



## Testing the ‘uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli



Jari Kätsyri<sup>a,b,\*</sup>, Meeri Mäkäräinen<sup>b</sup>, Tapio Takala<sup>b</sup>

<sup>a</sup> Brain and Emotion Laboratory, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands

<sup>b</sup> Department of Computer Science, School of Science, Aalto University, Finland

### ARTICLE INFO

**Keywords:**

Uncanny valley  
Anthropomorphism  
Human-computer interaction  
Computer animation  
Animation films

### ABSTRACT

The uncanny valley (UV) hypothesis, which predicts that almost but not fully humanlike artificial characters elicit negative evaluations, has become increasingly influential. At the same time, the hypothesis has become associated with many computer-animated films that have aimed at high realism. In the present investigation, we tested whether semirealistic animated film characters do in fact elicit negative evaluations. Fifty-four participants were asked to evaluate five matched film excerpts from each of cartoonish, semirealistic, and human-acted films. Mixed model analyses were conducted to reduce the effects of participant and stimulus related confounds. Explicit selections made after the experiment confirmed that participants associated semirealistic film characters correctly with the UV. Semirealistic animated characters also received higher eeriness ratings than the other film characters. In particular, two semirealistic films '*Beowulf*' and '*The Polar Express*' were selected the most often explicitly, and '*Beowulf*' also received higher eeriness ratings than any other film. Somewhat unexpectedly, cartoonish characters received the highest strangeness ratings and (after confound correction) the lowest likability ratings. Taken together, the present findings demonstrate that semirealistic animated film characters are more eerie than cartoonish characters or real actors, and hence provide evidence for the existence of the UV in animated film characters.

### 1. Introduction

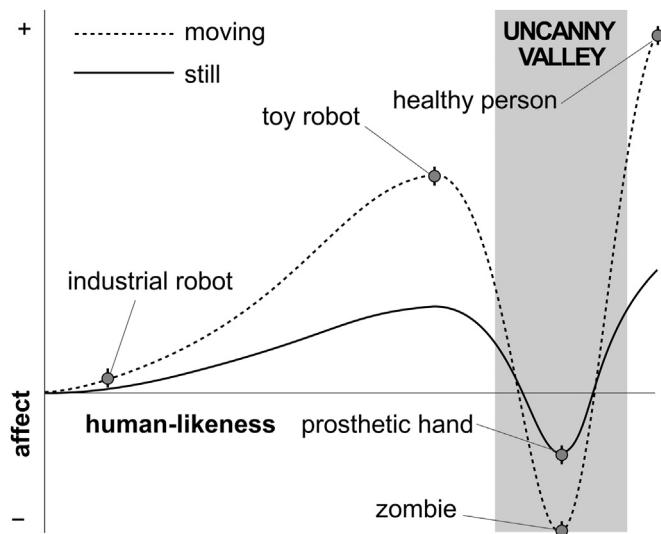
Masahiro Mori, a Japanese robotics professor, predicted already in the 1970s that although increasingly humanlike robots would elicit positive affects, robots and other artificial devices that reached a threshold of being almost but not fully humanlike could elicit a profound sense of eeriness (Mori, 1970/2012) (Fig. 1). Based on the shape of this hypothetical evaluative curve, Mori coined his hypothesis as the uncanny valley (UV). The UV hypothesis has been rediscovered during the ongoing millennium (Brenton et al., 2005; Gee et al., 2005; Hanson, 2005; MacDorman, 2005), and it is at the present particularly relevant for computer graphics and animation technologies, which can arguably be used to produce the most realistic humanlike characters of today (e.g., Alexander et al., 2010). Although realistic computer-generated faces and characters are not necessarily interactive; realistic, emotionally expressive, and virtually interactive animated characters can already be found in the cinema. In the present empirical study, we investigate whether semirealistic animated film characters show evidence of the UV hypothesis..

The UV hypothesis would predict that some animated film char-

acters that are intended to appear realistic elicit negative affective reactions in viewers. Consistently, computer-animated films using state-of-the-art animation techniques, such as *Final Fantasy: The Spirits Within* (Aida et al., 2001), *The Polar Express* (Goetzman et al., 2004), and *Beowulf* (Rapke et al., 2007), have aroused critical reviews in the media. For example, the critics have noted that the characters of *Final Fantasy* “look so real that it's creepy” (Kempley, 2001), that “watching the humans in *The Polar Express* is like watching people through a smoky car windscreen” (Savlov, 2004), and that “motion capture [the animation technique used] in *Beowulf* comes across as an unsatisfying compromise between animation and live action” (Anslen, 2007). These and other similar films have been explicitly considered in the UV context in later film reviews (e.g., Gleiberman, 2011; Hill, 2011; Phillips, 2011; Robinson, 2007; Stevens, 2011) and technologically oriented magazine articles (e.g., Plantec, 2007; Perry, 2014; Weschler, 2002). Such observations from film and technology experts provide anecdotal evidence for the existence of the UV in computer-animated films. Although anecdotal, this association has been repeatedly mentioned in empirical research as well (e.g., Bartneck et al., 2009; Brenton et al., 2005; Burleigh et al., 2013;

\* Correspondence to: Oxfordlaan 55, 6229 ER Maastricht, The Netherlands.

E-mail address: [jari.katsyri@maastrichtuniversity.nl](mailto:jari.katsyri@maastrichtuniversity.nl) (J. Kätsyri).



**Fig. 1.** The characteristic uncanny valley curve between affective evaluations and human-likeness, as predicted by Mori (1970/2012). Some of Mori's original examples have been highlighted on the curve for moving characters.

Chaminade et al., 2007; Kaba, 2013; Looser and Wheatley, 2010; MacDorman et al., 2009; McDonnell et al., 2012; Misselhorn, 2009; Piwek et al., 2014; Pollick, 2010; Saygin et al., 2012; Steckenfänger and Ghazanfar, 2009; Tinwell et al., 2011; Tondu, 2012), which indicates considerable academic interest in such possibility.

Demonstrating that semirealistic animated film characters do elicit negative affective reactions in viewers would strengthen the UV hypothesis, which has to date received inconsistent empirical evidence (for recent reviews, see Kätsyri et al., 2015; Pollick, 2010; Wang et al., 2015). This inconsistency may originate from the lack of consensus on the conceptual and operational definitions of the UV hypothesis – in fact, a characteristic of the original UV formulation is that it is a “broadly applicable guidepost to designers in a variety of domains” (Pollick, 2010, pp. 70–71) rather than a precisely defined experimental hypothesis. We will first consider evidence from studies that have used strictly controlled stimulus continua ranging from fully artificial to fully realistic, such as those generated by image morphing (e.g., Cheetham et al., 2011; MacDorman, 2006; Yamada et al., 2013), computer-generated imagery (CGI) (e.g., Burleigh et al., 2013; MacDorman et al., 2009), and motion manipulation methods (e.g., Piwek et al., 2014; Thompson et al., 2011). Although the earliest image morphing studies provided evidence in favour of the UV hypothesis (Hanson, 2006; MacDorman and Ishiguro, 2006), these findings could also be explained by uncontrolled image morphing artifacts (cf. MacDorman et al., 2009). The majority of recent studies have demonstrated that, contrary to the UV hypothesis, increasing human-likeness elicits increasingly positive evaluations (e.g., Experiment 1 in Burleigh et al., 2013; Cheetham et al., 2014; Looser and Wheatley, 2010; MacDorman et al., 2009; Piwek et al., 2014; Seyama and Nagayama, 2007; Thompson et al., 2011). However, a minority of studies have demonstrated nonlinear changes that are consistent with the UV hypothesis (Experiment 2 in Burleigh et al., 2013; Ferrey et al., 2015; Yamada et al., 2013).

The inconsistency of the above findings could possibly originate from the fact that the UV would manifest itself only under very specific experimental conditions. A careful reading of Mori's original article (Mori, 1970/2012) reveals that he did not explicitly state that all kinds of possible human-likeness manipulations would lead to the UV. One possibility is that the UV is caused by a perceptual mismatch between artificial and realistic features. This suggestion is consistent with Mori's illustrative examples, such as a myoelectric hand that looks but does not feel human, and it has also received support from empirical studies

(for a review, see Kätsyri et al., 2015). For example, Seyama and Nagayama (2007) showed that a greater mismatch between the realism of the eyes and the rest of the face elicits more negative evaluations, with the most negative evaluations occurring for fully artificial eyes placed on a fully realistic face or vice versa. The authors also demonstrated that unrealistically large eyes appeared the most eerie on the most realistic faces. MacDorman et al. (2009) demonstrated similar eyes–face mismatch and eye enlargement effects for CGI faces. Mäkäräinen et al. (2014) found that exaggerated facial expressions are acceptable on cartoonish faces but appear increasingly strange on increasingly humanlike faces. Recently, MacDorman and Chattopadhyay (2016) demonstrated that inconsistency between computer animated and real features causes humans and animals, but not objects, to appear eerier and colder. Other studies have demonstrated that individuals show increasing consensus when judging the range of aesthetic facial proportions on increasingly realistic faces (Green et al., 2008; MacDorman et al., 2009).

Although the above findings could be taken to imply that the UV exists and can be caused by either a perceptual mismatch between realistic and artificial features or a heightened sensitivity to deviations from human norms in highly realistic characters, this suggestion is not without problems. First, one should be careful in generalizing results from these relatively few experimental manipulations to all possible kinds of perceptual mismatches. Second, the above explanations cannot exclude the possibility that the UV could also be caused by yet some other explanatory mechanisms. Third, it remains uncertain whether the above experimental results can be generalized to natural stimuli. Rigorously controlled experimental UV studies are by necessity tied to narrow stimulus manipulations; for example, the above studies have focused predominantly on facial feature modifications. Testing whether the UV is caused by a specific stimulus manipulation out of various imaginable possibilities could be said to represent a “bottom up” approach for testing the UV hypothesis. Given that the UV still remains poorly understood, the risk is that the adopted stimulus manipulations are not fully relevant for the phenomenon.

An alternative “top down” approach would be to first test whether the UV phenomenon exists for natural stimuli and then investigate which specific features have caused it. Two recent studies have already provided positive evidence for the UV in images of human, prosthetic, and robot hands (Poliakoff et al., 2013) and images of real-world robot faces (Mathur and Reichling, 2016). Other studies have already provided tentative evidence for the existence of the UV in video game characters. McDonnell et al. (2012) demonstrated that one of their most realistic rendering styles for computer-generated faces elicited low appeal. Schneider and Yang (2007) showed that almost but not fully human video game characters tend to receive low attractiveness ratings. Tinwell et al. (2010) showed that two of their studied video game characters that were not intended to appear eerie nevertheless received lower familiarity ratings than other similar characters. Furthermore, they also reported a negative correlation between audio-visual asynchrony and familiarity. Flach et al. (2012) studied the UV using animated film characters; however, they also included materials from various other sources, and their results were based only on visual inspection of data. Ho and MacDorman (2010) included film excerpts from two semirealistic animated films as a part of their questionnaire development; however, these films were not explicitly compared to other stimuli. To the best of our knowledge, no studies have yet studied the existence of the UV exclusively in animated film characters by comparing such characters to matched cartoonish and human stimuli.

To summarize, animated film characters could be used to test the validity of the UV hypothesis without making strict *a priori* assumptions about which specific features, mechanisms, or explanations cause the phenomenon. This is important because at the present the evidence for the UV appears to vary depending on the specific methods and assumptions (e.g., perceptual mismatch) adopted in each particular study. Although animated films are by no means the only possible

naturalistic stimuli that can be used for testing the UV, they are particularly interesting because their anecdotal connection with the UV has been frequently noted in academia. In the present study, we test whether evaluations of animated film characters do show evidence of the UV. Specifically, we intend to compare cinematic materials extracted from semirealistic animated films to those of matched cartoonish animated films and human-acted films. Our aim is to include a comprehensive set of commercial motion-capture animated film characters into the semirealistic animation category. Motion-capture animation techniques are used to track a real actor's artistic performance and to replicate it on a computer-generated character. We focus specifically on motion-capture animation following the observation that the typically mentioned anecdotal examples of the UV – *Final Fantasy* (Aida et al., 2001), *The Polar Express* (Goetzman et al., 2004), and *Beowulf* (Rapke et al., 2007) in particular (see above for representative citations in academia) – have applied this technique. With cartoonish animations, we refer to conventional computer animations that follow the common animation principle of exaggeration to produce simplified but appealing characters (Lasseter, 1987).

A challenge of using animated film characters as research stimuli is that complete control over all possible confounding factors influencing their evaluations is not possible. Difficulties associated with the heterogeneity of research stimuli in studies with natural stimuli were noted already in the first empirical UV studies (MacDorman, 2006). Some confounds should be expected for film stimuli in particular. Participants' expertise with animation technologies, previous familiarity with the UV phenomenon, and familiarity with the animated film genre are likely to influence their evaluations. Even more critically, confound factors are also likely to vary at the level of individual stimuli. For example, participants' previous familiarity with and liking of specific films are likely to influence how they evaluate individual films. Furthermore, the evaluations of specific film characters may be influenced by how appealing or appalling they have been intended to appear. Although absolute experimental control of such confound factors is not possible, their effects could be mitigated by including appropriate confound variables in the analyses. Ideally, confound variables that depend on both films and participants (e.g., whether a specific film has been seen by a specific participant) should be collected from the same participants who do the evaluations.

A necessary condition for demonstrating the existence of the UV in film characters would be that, after participants have been explained the UV hypothesis, they will associate it correctly with the semirealistic animated characters. This prediction can be tested by asking participants to explicitly indicate which characters in their opinion appear eerie in the sense of the UV, after they have been explained the UV hypothesis in detail. Such explicit evaluations can be collected after participants have first rated their immediate impressions of the film characters, so that these immediate evaluations are not confounded by the explicit explanation for the UV hypothesis. This leads to our first hypothesis.

**H1.** After participants have been explained the UV hypothesis, they will explicitly select semirealistic animated film characters as eerie more often than cartoonish animated film characters or human actors.

Empirical studies have adopted specific self-report items from Mori's (1970/2012) original Japanese terms *bukimi* and *shin-wakan* to study the affective experience of the UV. *Bukimi* translates quite unequivocally as eeriness; however, *shin-wakan* does not have an undisputable translation and it has been translated varyingly as familiarity, affinity, warmth, and likability (Bartneck et al., 2009; Ho and MacDorman, 2010). Familiarity would be a particularly problematic item for animated films, given that it is trivially confounded with participants' previous exposure with specific films and film characters. Most studies have adopted either eeriness (Hanson, 2006; MacDorman et al., 2009; MacDorman and Ishiguro, 2006; Thompson et al., 2011), likability or its synonyms (Ferrey et al., 2015; Looser and Wheatley,

2010; Seyama and Nagayama, 2007), or both eeriness and likability (Burleigh et al., 2013; McDonnell et al., 2012) to study the affective dimension of the UV. Following these previous conventions, we have adopted likability and eeriness to tap into the positive and negative reactions elicited by UV, respectively.

Based on the previous research, we made two opposing predictions. If greater human-likeness elicits greater positive affect, the positivity of evaluations should increase across cartoonish, semirealistic, and human characters. However, if the UV hypothesis holds true, semirealistic animated characters should elicit more negative evaluations than cartoonish and human characters. These alternative hypotheses can be formulated as follows.

**H2a.** Cartoonish animated film characters will be evaluated as more eerie and less likable than semirealistic animated film characters, which will be evaluated as more eerie and less likable than human actors.

**H2b.** Semirealistic animated film characters will be evaluated as more eerie and less likable than cartoonish animated film characters and human actors.

One tempting explanation for the UV is that stimuli falling between clearly artificial and clearly human categories elicit eeriness because they are difficult to categorize. Cheetham et al. (2011) were the first to demonstrate that image morphs between artificial and human faces are indeed perceived categorically. Formally, categorical perception requires demonstrating enhanced perceptual discrimination performance for stimuli that straddle a category boundary than for equally spaced stimuli that reside on the same side of the boundary (Goldstone and Hendrickson, 2010). Studies fulfilling this strict criterion have not yet demonstrated that categorically ambiguous stimuli would elicit negative evaluations (for negative results, see Cheetham et al., 2014; Looser and Wheatley, 2010). However, other studies using bistable images (Ferrey et al., 2015) and images of faces (Burleigh et al., 2013; Ferrey et al., 2015; Yamada et al., 2013) have demonstrated that some intermediate stimuli located between artificial and realistic categories do elicit negative evaluations (however, see also MacDorman and Chattopadhyay, 2016). In the present context, it is possible that semirealistic film characters would appear eerie because they are difficult to categorize as clearly artificial or clearly human. Although it is obviously not possible to test perceptual discrimination performance formally with natural film excerpts, we nevertheless expected that participants' subjective classification difficulty evaluations would show evidence of heightened difficulty for semirealistic characters. This led to our third hypothesis:

**H3.** Semirealistic animated film characters will be subjectively more difficult to categorize as human or nonhuman than cartoonish animated film characters or human actors.

## 2. Methods

### 2.1. Participants

54 participants (30 women) with a mean age of 23.4 years ( $SD=4.9$ ) were recruited for the study. A majority of the participants were undergraduate (88%) or postgraduate (4%) students in Finnish universities or universities of applied sciences; the remaining participants (7%) had either a university degree or a higher vocational diploma and were employed in various occupations. All participants were native or fluent Finnish speakers and reported normal hearing and eyesight. Participants received two movie tickets for their participation. The present study was approved by the Aalto University Research Ethics committee, and it adhered to the tenets of the World Medical Association Declaration of Helsinki and the ethical principles established by the Finnish Advisory Board on Research Integrity (<http://www.tenk.fi/en/>).

All participants reported viewing full-length feature films on a regular basis. The majority of participants reported viewing feature films regularly on a monthly (50%) or weekly (39%) basis; the remaining participants viewed feature films less frequently than once per month (11%). In an evaluation of the three most liked and the three most disliked film genres, participants selected adventure (43%), comedy (37%), and drama (37%) as the most commonly liked ones; and horror (44%), family (44%), and romance (33%) as the most commonly disliked ones. Animation films were selected infrequently as either one of the most liked (15%) or one of the most disliked (7%) genres. About one half of the participants reported viewing animated feature films or television series regularly on a monthly (33%), weekly (20%), or daily (4%) basis; the remaining participants (41%) viewed animations less frequently than once per month; and one participant (2%) reported never viewing animations.

## 2.2. Design

The experiment used a  $3 \times 5 \times 4 \times 3$  mixed design with film type (cartoonish animation, semirealistic animation, and real human actors), film (five films; nested within actor type), and film clip (four film clips; nested within film) as within-subjects factors and experimental group (three groups) as a between-subjects factor.

## 2.3. Film stimuli

Five films were included for each of the cartoonish animation (films C1 to C5), semirealistic animation (R1 to R5), and real human (H1 to H5) film type categories. Four short film clips ( $M=40$  s, range 30–56 s) were extracted from each individual film. Selected films are listed in Table 1, and full details on the selected films and film clips are available in Tables S1 and S2, respectively. All films were produced in North American film studios and were spoken in English. To create as naturalistic viewing conditions as possible and to avoid the extra cognitive burden of having to follow speech dialogue in a foreign language, Finnish subtitles were displayed on the bottom of the screen, which is the predominant norm in the Finnish cinema and television.

For the semirealistic animation films, our inclusion criteria were that the films should be fully computer-animated, implemented with motion-capture animation techniques, and intentionally aimed at high levels of human-likeness. Our decision to adopt motion-capture animation as an inclusion criteria was based on the observation that most animation films such as the *The Polar Express* (Goetzman et al., 2004) that have received common criticism in the UV context have used such techniques. We accepted only motion-capture animated films into the study to reduce variation in the animation techniques. A comprehensive list of all (seven) motion-capture animation films published since the year 2000 was first extracted from an Internet database (Box Office Mojo; <http://www.boxofficemojo.com>). One of these films (*A Christmas Carol*; Rapke et al., 2009) was excluded because its main character was an obvious caricature. Cartoonish animation and human-acted films (i.e., conventional feature films) were selected by matching them with the six semirealistic animation films on the basis of specific criteria. The main criterion was that the main actors' ages and genders should match those of the semirealistic films. For the cartoonish animations, films depicting other than human characters and films implemented with other than computer animation techniques (e.g., cel animation) were excluded. In an attempt to reduce variation in participants' previous familiarity with different films, films were matched with respect to their publication year, non-domestic gross profit (i.e., box office profit outside USA), and critical evaluations from both professionals and laymen.

Selection criteria for the four film clips were that they should depict social interactions between the main character and the other film characters, they should display the main character's face clearly (i.e., from front and from a close range), they should not display nudity or

**Table 1**  
List of Selected Films.

Film type	Label	Film title (main character)	Citation
Semirealistic animation	R1	<i>Final Fantasy: The Spirits Within</i> ("Aki Ross")	Aida, Lee, Sakai, Sakaguchi, & Sakakibara (2001)
	R2	<i>The Polar Express</i> (unnamed boy)	Goetzman, Starkey, Teitler, & Zemeckis (2004)
	R3	<i>Beowulf</i> ("Beowulf")	Rapke, Starkey, & Zemeckis (2007)
	R4	<i>Mars Needs Moms</i> ("Milo")	Boyd, Rapke, Starkey, & Wells (2011)
	R5	<i>The Adventures of Tintin: The Secret of the Unicorn</i> ("Tintin")	Jackson, Kennedy, & Spielberg (2011)
Cartoonish animation	C1	<i>The Incredibles</i> ("Mr. Incredible")	Walker & Bird (2004)
	C2	<i>Meet the Robinsons</i> ("Lewis")	McKim & Anderson (2007)
	C3	<i>Cloudy with a Chance of Meatballs</i> ("Flint Lockwood")	Marsden, Lord, & Miller (2009)
	C4	<i>Arthur Christmas</i> ("Arthur Christmas")	Pegram, Smith, & Cook (2011)
	C5	<i>Epic</i> ("M.K.")	Davis, Forte, & Wedge (2013)
Human actors	H1	<i>Gladiator</i> ("Maximus")	Franzoni, Lustig, Wick, & Scott (2000)
	H2	<i>Lara Croft Tombraider: The Cradle of Life</i> ("Lara Croft")	Gordon, Levin, & de Bont (2003)
	H3	<i>Zathura: A Space Adventure</i> ("Danny")	Kroopf, de Luca, Teitler, & Favreau (2005)
	H4	<i>Bridge to Terabithia</i> ("Jesse Aarons")	Levine, Lieberman, Paterson, & Csupo (2007)
	H5	<i>Stardust</i> ("Tristan Thorne")	di Bonaventura, Dreyer, Gaiman, & Vaughn (2007)

violence, and that their events should be reasonably understandable on their own. In an attempt to control for the emotional contents of films, a further selection criterion was that two of the selected clips should elicit pleasant emotions and two of the clips should elicit unpleasant emotions in the viewers. Valence (i.e., unpleasantness-pleasantness) was adopted as the selection criterion because a more fine-grained emotion classification would have required a much larger set of films (cf. Gross and Levenson, 1995). Fourteen pre-test participants who did not take part in the main study evaluated the valence and arousal (see Section Measured variables) of the initial stimuli sequentially in three pilot sessions (4–5 participants per session). After each session, film clips whose median valence ratings fell on the opposite side of the valence scale than expected were replaced with new clips (in total, 8 out of 64 clips).

Thirteen additional participants who did not take part in the main study participated in a pilot evaluation. The human-likeness of film characters was evaluated using a similar index as in the main study (see Measured variables), except that evaluations were given on a 9-step rather than a 7-step scale. The results confirmed that human actors were more humanlike than the semirealistic animated characters, which were more humanlike than the cartoonish animated characters ( $M_s = 8.87, 4.74$ , and  $2.39$ ), LMM analysis (see Section Analyses):  $F(2, 17)=922.85$ ,  $p < 0.001$ . One of the allegedly semirealistic animated films (*Monster House*; Rapke et al., 2006) was dropped from the final selection because it received equally low human-likeness ratings ( $M=2.39$ ) as the cartoonish characters, and hence did not fulfil the explicit inclusion criterion of high human-likeness. Correspondingly, one cartoonish animation (*Jimmy Neutron*; Hecht et al., 2001) and one

**Table 2**

Median (and Range) Statistics for Film Types.

Measure	Cartoonish animations	Semirealistic animations	Human actors	$\chi^2$ (2)	p
Publication year	2008 (9)	2007 (10)	2005 (7)	3.23	0.199
Gross profit (non- domestic) <sup>a</sup>	118.1 (298.4)	114.1 (178.8)	90.8 (234.9)	1.68	0.432
Critical evaluation <sup>b</sup>					
Critics	66 (38)	59 (19)	66 (31)	2.10	0.351
Individuals	7.4 (2.2)	6.5 (2.2)	7.2 (3.4)	1.15	0.563

Note. Statistics are from Kruskall-Wallis H tests.

<sup>a</sup> Extracted from Box Office Mojo database (<http://boxofficemojo.com>). Unit is millions of dollars.

<sup>b</sup> Extracted from Metacritic database (<http://www.metacritic.com>). Critics' evaluations refer to the "Metascore" metric (range 0–100) and individuals' evaluations refer to average ratings from the Metacritic users (range 0–10).

human-acted film (Hugo; Depp et al., 2011) were dropped from the other categories, leaving five films for each film type. Statistics for these final films are shown in Table 2. Non-parametric median tests confirmed that publication years, non-domestic gross profits, and critical evaluations did not differ significantly between film types.

We used official DVD releases for all films. All DVDs used PAL format and were anamorphically encoded for widescreen displays. The films' aspect ratios varied between 1.78 and 2.40 depending on the film (Table S1).

#### 2.4. Procedure

Participants arrived to the experiment in three groups ( $N_s = 14, 19,$  and 21) such that all members in a specific group viewed and evaluated films at the same time. Experiments took place in a square 95 m<sup>2</sup> auditorium designed specifically for audio-visual presentations. The auditorium had black-painted walls and sound insulation panels installed on the walls and ceiling. Lights were switched off during the experiment, with the exception of dim lights at the rear wall that were used to provide sufficient lighting for responding to the questionnaires using pen and paper. Film media were projected on a wide 8.5 m × 5.5 m silver screen using Eiki LC-SX4 L video projector at its native horizontal ratio of 1280 pixels and a vertical resolution that depended on each film's aspect ratio (e.g., 1280 × 720 pixels for a film with 1.78:1 aspect ratio). VLC Player software (<http://www.videolan.org/vlc/>) running on a desktop PC was used for handling all stimulus display. Participants were positioned in several rows located approximately 5–9 m from the screen. Film sound tracks were played back using a 5.1 multichannel audio system at a loud but comfortable sound volume level.

After arrival, participants were welcomed to the experiment, given written instructions, and asked to fill a background questionnaire and an informed consent form. The instructions were repeated verbally, after which participants completed two practice trials for two excerpts from a cartoonish animated film, *Brave* (Sarafian et al., 2012), and a conventional film using photorealistic CGI effects, *Tron: Legacy* (Bailey et al., 2010). These stimuli were not included in the main study. After the practice, participants were encouraged to ask clarifications for any unclear questions. Films were presented in one of three pseudo-randomized orders depending on the experimental group, with the restriction that no more than two films of the same type could be presented in succession. Each film trial began with the presentation of a blank screen (random duration between 1–3 s), presentation of the film title (3 s), and a short description of the film ( $M=19$  words, range 16–22 words; duration 15 s). After this, the four film clips were presented one after another. Clips were presented in the same order as they occurred in the original film, except in one case in which

modifying this order made the narrative easier to understand (film H4; see Table S2). A blank screen was presented at the beginning of each clip (1–3 s). After the video clip, blank screen was presented again (for 1–3 s), after which a verbal request for evaluating the emotional contents of the clip was shown (20 s). After all clips had been evaluated, a new verbal request for evaluating the whole film was presented (75 s). Brief flashes from white to black (5 s) were displayed at the end of all verbal requests to draw the participants' attention back to the screen.

After evaluating all films, participants were given a verbal debriefing of the UV phenomenon and the purpose of the study, and asked to fill a post-experimental evaluation questionnaire. Specifically, participants were asked to indicate whether they had been familiar with the UV hypothesis before this study (yes/no choice). Additionally, participants were presented thumbnail pictures of film characters, and were asked to select those characters which in their opinion were the most eerie in the sense of the UV phenomenon (or to skip this task if none of them were). Participants were also able to provide open feedback on the film viewing or experimental arrangements in writing. After post-evaluation, participants were debriefed and thanked for their participation. In total, the experiment lasted approximately one-and-a-half hours.

#### 2.5. Measured variables

##### 2.5.1. Dependent variables

Dependent variables consisted of explicit eeriness selection, human-likeness, likability, strangeness, eeriness, and subjective classification difficulty. Explicit eeriness selection was adopted from the post-experimental questionnaire and coded dichotomously (0=not selected, 1=selected). Other evaluations were rated on a 7-step Likert scale ranging from total disagreement to total agreement during the experiment. Human-likeness consisted of three items (Cronbach's  $\alpha=0.89$ ): the extent to which the character appeared genuinely human, cartoonish, and exaggerated (the two last items were reverse-coded). For likability, three items were evaluated ( $\alpha=0.90$ ): the extent to which the character appeared aesthetic and pleasant, and the extent to which the participant liked the character's appearance. Eeriness was evaluated using three items: the extent to which the character appeared strange, unsettling, and eerie; however, because of poor reliability ( $\alpha=0.64$ ), strangeness was dropped from this scale. After this change, the scale had an acceptable reliability ( $\alpha=0.70$  and  $R=0.58$ ). Classification difficulty referred to the subjective assessment of the extent to which the character was "difficult to classify as being computer-animated or human". Human-likeness, likability, and eeriness items were averaged to form aggregate indexes. Strangeness was included as a separate dependent variable.

**Table 3**  
Tested Confound Variables.

Variable <sup>a</sup>	Description	Levels <sup>b</sup>
PartGender	Gender	2
PartAnimExpert	Animation expertise	C
PartAnimFreq	Animation viewing frequency	C
PartKnewUV	Previous familiarity with UV	2
FilmSeen	Film seen previously	2
FilmLikability	Film likability	C
FilmValence	Film valence (mean across clips)	C
FilmArousal	Film arousal (mean across clips)	C
FilmConfusingness	Confusingness of film events	C
CharFamiliarity	Familiarity with character	C
CharIntentLikability	Character's intentional likability	C
CharIntentEeriness	Character's intentional eeriness	C

Note. UV—Uncanny valley.

<sup>a</sup> Variables with prefix "Part" varied only across participants, whereas variables with prefixes "Film" and "Char" also varied across films and characters.

<sup>b</sup> Number of levels; "C" if continuous.

**Table 4**

Estimated Marginal Means for Film and Character Confound Variables by Film Type.

Variable	Cartoonish	Semirealistic	Human	df <sup>a</sup>	Value <sup>b</sup>	p
FilmSeen	18 <sub>a</sub> (6%)	11 <sub>b</sub> (6%)	24 <sub>a</sub> (25%)	2, 54	21.26	< 0.001
FilmLiking	4.55 <sub>a</sub> (0.97)	4.11 <sub>b</sub> (1.08)	4.77 <sub>c</sub> (0.93)	2, 50.2	14.58	< 0.001
FilmValence	5.28 <sub>a</sub> (0.58)	4.86 <sub>b</sub> (0.64)	5.03 <sub>a</sub> (0.63)	2, 57.1	21.75	< 0.001
FilmArousal	3.50 <sub>a</sub> (1.11)	3.61 <sub>a</sub> (1.02)	3.91 <sub>b</sub> (1.16)	2, 37.4	17.05	< 0.001
FilmConfusingness	3.02 <sub>a</sub> (1.27)	3.21 <sub>a</sub> (1.20)	3.06 <sub>a</sub> (1.02)	2, 26.4	1.38	0.269
CharFamiliarity <sup>c</sup>	2.26 <sub>a</sub> (0.84)	2.77 <sub>b</sub> (0.82)	3.98 <sub>c</sub> (0.97)	2, 225.4	48.91	< 0.001
CharIntentLikability	5.50 <sub>a</sub> (0.84)	5.48 <sub>a</sub> (0.92)	5.83 <sub>b</sub> (0.83)	2, 22.9	8.04	0.002
CharIntentEeriness	1.68 <sub>a</sub> (0.75)	1.40 <sub>b</sub> (0.68)	1.38 <sub>b</sub> (0.77)	2, 198.9	7.09	0.001

Note. Standard deviations are in parentheses. Means in each row sharing a common subscript are not statistically different at  $\alpha=0.05$ . Statistics are from Generalized Estimating Equations (GEE) analysis for the categorical FilmSeen variable and Linear Mixed Model (LMM) for all other variables (see Analyses). Film valence and arousal were recorded on 9-step pictorial scales from unpleasantness to pleasantness and low to high visceral arousal, and other non-categorical variables on a 7-step scale ranging from total disagreement to total agreement.

<sup>a</sup> Degrees of freedom and sample size are shown for GEE analysis. Degrees of freedom for LMM analyses are based on Welch-Satterthwaite approximation.

<sup>b</sup> Value is  $\chi^2$  statistic for GEE analysis and F-statistic for LMM analyses.

<sup>c</sup> Error covariance matrix for random variable film type was specified as identity (ID) matrix in the LMM analysis.

### 2.5.2. Confound variables

Table 3 shows tested confound variables. Mean values for confound variables by film type are available in Table 4. Confound variables for participants included gender, animation expertise, the frequency of viewing animated films or animated television series, and previous familiarity with the UV hypothesis. Participants were asked to evaluate their animation expertise by rating the extent to which they considered themselves computer animation experts, had studied computer animation, and had worked with computer animations (three items;  $\alpha=0.79$ ). The ratings were made on a scale ranging from 1 (none at all) to 5 (very much). Animation viewing frequency was evaluated on a scale ranging from 1 (never) to 5 (daily). Previous familiarity with the UV was adopted from the post-evaluation questionnaire and coded dichotomously (0=not familiar, 1=familiar).

Confound variables for the films included likability, emotional contents in terms of valence and arousal, film familiarity, and film confusingness. Participants evaluated their liking of each film on a 7-step scale ranging from total disagreement to total agreement. Valence and arousal were evaluated separately for each film clip using 9-step pictorial scales similar to the Self-Assessment Manikin (SAM) (Bradley and Lang, 1994), and averaged across the film clips for each film. The valence scale ranged from unpleasantness to pleasantness, and the arousal scale ranged from low to high visceral arousal. Participants were explicitly instructed to evaluate their own authentic reactions to film events rather than emotions expressed by the film actors or emotions intentionally conveyed by the film producers. Film familiarity was coded dichotomously (0=participant had not seen and 1=participant had seen the film previously; missing values were replaced with zeros). Confusingness of film events was evaluated on a 7-step scale from total disagreement to total agreement.

Confound variables for film characters included familiarity, intentional likability, and intentional eeriness. Specifically, participants evaluated on a 7-step scale ranging from total disagreement to total agreement their previous familiarity with each character and the extent to which they thought the character had been intended to appear “likable” or “strange, eerie, and/or unsettling” by the film producers.

### 2.6. Analyses

Because the present data violated the homoscedasticity and sphericity assumptions of ANOVA (i.e., both film types and films had heterogeneous variances and covariances), the data were analyzed using Linear Mixed Model (LMM) procedure, which can be considered as a generalization of ANOVA analysis (for tutorials, see Hayes, 2006; Hoffman and Rovine, 2007; Quené and van den Berg, 2004). Importantly, the present approach also made it possible to include

continuous confound variables that varied both at the level of participants and individual stimuli (Hoffman and Rovine, 2007). We adopted LMM procedure with restricted maximum-likelihood estimation in SPSS (version 23). Fixed effects included film type, individual film (nested within film type), experimental group, and interaction between film type and group. Film type was included as a random variable with unspecified covariance matrix (UN), which allowed modeling the presence of heterogeneous variances and covariances for different film types (Snijders and Bosker, 1999). Simpler covariance matrices were used when they improved the model fit over UN or in case of convergence problems. To allow heterogeneous variances across individual films, diagonal error covariance matrix was additionally specified for films across participants.<sup>1</sup> Importantly, main findings from the present analyses did not differ from those that would have been obtained using conventional ANOVA analyses (Table S3). Furthermore, conservative model fit statistics indicated that the present mixed model specifications provided better fit to the data than simpler models resembling ANOVA, and that the inclusion of confounds further improved model fit over the original models (Table S4).

Because dichotomous variables could not be modeled using LMM procedure, explicit eeriness selections were analyzed using the Generalized Estimating Equations (GEE) procedure in SPSS. This procedure is an extension of LMM that allows modeling non-normally distributed dependent variables via a specific link function between the variable and model predictors. Binomial distribution was specified for the dichotomous responses, and probit link function was selected based on the best model fit. Human actors, which were practically never selected as eerie (with only two individual exceptions) were excluded from this analysis. Compound symmetry error covariance matrix was specified for films across participants.<sup>2</sup>

<sup>1</sup> Sample SPSS syntax for the LMM analysis including one confound variable:  
MIXED likability BY filmtyp filmid group WITH filmlikability  
/CRITERIA=MXSTEP(50)  
/FIXED=filmtyp filmid(filmtyp) group group\*filmtyp filmlikability  
/METHOD=REML  
/RANDOM=filmtyp | SUBJECT(partid) COVTYPE(UN)  
/RANDOM=filmlikability | SUBJECT(partid)  
/REPEATED=filmid | SUBJECT(partid) COVTYPE(DIAG).

<sup>2</sup> Sample SPSS syntax for the GEE analysis with one confound variable:  
GENLIN selection (REFERENCE=FIRST) BY filmtyp filmid group WITH filmlikability  
/MODEL filmtyp filmid(filmtyp) group group\*filmtyp filmlikability  
DISTRIBUTION=binomial LINK=PROBIT  
/REPEATED SUBJECT=partid WITHINSUBJECT=filmid  
CORRTYPE=EXCHANGEABLE COVB=ROBUST  
/MISSING CLASSMISSING=INCLUDE.

**Table 5**  
Statistical Analyses for Dependent Variables.

Effect	df <sup>a</sup>	Value <sup>b</sup>	p
Human-likeness			
Film type	2, 51.7	500.34	< 0.001
Film	12, 74.5	17.62	< 0.001
Likability			
Film type	2, 51.8	22.96	< 0.001
Film	12, 75.1	22.39	< 0.001
Strangeness			
Film type	2, 53.8	84.10	< 0.001
Film	12, 71.7	8.89	< 0.001
Eeriness			
Film type	2, 55.6	3.49	0.037
Film	12, 82.2	7.57	< 0.001
Classification difficulty <sup>c</sup>			
Film type	2, 104.5	104.98	< 0.001
Film	12, 77.4	5.11	< 0.001
Explicit selection			
Film type	1, 54	29.32	< 0.001
Film	8, 54	41.64	< 0.001

Note. Statistics are from Generalized Estimating Equations (GEE) analysis for explicit selection variable and Linear Mixed Model (LMM) for all other variables (see Analyses).

<sup>a</sup> Degrees of freedom and sample size are shown for GEE analysis. Degrees of freedom for LMM analyses are based on Welch-Satterthwaite approximation.

<sup>b</sup> Value is  $\chi^2$  statistic for GEE analysis and F-statistic for LMM analyses.

<sup>c</sup> Error covariance matrix for random variable film type was specified as compound symmetry (CS) matrix in the LMM analysis.

**Table 6**  
Estimated Marginal Means for Dependent Variables by Film type.

Variable	Cartoonish	Semirealistic	Human
Human-likeness	2.13 <sub>a</sub> (0.69)	4.53 <sub>b</sub> (0.82)	6.69 <sub>c</sub> (0.56)
Likability	4.60 <sub>a</sub> (0.79)	4.74 <sub>a</sub> (0.87)	5.26 <sub>b</sub> (0.70)
Strangeness	3.27 <sub>a</sub> (1.15)	2.40 <sub>b</sub> (1.13)	1.49 <sub>c</sub> (0.74)
Eeriness	1.37 <sub>a</sub> (0.60)	1.49 <sub>b</sub> (0.66)	1.29 <sub>a</sub> (0.42)
Classific. difficulty	1.40 <sub>a</sub> (0.94)	2.90 <sub>b</sub> (2.01)	1.12 <sub>c</sub> (0.60)
Explicit selection	5 <sub>a</sub> % (16%)	38 <sub>b</sub> % (28%)	< 1%

Note. Standard deviations are in parentheses. Means in each row sharing a common subscript are not statistically different at  $\alpha=0.05$ . Explicit selection was recorded dichotomously, and all other variables were recorded on a 7-step scale ranging from total disagreement to total agreement.

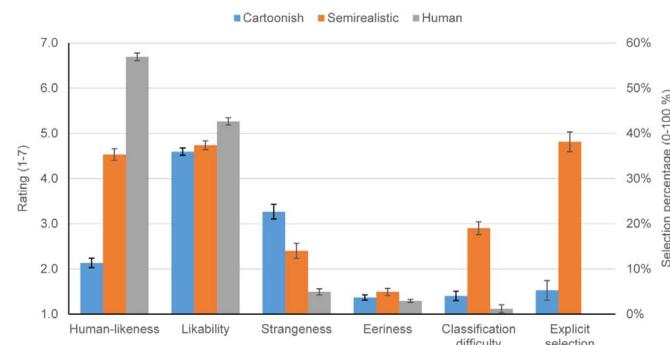


Fig. 2. Original evaluation results by film type. Error bars denote one SEM.

### 3. Results

#### 3.1. Unadjusted results

Table 5 shows statistical analysis results for film type and film effects. Mean values for dependent variables by film type are shown in Table 6 and illustrated in Fig. 2. The effect of film type was statistically significant for all dependent variables. Human actors received significantly higher human-likeness ratings than semirealistic animations, which received significantly higher human-likeness ratings than car-

toonish animations, 95% CIs for the differences [1.91, 2.41] and [2.14, 2.66]. Hence, the present film selection evoked expected changes in human-likeness..

Hypothesis H1 predicted that after being explained the UV concept, the participants would explicitly select semirealistic animations as eerie more often than the other films. In support of this hypothesis, participants selected semirealistic characters significantly more often as eerie than cartoonish animations, 95% CI for the difference [24%, 42%]. Human actors, which were excluded from this analysis, were practically never selected as eerie ( $M < 1\%$ ).

Hypotheses H2a and H2b made opposite predictions for the effects of film type on the likability and eeriness of film characters. Likability ratings failed to support neither one of these hypotheses: although semirealistic animations received slightly higher likability ratings than cartoonish animations, this difference was not statistically significant ( $p=0.191$ ), 95% CI [-0.07, 0.36]. Human actors received higher likability ratings than cartoonish and semirealistic animations, 95% CIs for the differences [0.46, 0.88] and [0.32, 0.73]. Strangeness ratings supported hypothesis H2a, which predicted decreasing strangeness across increasing human-likeness. That is, cartoonish animations evoked significantly higher strangeness ratings than semirealistic animations, which evoked significantly higher strangeness ratings than human actors, 95% CIs for the differences [0.36, 0.94] and [0.66, 1.29]. Eeriness ratings provided support for hypothesis H2b: semirealistic animations evoked significantly higher eeriness ratings than either cartoonish animations or human actors, 95% CIs for the differences [0.00, 0.25] and [0.05, 0.35].

Hypothesis H3 predicted that semirealistic animations would be more difficult to classify as human or nonhuman than the other film types. Subjective evaluations supported this hypothesis: Semirealistic animations received higher classification difficulty ratings than cartoonish animations and human actors, 95% CIs for the differences [1.24, 1.76] and [1.54, 2.03].

#### 3.2. Confound selection

Confound variables in Table 3 were tested both individually and jointly for inclusion in the adjusted analysis of each dependent variable. Main effects were tested for film and character confounds, and interaction effects with film type were tested for participant confounds (main effects were always included for significant interactions but considered unimportant by themselves). Random term across participants was always included for film and character confounds when estimable (cf. Kenny et al., 2006, p. 349).

Individual confound variable was retained if its inclusion improved the model fit for the dependent variable in question (as measured with Corrected Akaike Information Criterion [AICC] for LMM analysis and Corrected Quasi Likelihood under Independence model Criterion [QICC] for GEE analysis) and its effect was statistically significant ( $p < 0.05$ ). For brevity, we will consider here only the most consequential confounds; full list of included confound variables is available in Table S5. Likability ratings (dependent variable) were associated with intentional likability ratings, film likability ratings, and film valence ratings (confound variables), 95% CIs for the slopes [0.48, 0.64], [0.28, 0.41], and [0.30, 0.48], respectively. Similarly, eeriness ratings were associated with intentional eeriness ratings and negatively with film likability and film valence ratings, 95% CIs for the slopes [0.30, 0.46], [-0.02, -0.08], and [-0.03, -0.13]. Strangeness ratings were associated with intentional eeriness and film likability ratings, 95% CIs for the slopes [0.38, 0.66] and [-0.04, -0.16]. For classification difficulty ratings, semirealistic characters received lower ratings from participants with higher animation expertise scores, 95% CI for the slope [-0.05, -0.42], and higher ratings from female than from male participants ( $M=3.22$  and  $2.29$ ,  $SD=2.46$  and  $2.89$ ), 95% CI for the difference [0.42, 1.43]. Explicit selections showed a negative association with film likability ratings, 95% CI for the slope [-0.03, -0.22]

**Table 7**

Statistical Analyses for Dependent Variables after Adjustment for Confounds (from Joint Confound Models).

Effect	df <sup>a</sup>	Value <sup>b</sup>	p
Human-likeness <sup>c</sup>			
Film type	2, 51.7	500.34	< 0.001
Film	12, 74.5	17.62	< 0.001
Likability			
Film type	2, 46.2	13.79	< 0.001
Film	12, 88.9	11.92	< 0.001
CharIntLikability	1, 47.0 <sup>d</sup>	204.86	< 0.001
FilmLikability	1, 54.8 <sup>d</sup>	70.54	< 0.001
Strangeness			
Film type	2, 50.4	62.14	< 0.001
Film	12, 75.5	5.11	< 0.001
CharIntEeriness	1, 45.6 <sup>d</sup>	50.45	< 0.001
FilmLikability	1, 52.8 <sup>d</sup>	8.00	0.007
Eeriness			
Film type	2, 70.0	5.11	0.008
Film	12, 68.4	3.57	< 0.001
CharIntEeriness	1, 54.4 <sup>d</sup>	78.83	< 0.001
FilmLikability	1, 216.9	11.56	0.001
Classif. difficulty <sup>e</sup>			
Film type	2, 113.2	94.32	< 0.001
Film	12, 73.2	4.47	< 0.001
PartAnimExpertise	1, 51.5	0.06	0.816
PartAnimExpertise × Film type	2, 67.2	9.30	< 0.001
Gender	1, 51.5	8.72	0.005
Gender × Film type	2, 67.2	3.49	0.036
Explicit selection			
Film type	1, 54	28.14	< 0.001
Film	8, 54	33.28	< 0.001
FilmLikability	1, 54	6.23	0.013

Note. Statistics are from Generalized Estimating Equations (GEE) analysis for explicit selection variable and Linear Mixed Model (LMM) for all other variables (see Analyses).

<sup>a</sup> Degrees of freedom and sample size are shown for GEE analysis. Degrees of freedom for LMM analyses are based on Welch-Satterthwaite approximation, and depend on the included random terms.

<sup>b</sup> Value is  $\chi^2$  statistic for GEE analysis and F-statistic for LMM analyses.

<sup>c</sup> Results are identical to those of unadjusted analyses.

<sup>d</sup> The model included random term for this effect.

<sup>e</sup> Error covariance matrix for random variable film type was specified as compound symmetry (CS) matrix in the LMM analysis.

probit units.

Joint confound model specification was made loosely on the basis of Snijders and Bosker (1999); however, forward- rather than backward-selection was used for parsimony (Janssen, 2012; Nezlek, 2008). Models were built by the stepwise addition of significant individual confounds (Table S5). In each step, the considered confound was added only if the more complex model showed a better fit to the data (as measured with AICC or QICC criterion) and if its effect remained statistically significant ( $p < 0.05$ ). Film and character confounds were tested before participant confounds; otherwise, confounds were tested in the order of the best model fit. Only the better fitting variable out of intentional likability and eeriness confounds (CharIntLikability/Eeriness) was included for any dependent variable.

### 3.3. Adjusted results

As can be seen in Table 7, the effect of film type remained statistically significant for all dependent variables even after the best fitting confounds were included jointly into the statistical analyses. The pattern of significant differences between film types also remained similar for all dependent variables with the exception of likability (Table 8). Specifically, after adjustment for confounds, cartoonish animations now received significantly lower likability ratings than semirealistic animations, 95% CI for the difference [0.08, 0.42], thereby tentatively supporting hypothesis H2a.

To better understand the above change and to test the effects of confounds that were excluded from the joint models, we next tested the

**Table 8**

Estimated Marginal Means for Dependent Variables after Adjustment for Confounds (from Joint Confound Models).

Variable	Cartoonish	Semirealistic	Human
Human-likeness <sup>a</sup>	2.13 <sub>a</sub> (0.69)	4.53 <sub>b</sub> (0.82)	6.69 <sub>c</sub> (0.56)
Likability	4.58 <sub>a</sub> (0.53)	4.83 <sub>b</sub> (0.59)	5.01 <sub>c</sub> (0.51)
Strangeness	3.14 <sub>a</sub> (1.07)	2.42 <sub>b</sub> (1.09)	1.54 <sub>c</sub> (0.49)
Eeriness	1.33 <sub>a</sub> (0.40)	1.51 <sub>b</sub> (0.55)	1.31 <sub>a</sub> (0.26)
Classif. difficulty	1.39 <sub>a</sub> (0.94)	2.79 <sub>b</sub> (1.90)	1.10 <sub>c</sub> (0.56)
Explicit selection	5 <sub>a</sub> % (16%)	37 <sub>b</sub> % (29%)	- <sub>b</sub>

Note. Standard deviations are in parentheses. Means in each row sharing a common subscript are not statistically different at  $\alpha=0.05$ . Explicit selection was recorded dichotomously, and all other variables were recorded on a 7-step scale ranging from total disagreement to total agreement.

<sup>a</sup> Adjusted values are identical to unadjusted values.

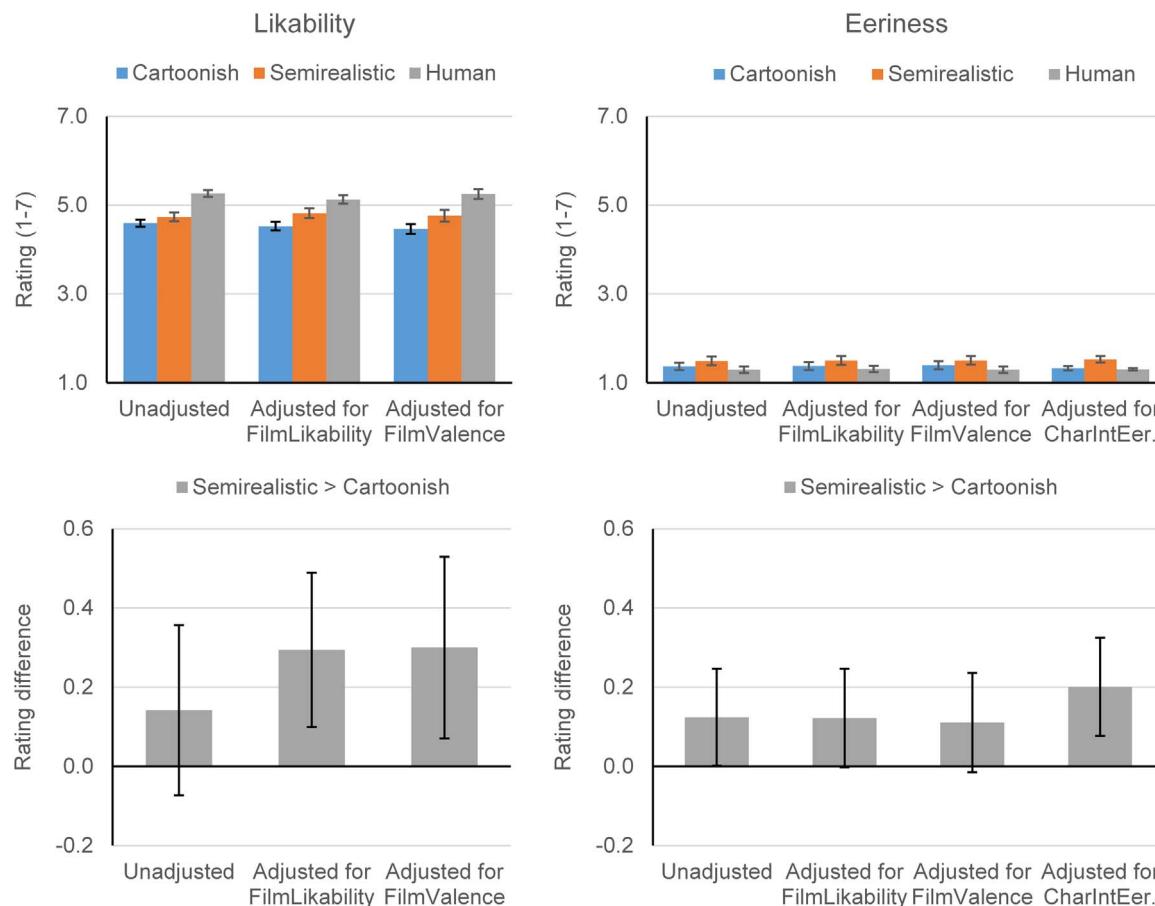
<sup>b</sup> Adjusted values were not calculated.

effect of each confound variable in Table S5 individually. Findings from these individual analyses are illustrated in Fig. 3. The change in likability ratings was explained by the inclusion of either FilmLikability or FilmValence confound. Specifically, the difference between semirealistic and cartoonish animations became significant after adjustment for FilmLikability ( $M=4.82$  and  $4.53$ ;  $SD=0.75$  and  $0.70$ ;  $p=0.004$ ) or FilmValence ( $M=4.76$  and  $4.46$ ;  $SD=0.88$  and  $0.79$ ;  $p=0.012$ ), 95% CIs for the difference [0.10, 0.49] and [0.07, 0.53]. Although eeriness results from unadjusted and joint confound models were similar (Tables 6, 8), the difference between semirealistic and cartoonish animations narrowly missed significance after the individual inclusion of either FilmLikability ( $M=1.50$  and  $1.38$ ;  $SD=0.61$  and  $0.54$ ;  $p=0.053$ ) or FilmValence ( $M=1.51$  and  $1.40$ ;  $SD=0.74$  and  $0.65$ ;  $p=0.081$ ) confound, 95% CIs for the difference [-0.00, 0.25] and [-0.01, 0.24]. However, the difference between semirealistic and cartoonish animations remained marginally significant ( $p < 0.10$ ) and qualitatively similar to unadjusted results. Furthermore, the difference between semirealistic and cartoonish animations was again significant after adjustment for CharIntEeriness confound ( $M=1.53$  and  $1.33$ ;  $SD=0.55$  and  $0.41$ ;  $p=0.002$ ), 95% CI [0.08, 0.33]. Intentional eeriness ratings showed a larger correlation with eeriness ratings than either film likability or valence (Section 3.2). Importantly, when both CharIntEeriness and FilmLikability were included jointly, the difference between semirealistic and cartoonish animations was significant ( $M=1.51$  and  $1.33$ ;  $SD=0.54$  and  $0.41$ ;  $p=0.003$ ), 95% CI [0.06, 0.31]..

### 3.4. Individual films

As can be seen in Table 5, the effect of individual films was significant for all dependent variables. Given that pairwise comparisons between films were considered exploratory, correction for multiple comparisons was not applied in pairwise comparisons between films. Three human actors that were never selected as eerie could originally not be included in the GEE analysis because of its probit link function did not allow zero values for predictors. To include all films into this analysis, three randomly selected responses for human actors were changed from "not selected" to "selected" for the pairwise film comparison.

Pairwise differences between individual films are illustrated in Fig. 4. As expected, human actors received higher human-likeness ratings than semirealistic characters, which received higher ratings than cartoonish characters. Character likability ratings for individual films were inconsistent within the same film types and did hence not provide clear evidence in favour of either hypothesis H2a or H2b. Strangeness ratings were consistent with hypothesis H2a: human actors tended to receive the lowest ratings, cartoonish characters received the highest ratings (except for C5), and semirealistic characters tended to fall in between them. All eeriness ratings were close to floor values with few significant differences between them; however, in



**Fig. 3.** Likability and eeriness evaluations as measured originally and as adjusted for specific individual confounds. Upper row shows ratings by film type (error bars denote one SEM) and lower row shows rating differences between semirealistic and cartoonish animations (error bars denote 95% CIs).

support of hypothesis H2b, film R3 (*Beowulf*) received higher eeriness ratings ( $M=2.03$ ,  $SD=1.39$ ) than any other film. Consistently with hypothesis H3, semirealistic characters received higher classification difficulty ratings than other characters (with one nonsignificant difference between R2 and C5). Explicit selections were consistent with hypothesis H1, given that semirealistic characters received the highest selection rates (with two nonsignificant pairwise differences R5–C2 and R5–C4). After sequential Holm-Bonferroni correction (Holm, 1979) for multiple comparisons, film R2 (*Polar Express*) was selected significantly more often ( $M=67\%$ ,  $SD=45\%$ ; corrected  $p < 0.002$ ) than any other film except R3 (*Beowulf*;  $M=50\%$ ,  $SD=52\%$ ; corr.  $p=1.00$ ), and R3 (*Beowulf*) was selected significantly more often than any other cartoonish or human film (corr.  $p < 0.001$ ).

Result patterns for individual films remained qualitatively similar after adjustment for confounds (joint confound models). In particular, film R3 (*Beowulf*) still received higher eeriness ratings ( $M=1.78$ ,  $SD=0.96$ ) than all other films (uncorrected  $p < 0.04$ ) with the exception of R2 ( $M=1.63$ ,  $SD=1.19$ ; uncorr.  $p=0.357$ ) and the marginally nonsignificant exception of H1 ( $M=1.51$ ,  $SD=0.59$ ; uncorr.  $p=0.054$ ). Similarly, after correction for multiple comparisons, film R2 (*Polar Express*) was still selected as eerie significantly more often than any other film ( $M=68\%$ ,  $SD=48\%$ ; corr.  $p < 0.006$ ) except for R3 (*Beowulf*;  $M=58\%$ ,  $SD=58\%$ ; corr.  $p=1.00$ ), and film R3 was selected as eerie significantly more often (corr.  $p < 0.004$ ) than any other cartoonish or human-acted film.

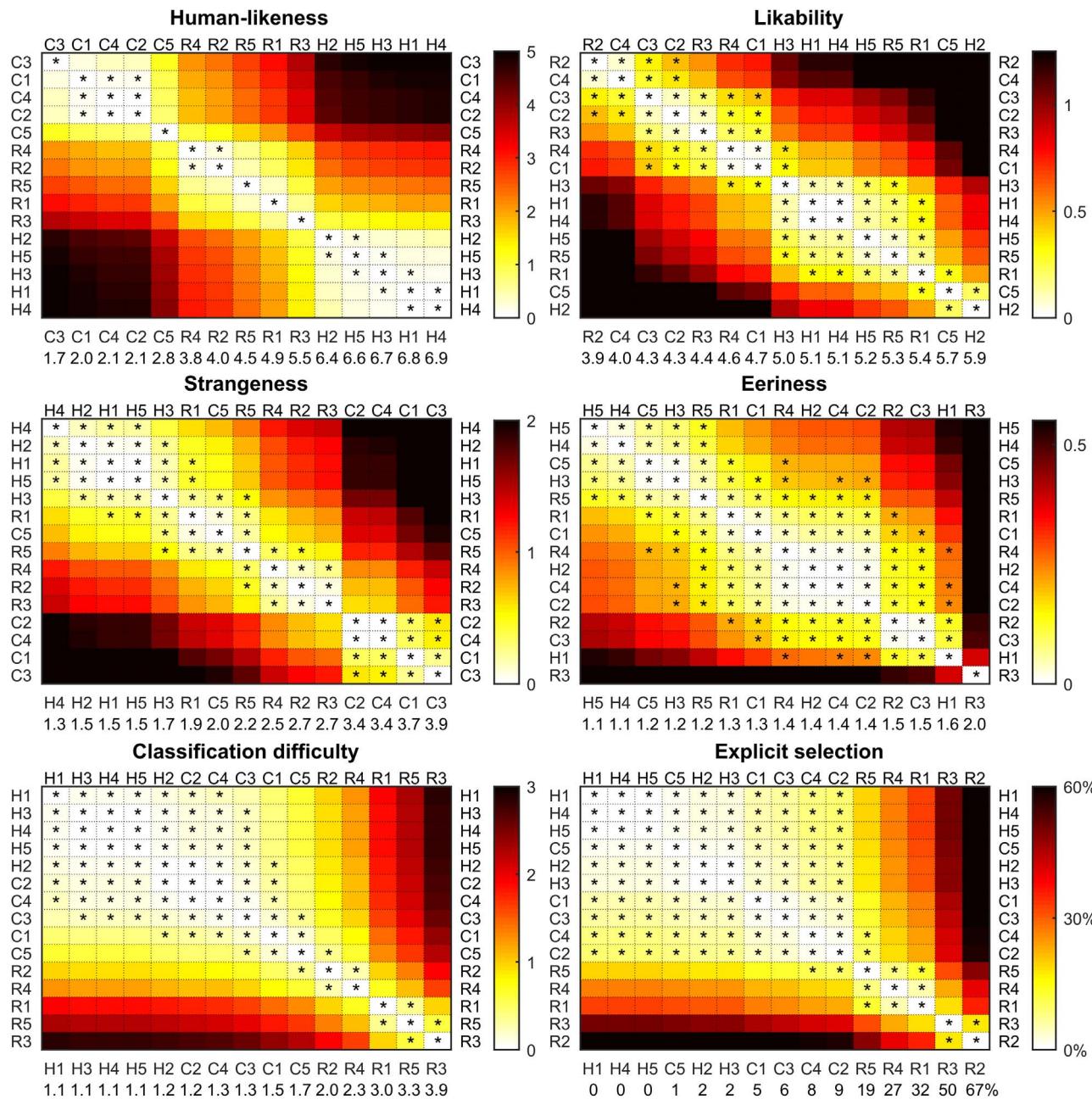
#### 4. Discussion

To the best of our knowledge, the present investigation is the first comprehensive attempt to evaluate the UV hypothesis for animated

film characters. Specifically, we collected a comprehensive sample of semirealistic film characters animated with motion-capture techniques as well as a matched sample of cartoonish characters and human actors, and we then asked participants to rate these materials with respect to conventional self-report items used in previous UV research. Participants were also asked to explicitly select the most representative film characters for the UV phenomenon in a post-experimental evaluation. Mixed model analyses were carried out in an attempt to control for most plausible confounds in film stimuli. Explicit selections and eeriness ratings provided positive support for the UV hypothesis; however, likability and strangeness ratings appeared to provide opposite evidence that increasing human-likeness elicits more positive evaluations in a monotonically increasing manner.

We originally considered that matching semirealistic animated characters correctly with the UV concept in the explicit selection task would be a necessary prerequisite for demonstrating the UV in animated films. Although this finding is not sufficient by itself to demonstrate the UV, implicit eeriness evaluations supported these findings. Because participants received a full briefing of the UV concept and made the explicit selections only after the rating experiment, it is not possible that the eeriness ratings would have been influenced by the explicit selection task. The present confound analyses also failed to show significant eeriness rating differences between participants who knew and who did not know the UV concept prior to this study. Eeriness rating difference between semirealistic and cartoonish animations became nonsignificant when either film likability or film valence were included individually as confound variables; however, the difference remained marginally significant, and again reached significance in the best fitting joint confound model.

Our other evaluations provided the seemingly inconsistent results



**Fig. 4.** Pairwise rating similarities between individual film characters. Values are inverse coded such that lighter and “hotter” colors denote smaller differences equivalent to higher similarities (color scale is different for each panel). Rows and columns are sorted in ascending order based on estimated marginal mean ratings, which are also displayed below the horizontal axes. Asterisks (\*) denote nonsignificant ( $p > 0.05$ ) differences.

that cartoonish characters, not semirealistic ones, appear the least likable and the strangest out of the different film characters. A shortcoming of likability results is that cartoonish characters received lower likability ratings than semirealistic characters only after the evaluations were adjusted for either film likability or film valence confounds. This change clearly resulted from the fact that cartoonish films were considered more likable than semirealistic films, and that the appearance of characters in more likable films also tended to be more likable (the same pattern held true for film valence). Once the evaluations were adjusted for this pattern, the likability of cartoonish characters decreased and the likability of semirealistic characters increased, leading to a significant relative difference. Given that the original difference was not significant, however, this finding could have been a spurious effect caused by the confound adjustment.

Strangeness, on the other hand, was considered separately from

eeriness in the present study only because the combined index suffered from inadequate reliability. Although strangeness has been used in some previous studies either alone (e.g., Mäkäräinen et al., 2014) or as an opposite end to “familiarity” (e.g., MacDorman, 2006; Tinwell et al., 2011), it is not as commonly used as eeriness or creepiness. Furthermore, tentative evidence exists that eeriness or creepiness would better capture the emotional aspects of UV than strangeness, which itself can be considered more cognitive in nature (Ho et al., 2008). The cognitive nature of strangeness could explain the apparently counterintuitive finding that cartoonish animated characters, which according to common sense should have appeared the most appealing, on the contrary were considered the most strange. Specifically, higher strangeness ratings could have resulted from the cognitive observation that cartoonish characters were exaggerated beyond human norms (cf. the principle of exaggeration in traditional animation; Lasseter, 1987).

That cartoonish animations received the highest intentional eeriness ratings is also consistent with this interpretation. In general, eeriness seems to map more directly into Mori's (1970/2012) original concepts than is the case for likability (cf. Bartneck et al., 2009; Ho and MacDorman, 2010; MacDorman and Chattopadhyay, 2016).

Given that the unadjusted likability ratings did not differ between cartoonish and semirealistic characters and the strangeness findings could be explained away as a predominantly cognitive effect, the present findings hence do not support the prediction that the positivity of evaluations increases monotonically across human-likeness, as has been observed in several previous studies (Burleigh et al., 2013; Cheetham et al., 2014; Looser and Wheatley, 2010; MacDorman et al., 2009; Piwek et al., 2014; Seyama and Nagayama, 2007; Thompson et al., 2011). Instead, explicit selections and eeriness ratings provided positive evidence for the prediction that semirealistic animations elicit more negative evaluations than cartoonish or human-acted films. These results are consistent with some previous studies that have demonstrated UV for controlled (Experiment 2 in Burleigh et al., 2013; Ferrey et al., 2015; Yamada et al., 2013) and naturalistic stimuli (Mathur and Reichling, 2016; Poliakoff et al., 2013), and they are also consistent with similar findings for video game and other digital characters (e.g., McDonnell et al., 2012; Schneider and Yang, 2007; Tinwell et al., 2010). The present results can hence be taken as positive evidence for the UV hypothesis in animated film characters.

According to the participants' subjective evaluations, semirealistic animated characters were more difficult to categorize than cartoonish animated characters or human actors. These results tentatively support the notion that semirealistic animated characters are more categorically ambiguous than cartoonish animated characters or human actors. This evidence is only tentative, however, because (i) categorical ambiguity was not included as an independent variable in the experimental design and (ii) categorical perception could not be tested formally with the present film stimuli. Previous categorical perception studies have demonstrated that gradual human-likeness continua are indeed perceived categorically (Cheetham et al., 2011; Looser and Wheatley, 2010). However, categorical perception studies have not yet demonstrated that the most ambiguous stimuli would elicit negative reactions (Cheetham et al., 2014; Looser and Wheatley, 2010). Although some previous studies have demonstrated that intermediate levels between artificial and natural stimuli elicit the most negative evaluations (Burleigh et al., 2013; Ferrey et al., 2015; Yamada et al., 2013), a recent study comparing perceptual mismatch and categorization difficulty hypotheses did not support this finding (MacDorman and Chattopadhyay, 2016).

A disadvantage of natural research stimuli is that total control over all confounding factors is not possible. In the present study, almost all of the considered confound variables differed between film type categories despite our attempts to match the films across categories to the extent possible. In the present mixed model approach, however, we were able to test whether these confounds influenced our results and to reduce their effects accordingly. Confound analyses demonstrated several confound effects. In particular, characters' intentional appearance (whether they had been intended to appear likable or eerie) was strongly associated with the film characters' likability, strangeness, and eeriness. On the other hand, although participants' evaluations of films and film characters should have been separated from each other, in reality we observed a strong association between the overall likability of films and the likability of film characters' appearance, in particular.

As discussed above, the inclusion of confound factors had some effects on the pattern of significant results. First, likability ratings differed significantly between semirealistic and cartoonish animations only after the results were adjusted for film likability or film valence. This change was not considered important, given that our purpose was to use confound factors to exclude their effects on otherwise significant results rather than to identify new ones. Second, similar difference for eeriness ratings narrowly missed significance after the inclusion of

either of these confounds. However, after intentional eeriness, which had the greatest influence on eeriness ratings out of the tested confounds, was included in the confound model either individually or jointly with other confounds, the original difference became substantially larger. Hence, eeriness findings were not changed when all of the important confounds were taken into consideration. Eeriness findings were not changed by any other plausible confounds, which increased our confidence in their validity.

The present mixed model approach also allowed considering differences between individual films. Two semirealistic films, *The Polar Express* (Goetzman et al., 2004) and *Beowulf* (Rapke et al., 2007), received the highest explicit selection percentages in post-experimental evaluations. These findings survived conservative correction for multiple comparisons and were hence robust. More exploratory findings demonstrated that *Beowulf* also received higher eeriness ratings than any other film, whereas the remaining semirealistic films did not differ clearly from other films. This suggests that the present eeriness findings may have been driven mainly by the film *Beowulf*. Interestingly, the main character in *Beowulf* also received the highest human-likeness ratings and it tended to receive the highest classification difficulty ratings (Fig. 4), which suggests that this film may best fit the UV hypothesis out of the studied animated characters. This interpretation should be taken with some caution, however, given that it is based on exploratory analyses.

We consider several methodological limitations for the present findings. First, our findings supporting the UV hypothesis rely on explicit selections and eeriness self-report items that could both be criticized. Although explicit selections provided robust results, we cannot fully exclude the possible influence of demand characteristics. Specifically, it is possible that participants perceived the UV phenomenon in semirealistic films only after they had been explicitly told about this concept. Eeriness ratings, on the other hand, tended to be close to minimum for all film types. This floor effect could have weakened any genuine effects in the experimental stimuli, either in favour or against the UV hypothesis. More importantly, the term "eeriness" is not as theoretically justified for studying emotional responses as for example conventional emotion self-report items (cf. Cheetham et al., 2015). However, it should be noticed that our goal was to test the UV hypothesis in its present form. The adopted self-report items, eeriness in particular, were hence well-motivated because they are consistent with Mori's original formulation (Mori, 1970/2012) and recent empirical studies.

Another limitation is that only five semirealistic films (fifteen in total) survived our preselection procedure. Given the relatively small number of films, we were for example not able to include an equal number of male and female characters into the study. Notably, the three female characters received the highest likability ratings (R1, C5, and H2 in Fig. 4), which suggests that character gender was an uncontrolled confound in the present study. Importantly, gender distribution was matched across film types, however, which means that it could not have influenced the observed differences between film types. A further potential limitation is that all semirealistic stimuli were relatively easy to categorize as nonhuman. This is a limitation because more realistic stimuli could have led to more robust UV findings. During piloting, we experimented with some feature films that have used photorealistic CGI techniques to reconstruct real actors' faces, namely *The Curious Case of Benjamin Button* (Chaffin et al., 2008) and *Tron: Legacy* (Bailey et al., 2010). However, we decided not to include such stimuli in the final experiment, given that we were not able to gather enough photorealistic film material for this additional film type category. The present approach can be justified by the decision to include as homogeneous research stimuli to the evaluation as possible – here, a comprehensive sample of fully animated films using motion-capture technologies and aiming intentionally at (semi)realistic characters. Furthermore, it can be argued that more realistic animations would not have been strictly necessary, as the present semirealistic

animations already provided evidence for the UV phenomenon.

We also acknowledge that the statistical power of the present analyses may have been limited by the small participant and stimulus samples. Although the mixed model analyses were reasonably similar to conventional variance analysis, they also included additional fixed and random parameters for individual films and film types. On the other hand, adopting conventional variance analysis would not have changed the present statistical conclusions, and the more complex analyses were justified by a better fit to the data. Regardless of the adopted analysis paradigm, however, statistical power may have been an issue for the confound analyses because full retrospective control over all uncontrolled confound variables is clearly not possible, especially for the present small number of films surviving the preselection. Future studies should, when possible, consider using larger stimulus sets to average out confound effects originating from individual stimuli.

In the present study, participants viewed film materials with subtitles shown at the bottom part of the screen in their native language. This methodological decision can be justified in terms of ecological validity because viewing foreign films with subtitles was the predominant norm for our participants. Subtitles also reduced the cognitive effort of having to follow spoken narrative in a nonnative language. Although it is possible that subtitles drew some attention away from the animated characters, this effect should at worst have weakened the present observed effects. In particular, it is unlikely that the presence of subtitles would explain the obtained positive findings for explicit selections or eeriness ratings. A technical limitation is that we displayed the present film materials at a lower resolution (standard-rather than high-definition) than would be customary in the cinema. Similarly as above, this effect should have weakened the observed effects rather than caused them. It is also noteworthy that, even though some of our participants were knowledgeable about films and display technologies, display resolution issues were not explicitly mentioned in any of the participants' open feedback.

Taken together, the present findings demonstrate that semirealistic animated films are more eerie than cartoonish or human-acted films. Although an anecdotal connection between specific animated films and the UV hypothesis has been presented repeatedly, and some previous studies have included such films as research stimuli (e.g., Flach et al., 2012; Ho and MacDorman, 2010), to the best of our knowledge no previous studies have yet explicitly compared semirealistic animated films against matched film stimuli. Hence, the present findings are important because they provide empirical evidence for the previously anecdotal connection between semirealistic animated films and the UV. The findings are also important because they provide additional support for the UV hypothesis itself, whose empirical evidence has remained inconsistent. In addition to the limitations discussed above, it should be noticed that all animated films in the present study, including semirealistic ones, received low eeriness ratings. This overall floor effect suggests that the subjective experience of the UV in animated film characters still requires further elaboration. Nevertheless, the present findings suggest that the negative emotional experiences elicited by some animated film characters – in particular *The Polar Express* and *Beowulf* – are worthy of further research attention.

## Acknowledgments

JK was supported by an individual research fellowship grant from The Emil Aaltonen Foundation, and his work also received funding from the European Union's Horizon 2020 research and innovation programme within the framework of the Marie Skłodowska-Curie Individual Fellowship (IF-EF) under grant agreement No 703493. MM was supported by the Graduate School in User-Centered Information Technology. We thank M.Sc. (Tech.) Jussi Tarvainen for lending his technical and cinematic expertise to help establish the

experimental setup. We also want to thank Legal Counsel (IPR) Maria Rehbinder and the Finnish Art University Copyright Advice consortium for copyright advice, and M & M Viihdepalvelu (<http://www.elokuvalisensi.fi>) for providing a commercial license for presenting the copyrighted DVD materials.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.ijhcs.2016.09.010>.

## References

- Aida, J., Lee, C., Sakai, A. (Producer), Sakaguchi, H., Sakakibara, M. (Director), 2001. *Final Fantasy: The Spirits Within* [Motion Picture]. Columbia Pictures, USA, Japan.
- Alexander, O., Rogers, M., Lambeth, W., 2010. The digital emily project: Achieving a photorealistic digital actor. *IEEE Comput. Graph. Appl.* 30, 20–31. <http://dx.doi.org/10.1109/MCG.2010.65>.
- Ansen, D., 2007. Ansen on "Beowulf" and "Margot at the Wedding". *Newsweek*, November 16, 2007. Retrieved 20.09.16, from <<http://europe.newsweek.com/ansen-beowulf-and-margot-wedding-96207>>.
- Bailey, S., Lisberger, S., Silver, J. (Producer), Kosinski, J. (Director), 2010. *Tron: Legacy* [Motion Picture]. Walt Disney Pictures, USA.
- Bartneck, C., Kanda, T., Ishiguro, H., Hagita, N., 2009. My robotic doppelgänger – A critical look at the Uncanny Valley. In: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2009), pp. 269–276. <http://dx.doi.org/10.1109/ROMAN.2009.5326351>.
- Boyd, S.J., Rapke, J., Starkey, S. (Producer), Wells, S. (Director), 2011. *Mars Needs Moms* [Motion Picture]. Walt Disney Pictures, USA.
- Bradley, M.M., Lang, P.J., 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59.
- Brenton, H., Gillies, M., Ballin, D., Chatting, D., 2005. The Uncanny Valley: Does it exist? In: Proceedings of 19th British HCI Group Annual Conference, pp. 2–5.
- 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. *Comput. Hum. Behav.* 29, 759–771. <http://dx.doi.org/10.1016/j.chb.2012.11.021>.
- Chaffin, C., Kennedy, K., Marshall, F. (Producer), Fincher, D. (Director), 2008. *The Curious Case of Benjamin Button* [Motion Picture]. Warner Brothers Pictures, USA.
- Chaminade, T., Hodgins, J., Kawato, M., 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216. <http://dx.doi.org/10.1080/scan.ns017>.
- Cheetham, M., Suter, P., Jäncke, L., 2014. Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley". *Front. Psychol.* 5. <http://dx.doi.org/10.3389/fpsyg.2014.01219>.
- Cheetham, M., Suter, P., Jäncke, L., 2011. The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5. <http://dx.doi.org/10.3389/fnhum.2011.00126>.
- Cheetham, M., Wu, L., Pauli, P., Jäncke, L., 2015. Arousal, valence, and the uncanny valley: psychophysiological and self-report findings. *Front. Psychol.* 6. <http://dx.doi.org/10.3389/fpsyg.2015.00981>.
- Davis, J., Forte, L. (Producer), Wedge, C. (Director), 2013. *Epic* [Motion Picture]. 20th Century Fox Pictures, USA.
- Depp, J., Headdington, T., King, G. (Producer), Scorsese, M. (Director), 2011. *Hugo* [Motion Picture]. Paramount Pictures, USA.
- di Bonaventura, L., Dreyer, M., Gaiman, M. (Producer), Vaughn, M. (Director), 2007. *Stardust* [Motion Picture]. Paramount Pictures, USA.
- Ferrey, A.E., Burleigh, T.J., Fenske, M.J., 2015. Stimulus-category competition, inhibition, and affective devaluation: A novel account of the uncanny valley. *Front. Psychol.* 6. <http://dx.doi.org/10.3389/fpsyg.2015.00249>.
- Flach, L.M., Moura, R.H. de, Musse, S.R., Dill, V., Lykawka, C., Puers, F., 2012. Evaluation of the Uncanny Valley in CG characters. In: Proceedings of SBGames 2012, pp. 108–116.
- Franzoni, D., Lustig, B., Wick, D. (Producer), Scott, R. (Director), 2000. *Gladiator* [Motion Picture]. Universal Pictures, USA.
- Gee, F.C., Browne, W.N., Kawamura, K., 2005. Uncanny valley revisited. In: Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005), pp. 151–157. <http://dx.doi.org/10.1109/ROMAN.2005.1513772>.
- Gleiberman, O., 2011. The Adventures of Tintin. *Entertain. Wkly.*, December 22, 2011. Retrieved 20.09.16, from <<http://www.ew.com/article/2011/12/22/adventures-tintin>>.
- Goetzman, S., Starkey, S., Teitler, W. (Producer), Zemeckis, R. (Director), 2004. *The Polar Express* [Motion Picture]. Warner Brothers Pictures, USA.
- Goldstone, R.L., Hendrickson, A.T., 2010. Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 69–78. <http://dx.doi.org/10.1002/wcs.026>.
- Gordon, L., Levin, L. (Producer), de Bont, J. (Director), 2003. *Lara Croft Tombraider: The Cradle of Life* [Motion Picture]. Paramount Pictures, USA.
- Green, R.D., MacDorman, K.F., Ho, C.-C., Vasudevan, S., 2008. Sensitivity to the proportions of faces that vary in human likeness. *Comput. Hum. Behav.* 24, 2456–2474. <http://dx.doi.org/10.1016/j.chb.2008.02.019>.
- Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. *Cogn. Emot.* 9,

- 87–108. <http://dx.doi.org/10.1080/02699939508408966>.
- Hanson, D., 2006. Exploring the aesthetic range for humanoid robots. In: Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, pp. 39–42. < <http://www.androidscience.com/proceedings2006/6Hanson2006ExploringTheAesthetic.pdf> >
- Hanson, D., 2005. Expanding the aesthetic possibilities for humanoid robots. In: Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots. Hayes, A.F., 2006. A primer on multilevel modeling. *Hum. Commun. Res.* 32, 385–410. <http://dx.doi.org/10.1111/j.1468-2958.2006.00281.x>.
- Hecht, A., Oedecker, K. (Producer), Davis, J.A. (Director), 2001. *Jimmy Neutron: Boy Genius* [Motion Picture]. Paramount Pictures, USA.
- Hill, L., 2011. Movie Review: Mars Needs Moms - and Better Animation. *Vulture*, March 8, 2011. Retrieved 20.09.16, from < [http://www.vulture.com/2011/03/movie\\_review\\_the\\_alien\\_humans.html](http://www.vulture.com/2011/03/movie_review_the_alien_humans.html) > .
- Ho, C.-C., MacDorman, K.F., 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. <http://dx.doi.org/10.1016/j.chb.2010.05.015>.
- Ho, C.-C., MacDorman, K.F., Pramono, Z.D., 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. ACM, pp. 169–176.
- Hoffman, L., Rovine, M.J., 2007. Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behav. Res. Methods* 39, 101–117. <http://dx.doi.org/10.3758/BF03192848>.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70. <http://dx.doi.org/10.2307/4615733>.
- Jackson, P., Kennedy, K. (Producer), Spielberg, S. (Director), 2011. *The Adventures of Tintin: The Secret of the Unicorn* [Motion Picture]. Columbia Pictures, USA.
- Janssen, D.P., 2012. Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behav. Res. Methods* 44, 232–247. <http://dx.doi.org/10.3758/s13428-011-0145-1>.
- Kaba, F., 2013. Hyper-realistic characters and the existence of the uncanny valley in animation films. *Int. Rev. Soc. Sci. Humanit.* 4, 188–195.
- 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. *Front. Psychol.* 390, 6. <http://dx.doi.org/10.3389/fpsyg.2015.00390>.
- Kempley, R., 2001. "Final Fantasy": Virtually unwatchable. *Wash. Post*, July 11, 2011.
- Kenny, D.A., Kashy, D.A., Cook, W.L., 2006. *Dyadic Data Analysis*. The Guilford Press, New York.
- Kroopf, S., de Luca, M., Teitler, W. (Producer), Favreau, J. (Director), 2005. *Zathura: A Space Adventure* [Motion Picture]. Columbia Picture, USA.
- Lasseter, J., 1987. Principles of traditional animation applied to 3D computer animation. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), pp. 35–44.
- Levine, L., Lieberman, H., Paterson, D. (Producer), Csupo, G. (Director), 2007. *Bridge to Terabithia* [Motion Picture]. Walt Disney Pictures, USA.
- Looser, C.E., Wheatley, T., 2010. The tipping point of animacy: How, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. <http://dx.doi.org/10.1177/0956797610388044>.
- MacDorman, K.F., 2005. Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? In: CogSci-2005 Workshop: Toward Social Mechanisms of Android Science, pp. 108–118.
- MacDorman, K.F., 2006. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In: Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, pp. 26–29. < <http://www.androidscience.com/proceedings2006/8MacDorman2006Subjective.pdf> > .
- 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. <http://dx.doi.org/10.1016/j.cognition.2015.09.019>.
- MacDorman, K.F., Green, R.D., Ho, C.-C., Koch, C.T., 2009. Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. <http://dx.doi.org/10.1016/j.chb.2008.12.026>.
- MacDorman, K.F., Ishiguro, H., 2006. The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. <http://dx.doi.org/10.1075/is.7.3.03mac>.
- Mäkäräinen, M., Kätsyri, J., Takala, T., 2014. Exaggerating facial expressions: A way to intensify emotion or a way to the Uncanny Valley? *Cogn. Comput.* 6, 708–721. <http://dx.doi.org/10.1007/s12559-014-9273-0>.
- Marsden, P. (Producer), Lord, P., Miller, C. (Director), 2009. *Cloudy with a Chance of Meatballs* [Motion Picture]. Sony Pictures, USA.
- 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. <http://dx.doi.org/10.1016/j.cognition.2015.09.008>.
- McDonnell, R., Breidt, M., Bühlhoff, H.H., 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31, 1–11. <http://dx.doi.org/10.1145/2185520.2185587>.
- McKim, D. (Producer), Anderson, S. J. (Director), 2007. *Meet the Robinsons* [Motion Picture]. Walt Disney Pictures, USA.
- Misselhorn, C., 2009. Empathy with inanimate objects and the Uncanny Valley. *Minds Mach.* 19, 345–359. <http://dx.doi.org/10.1007/s11023-009-9158-2>.
- Mori, M., 1970/2012. The Uncanny Valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robot. Autom. Mag.* 19, 98–100. <http://dx.doi.org/10.1109/MRA.2012.2192811>.
- Nezlek, J.B., 2008. An introduction to multilevel modeling for social and personality psychology. *Soc. Personal. Psychol. Compass* 2, 842–860. <http://dx.doi.org/10.1111/j.1751-9004.2007.00059.x>.
- Pegram, S. (Producer), Smith, S., Cook, B. (Director), 2011. *Arthur Christmas* [Motion Picture]. Sony Pictures, USA.
- Perry, T., 2014. Leaving the uncanny valley behind. *IEEE Spectr.* 51, 48–53. <http://dx.doi.org/10.1109/MSPEC.2014.6821621>.
- Phillips, M., 2011. Disney motion-captures a baffling, wobbly dud. "Mars Needs Moms" – 1/2 stars. *Chic. Trib.*, March 10, 2011. Retrieved 20.09.16, from < [http://articles.chicagotribune.com/2011-03-10/entertainment/sc-mov-0309-mars-needs-moms-20110310\\_1\\_mars-needs-moms-mars-needs-moms-performance-capture](http://articles.chicagotribune.com/2011-03-10/entertainment/sc-mov-0309-mars-needs-moms-20110310_1_mars-needs-moms-mars-needs-moms-performance-capture) > .
- Piwiek, L., McKay, L.S., Pollick, F.E., 2014. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. <http://dx.doi.org/10.1016/j.cognition.2013.11.001>.
- Plantec, P., 2007. Crossing the Great Uncanny Valley. *Animat. World Netw.*, December 19, 2007. Retrieved 20.09.6, from < <http://www.awn.com/vfxworld/crossing-great-uncanny-valley> > .
- Poliakoff, E., Beach, N., Best, R., Howard, T., Gowen, E., 2013. Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception* 42, 998–1000. <http://dx.doi.org/10.1086/p7569>.
- Pollick, F.E., 2010. In search of the uncanny valley. *Lect. Notes Inst. Comput. Sci. Soc. Inform. Telecommun. Eng.* 40, 69–78. [http://dx.doi.org/10.1007/978-3-642-12630-7\\_8](http://dx.doi.org/10.1007/978-3-642-12630-7_8).
- Quené, H., van den Bergh, H., 2004. On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Commun.* 43, 103–121. <http://dx.doi.org/10.1016/j.specom.2004.02.004>.
- Rapke, J., Starkey, S. (Producer), Kenan, G. (Director), 2006. *Monster House* [Motion Picture]. Columbia Pictures, USA.
- Rapke, J., Starkey, S. (Producer), Zemeckis, R. (Director), 2009. *A Christmas Carol* [Motion Picture]. Walt Disney Pictures, USA.
- Rapke, J., Starkey, S. (Producer), Zemeckis, R. (Director), 2007. *Beowulf* [Motion Picture]. Paramount Pictures, USA.
- Robinson, T., 2007. Movie review: Beowulf. *AV Club*, November 16, 2007. Retrieved 20.09.16, from < <http://www.avclub.com/review/beowulf-3194> > .
- Sarafian, K. (Producer), Andrews, M., Chapman, B., Purcell, S. (Director), 2012. *Brave* [Motion Picture]. Walt Disney Pictures, USA.
- Savlov, M., 2004. The Polar Express. *Austin Chron.*, November 12, 2004. Retrieved 20.09.16, from < <https://www.austinchronicle.com/calendar/film/2004-11-10/the-polar-express/> > .
- Saygin, A.P., Chaminaud, T., Ishiguro, H., Driver, J., Frith, C., 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. <http://dx.doi.org/10.1093/scan/nsr025>.
- Schneider, E., Yang, S., 2007. Exploring the Uncanny Valley with Japanese video game characters. In: Proceedings of DiGRA 2007 Conference: Situated Play, pp. 546–549.
- Seyama, J., Nagayama, R.S., 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence Operators Virtual Environ.* 16, 337–351. <http://dx.doi.org/10.1162/pres.16.4.337>.
- Snijders, T.A.B., Bosker, R.J., 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications, London.
- Steckenfinger, S.A., Ghazanfar, A.A., 2009. Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci.* 106, 18362–18366. <http://dx.doi.org/10.1073/pnas.0910063106>.
- Stevens, D., 2011. Tintin, So So. Steven Spielberg's motion-capture adventure has its charms, but it's no Raiders. *Slate*, December 21, 2011. Retrieved 20.09.16, from < [http://www.slate.com/articles/arts/movies/2011/12/adventures\\_of\\_tintin\\_review\\_spielberg\\_s\\_motion\\_capture\\_adventure\\_has\\_its\\_charms\\_but\\_it\\_s\\_no\\_raiders\\_.html](http://www.slate.com/articles/arts/movies/2011/12/adventures_of_tintin_review_spielberg_s_motion_capture_adventure_has_its_charms_but_it_s_no_raiders_.html) > .
- Thompson, J.C., Trafton, J.G., McKnight, P., 2011. The perception of humanness from the movements of synthetic agents. *Perception* 40, 695–704. <http://dx.doi.org/10.1068/p6900>.
- Tinwell, A., Grimshaw, M., Williams, A., 2011. The Uncanny Wall. *Int. J. Arts Technol.* 4, 326–341. <http://dx.doi.org/10.1504/IJART.2011.041485>.
- Tinwell, A., Grimshaw, M., Williams, A., 2010. Uncanny behaviour in survival horror games. *J. Gaming Virtual Worlds* 2, 3–25. [http://dx.doi.org/10.1386/jgvw.2.1.3\\_1](http://dx.doi.org/10.1386/jgvw.2.1.3_1).
- Tondu, B., 2012. Anthropomorphism and service humanoid robots: an ambiguous relationship. *Ind. Robot. Int. J.* 39, 609–618. <http://dx.doi.org/10.1108/01439911211268840>.
- Walker, P. (Producer), Bird, B. (Director), 2004. *The Incredibles* [Motion Picture]. Walt Disney Pictures, USA.
- Wang, S., Lilienfeld, S.O., Rochat, P., 2015. The uncanny valley: Existence and explanations. *Rev. Gen. Psychol.* 19, 393–407. <http://dx.doi.org/10.1037/gpr0000056>.
- Weschler, L., 2002. Why is this man smiling? *Wired*, January 6, 2002. Retrieved 20.09.16, from < <https://www.wired.com/2002/06/face/> > .
- Yamada, Y., Kawabe, T., Ihaya, K., 2013. Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. <http://dx.doi.org/10.1111/j.1468-5884.2012.00538.x>.