by previous research (Blanton et al., 2015), this can only be considered a slight implicit association. No other D-measure reached the conventional level of strength (see Table 2). We found no sex differences in participants' implicit associations.

Table 2. *D-Measures for each ST-IAT*

ST-IAT	D-Measure
Fear - Disgust	-0,081094952
Disgust - Moral Disgust	-0,03929217
Moral Disgust - Fear	**-0,164542101**
Standard Moral Violations - Moral Disgust	0,059380563
Physical Illness - Mental Illness	-0,099479203

Discussion

Throughout the application of five ST-IATs, there was little support for the hypothesized effects. It is possible that the nature of the uncanny stimuli used in this study could cause a weakening of potential effects on the implicit measures. As Palomaki and colleagues (2018) have suggested, the lack of pre-validated stimulus materials known to reliably and robustly elicit the UF may have undermined the replicability of empirical results in this field. For instance, while they failed to replicate the uncanny valley effect when using non-photorealistic images of faces (e.g., half human/half robot; adapted from Ferrey et al., 2015), they did replicate the effect when using photorealistic stimulus material (e.g., actual photos of androids).

However, our results cast some doubts on the reliability of real android pictures to strongly elicit the UF. It is possible that the uncanny stimuli selected for the implicit task failed to elicit the complete ("full-blown") experience of uncanniness. Specifically, it may be that the use of female faces as uncanny stimuli may have weakened the strength of the UF. Indeed, some evidence suggests that men are more likely to be perceived as "creepy" than women (McAndrew & Koehnke, 2016; Watt, Maitland, & Gallagher, 2017), which is congruent with males being more physically threatening. Nevertheless, we found no sex differences in participants' implicit responses to the stimuli.

On the other hand, if the female android faces weren't appraised as real (i.e., plausibly found outside the lab), they may have failed to activate more fine-grained forms of cognition involved in the UF (e.g., perceptions of social threat; Olivera-La Rosa, 2018). As Smith (2014) has suggested, the uncanny may have a double perceived trajectory: from non-human (e.g., android) to human, and also from the natural and human towards the artificial self (e.g., over-users of Botox or cosmetic surgery). Therefore, it is possible that the moral connotations of the UF are stronger in connection with—or even limited to—the perception of "human" uncanny stimuli, which are presumably appraised as more

unnatural/grotesque than other (affectively subtler or more neutral) forms of non-human uncanny stimuli. Future research should address this question.

It is worth mentioning that participants' explicit ratings of likability/eeriness showed that all android faces were rated as "creepy" This suggests that, at least at an explicit level, our stimuli were good candidates to be perceived as uncanny. However, previous implicit attempts to capture the UF have found some explicit-implicit discrepancies, which indicate that IAT-like procedures may not be valid measures for capturing the UF (as suggested by Wang and Rochat, 2017, based on Zlotowski et al.'s, 2015, results). The results of the present study support this suggestion.

Further, the fact that we used female android faces which were highly symmetrical (and thus could potentially be seen as attractive if they belonged to real humans) may have created a confound in participants, for example by causing associations with sex dolls. It is difficult, however, to estimate at what degree explicit ratings of creepiness were affected by self-presentation bias, or if the stimuli indeed elicited the visceral aversive responses that are characteristic of the UF. Further studies are thus needed in order to test explicit-implicit discrepancies in the context of the UF. Another potential methodological factor explaining these results relates to the structure of the ST-IATs. The possibility cannot be ruled out that each ST-IAT was composed of two categories that were both strongly (but equally) associated with the UF. For instance, if participants were as afraid as they were disgusted by the android faces, their response latencies in both sorting tasks would have turned out comparably, leading to a D-measure close to zero. If this was the case, ST-IAT scores would not be very informative about UF implicit associations.

There are several general conclusions to be drawn from these results. One is that we found evidence for a slight association of the UF with moral disgust (relative to fear). This finding may be interpreted as preliminary evidence of an implicit link between the UF and moral appraisals, as suggested by some authors (Olivera-La Rosa, 2018; Tinwell et al., 2013). However, it must be remembered that this was only a small effect. Moreover our results failed to replicate the finding that perceptions of psychopathic traits were associated with the UF (Tinwell et al., 2013). Therefore, the state of knowledge about links between the UF and moral emotions or evaluations should be assessed with caution. Further research is needed to study how "humanlike" uncanny agents are perceived morally, given the potential social implications of this association.

Acknowledgments

We thank Evelyn Gil, Mateo Gutierrez, Estefania Jaramillo and Juan Sebastian Nassar for help with the experimental procedure. We also thank Maya Mathur for help with the stimuli selection and Jaume Rosselló for his helpful comments on this manuscript.

References

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380–393. doi:10.1037//0022-3514.83.2.380

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in cognitive sciences, 16*(12), 593-605.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology, 38,* 977–997. doi:10.1002/ejsp.487

Bohner, G., Siebler, F., González, R., Haye, A., & Schmidt, E. A. (2008). Situational flexibility of in-group-related attitudes: A single category IAT study of people with dual national identity. *Group Processes & Intergroup Relations, 11,* 301–317. doi:10.1177/1368430208090644

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science.* doi:10.1002/wcs.1501.

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior, 29,* 759–771.

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5*. doi:10.3389/fnhum.2011.00126

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 55, pp. 131–200). Amsterdam: Elsevier Inc.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441–1455. doi:10.1037/0022-3514.83.6.1441

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgments. *Psychological Science, 22,* 295–99.

Fraga, I., Guasch, M., Haro, J., Padrón, I., & Ferré, P. (2018). EmoFinder: The meeting point for Spanish emotional words. *Behavior Research Methods, 50,* 84–93. doi:10.3758/s13428-017-1006-3

Freud, S. (1964). The uncanny. In J. Strachey & A. Freud (Eds. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (pp. 217–256). London, UK: Hogarth Press. (Original work published 1919)

Goldenberg, J. L., Pyszczynski, T., Greenberg, J., Solomon, S., Kluck, B., & Cornwell, R. (2001). I am not an animal: Mortality salience, disgust, and the denial of human creatureliness. *Journal of Experimental Psychology: General, 130*(3), 427–435.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology,* 108(4), 553-561.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personalityand Social Psychology, 74,* 1464–1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. doi:10.1037/0022-3514.85.2.197

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychology Review, 108(4),* 814–834.

Hanson, D. (2005). Expanding the aesthetic possibilities for humanoid robots. In IEEE-RAS international conference on humanoid robots, Tsukuba, Japan.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78,* 837–852.

Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. (2008). Human emotion and the uncanny valley: A GLM, MDS, and Isomap analysis of robot video ratings. *In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, HRI '08* (pp. 169–176). Presented at the HRI '08, New York, NY.

Hodson, G., Kteily, N. S., & Hoffarth, M. R. (2014). Of filthy pigs and subhuman mongrels: Dehumanization, disgust, and intergroup prejudice. *Testing Psychometrics Methodology in Applied Psychology, 21,* 267–284.

Jentsch, E. (1997). On the psychology of the uncanny (1906). *Angelaki: Journal of the Theoretical Humanities, 2*(1), 7–16.

Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., . . . Narr, R. K. (2016). The effects of disgust on moral judgments: Testing moderators. *Social Psychological and Personality Science, 7,* 640-647. doi:10.1177/1948550616654211

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16-32. doi:10.1037/0022-3514.91.1.16

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00390

Laakasuo, M., Sundvall, J., & Drosinou, M. (2017). Individual differences in moral disgust do not predict utilitarian judgments, sexual and pathogen disgust do. *Scientific Reports, 7,* 45526. doi:10.1038/srep45526

Landy J. F., & Goodwin G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science, 10,* 518–536.

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146,* 190–205.

MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 16,* 141–172.

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25,* 695–710.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 7,* 297–337.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314–324. doi:10.3758/s13428-011-0168-7

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146,* 22–32.

McAndrew, F. T., & Koehnke, S. S. (2016). On the nature of creepiness. *New Ideas in Psychology, 43,* 10–15.

Mitchell, W. J., Szerszen, K. A., Sr., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & Macdorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2,* 10–12.

Mori, M. (1970/2005). The uncanny valley. (K. F. MacDorman, & T. Minato, Trans.). *Energy, 7,* 33–35.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior, 26*(1), 88–105.

Olivera-La Rosa, A. (2018). Wrong outside, wrong inside: A social functionalist approach to the uncanny feeling. *New Ideas in Psychology, 50,* 38–47.

Olivera-La Rosa, A., Villacampa. J., Amador, O., Corradi, G., Munar, E., & Rosselló, J. (2017). Implicit attitudes toward violence in a sample of adolescent offenders with conduct disorder. *Journal of Interpersonal Violence*. doi:10.1177/0886260517739287

Richetin, J., Costantini, G., Perugini, M., & Schönbrodt, F. (2015). Should we stop looking for a better scoring algorithm for handling Implicit Association Test data? Test of the role of errors, extreme latencies treatment, scoring formula, and practice trials on reliability and validity. *PLoS ONE, 10*(6), e0129601. doi:10.1371/journal.pone.0129601

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions (*3rd ed., pp. 757–776). New York: Guilford Press.

Schnall, S. (2016). Disgust as embodied loss aversion. *European Review of Social Psychology, 28,* 50–94.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2015). Landy and Goodwin confirmed most of our findings then drew the wrong conclusions. *Perspectives on Psychological Science, 10,* 537–538

Schwarz, N. & Clore, G. L. (2007). Feelings and phenomenal experiences. In A. Kruglanski & E. T. Higgins (Eds.), *Social psychology. Handbook of basic principles* (2nd ed., pp. 385–407). New York: Guilford Press.

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence, 16,* 337–351. doi:10.1162/pres.16.4.337

Simpson, E. A., Varga, K., Frick, J. E., & Fragaszy, D. (2011). Infants experience perceptual narrowing for nonprimate faces. *Infancy, 16*(3), 318–328.

Smith, P. (2014). Of 'near pollution'and non-linear cultural effects: Reflections on Masahiro Mori and the Uncanny Valley. *American Journal of Cultural Sociology, 2*(3), 329–347.

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science,* 26, 39–47.

Tello, N., Harika-Germaneau, G., Serra, W., Jaafari, N., & Chatard, A. (2018). Does the Implicit Association Test predict suicidal behavior? Direct replication of Nock, Park, Finn, Deliberto, Dour, & Banaji (2010). Retrieved from https://psyarxiv.com/esgf5/

Tinwell, A., Nabi, D. A., & Charlton, J. P. (2013). Perception of psychopathy and the uncanny valley in virtual characters. *Computers in Human Behavior, 29*(4), 1617–1625.

Todorov, A., Mende-Siedlecki, P. M., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology, 23,* 373–380. doi:10.1016/j.conb.2012.12 .010

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12,* 455–460. doi:10.1016/j.tics.2008 .10.001

Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: Individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology, 97,* 103–122.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*(1), 72–81. doi:10.1177/1745691614556679

van Dijke, M., van Houwelingen, G., De Cremer, D., & De Schutter, L. (2018). So gross and yet so far away: psychological distance moderates the effect of disgust on moral judgment. *Social Psychological and Personality Science, 9,* 689–701.

Villacampa, J., Ingram, G., & Martí-Vilar, M., & Olivera-La Rosa, A. (2018). An investigation of Facebook users' implicit associations between Facebook, sexual and prosocial behavior. *Heliyon, 4,* e00811. doi:10.1016/j.heliyon.2018.e00811

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393–407.

Wang, S., & Rochat, P. (2017). Human perception of animacy in light of the uncanny valley phenomenon. *Perception, 46*(12), 1386–1411.

Watt, M. C., Maitland, R. A., & Gallagher, C. E. (2017). A case of the "heeby jeebies": An examination of intuitive judgements of "creepiness". *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *49*(1), 58–69.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598.

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research, 55*, 20–32.

Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology, 6.* doi:10.3389/fpsyg.2015.00883

Received: 2.6.2019
Revised: 5.13.2019
Accepted: 5.14.2019