



Can Digital Face-Morphs Influence Attitudes and Online Behaviors?

Eyal Peer, *Bar-Ilan University*; Sonam Samat and Alessandro Acquisti,
Carnegie Mellon University

<https://www.usenix.org/conference/soups2018/presentation/peer>

**This paper is included in the Proceedings of the
Fourteenth Symposium on Usable Privacy and Security.**

August 12–14, 2018 • Baltimore, MD, USA

ISBN 978-1-939133-10-6

**Open access to the Proceedings of the
Fourteenth Symposium
on Usable Privacy and Security
is sponsored by USENIX.**

Can Digital Face-Morphs Influence Attitudes and Online Behaviors?

Eyal Peer

Graduate School of Business
Administration, Bar-Ilan University,
Israel
eyal.peer@biu.ac.il

Sonam Samat

Heinz College, Carnegie Mellon
University
Pittsburgh, PA, USA
sonamsamat@gmail.com

Alessandro Acquisti

Heinz College, Carnegie Mellon
University
Pittsburgh, PA, USA
acquisti@andrew.cmu.edu

ABSTRACT

Self-images are among the most prevalent forms of content shared on social media streams. Face-morphs are images digitally created by combining facial pictures of different individuals. In the case of self-morphs, a person's own picture is combined with that of another individual. Prior research has shown that even when individuals do not recognize themselves in self-morphs, they tend to trust self-morphed faces more, and judge them more favorably. Thus, self-morphs may be used online as covert forms of targeted marketing – for instance, using consumers' pictures from social media streams to create self-morphs, and inserting the resulting self-morphs in promotional campaigns targeted at those consumers. The usage of this type of personal data for highly targeted influence without individuals' awareness, and the type of opaque effect such artifacts may have on individuals' attitudes and behaviors, raise potential issues of consumer privacy and autonomy. However, no research to date has examined the feasibility of using self-morphs for such applications. Research on self-morphs has focused on artificial laboratory settings, raising questions regarding the practical, in-the-wild applicability of reported self-morph effects. In three experiments, we examine whether self-morphs could affect individuals' attitudes or even promote products/services, using a combination of experimental designs and dependent variables. Across the experiments, we test both designs and variables that had been used in previous research in this area and new ones that had not. Questioning prior research, however, we find no evidence that end-users react more positively to self-morphs than control-morphs composed of unfamiliar facial pictures in either attitudes or actual behaviors.

1. INTRODUCTION

Face composites, or face-morphs, consist of facial images merged together to produce a new, realistic-looking image of a person that contains some of the elements of the comprising facial images [11]. A substantial body of work has shown that individuals sometimes fail to consciously recognize themselves in face composites that contain their own picture [4], but tend to prefer such self-morphs, trusting them more and finding them more attractive [12,13] when compared to morphs that do not contain the individual's own facial image. Facial images are commonly used in advertising (e.g., of

models or celebrities), and if morphs are effective in influencing end-users' attitudes and behavior, that could have far reaching implications for marketers [28] but also for consumer privacy. Consider a marketer who has access to a consumer's Facebook profile. That marketer may use a picture of that consumer in an ad for a product. Such use of the consumer's picture may be deemed unethical (or even appalling), and would probably not promote the marketer's goals. However, what if the marketer instead used the consumer's picture to create a digital morph that combined that picture with an unknown face? The consumer might not consciously recognize this self-morph. The morph, however, could still evoke strong and positive emotional responses in the consumer, due to the familiar elements it contains. How would consumers react to this implicit, visceral mode of persuasion? Social media users make many types of personal information publicly available [2]. Firms use that information to learn more about potential customers and target advertisements accordingly [32], sometimes influencing end-users [20] without their explicit consent or awareness - a form of hidden "digital market manipulation" [7]. Leveraging individuals' innate attraction to self-morphs to promote products is an example of a targeted marketing strategy [10] that may influence end-users' actions while operating outside their awareness, raising potential yet significant privacy concerns.

Existing research has examined the impact of celebrity morphs on consumers' behavior [28], but not the potential impact of self-morphs as a covert and visceral forms of targeted marketing. Moreover, research on self-morphs has been limited to artificial laboratory settings, raising questions about the generalizability and applicability of the reported effects. We explore the uncharted territory of the impact of self-morphs on consumers' behavior in settings that more closely model real-life conditions. Unlike prior studies (that relied on taking photos of subjects in a lab, thus raising awareness among subjects about the purposes of the experiments), we examine whether self-morphs could be created using individuals' personal information from their social network profiles, and then used without subjects' awareness. Furthermore, unlike prior studies (that focused on participants' attitudes towards facial morphs, including trust) we examine to what extent self-morphs can affect also behavioral intentions, including purchase intentions. Our work thus ties into the privacy literature in two ways. First, it highlights how, due to the vast self-dissemination of personal information, public yet personal data can be used in interactions with consumers by both services and independent third parties surreptitiously – that is, without the former's awareness. Second, it highlights potential limits on individual autonomy [27] in decision making by examining the effectiveness of technologies that may covertly influence consumer decision making based on their own data – a form of "visceral targeting," so to say.

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2018.
August 12 -- 14, 2018, Baltimore, MD, USA.

In three online and lab experiments, we tested whether self-morphs impact consumers' attitudes and online behavior. We relied on real world data (images posted on social networking sites by experiment participants) and used various dependent variables and a combination of techniques and study designs. Unlike prior research, we found no differences in how consumers judge, or react to, a self-morph vs. a control-morph composed of unfamiliar facial pictures in both realistic settings and in settings that follow previous lab experiments, as well as both when focusing on attitudinal variables and when focusing on behavioral intentions. Indeed, across the experiments, we ended up testing both 1) new designs and variables that had not been used in previous research in this area (as we initially focused on extending prior work) and 2) designs and variables that had already been reported (when we switched to replicating prior research). Our research endeavor did not begin as a replication attempt: building upon the apparent robustness of results in existing literature (see Section 1.2), at the onset we focused on capturing behavioral metrics (such as behavioral intentions in Study 1, and actual self-disclosures in Study 2, as opposed to the attitudinal variables more common in previous studies) to measure to what extents face-morphs derived from social media could affect actual online behaviors. As both initial pilots and main experiments failed to find such an effect, we traced back our efforts to attempt to replicate existing results in the literature, without finding significant results. Nevertheless, such null results are worthy to be reported through the scientific community for several reasons. First, null findings (when backed by appropriate methodologies) can be important and enlightening [17], especially in light of the recent evidence of non-replicability of major findings revealed in many scientific areas [23]. Second, the results suggest that, if self-morphs have any effect on people's judgments and behavior in the lab, that effect may not robustly extend to other settings.

1.2 Related Work

Facial images are an exceptional type of perceptual stimuli. Evidence from neuroscience, in support of the 'face-specificity' hypothesis, suggests that the brain has specialized cognitive and neural mechanisms dedicated to face perception [19]. Further evidence suggests that the brain implicitly and automatically evaluates faces, thus enabling individuals to make social judgments about unfamiliar individuals from facial properties alone [11,14,31]. With the advent of computer graphics, face-morphing technology has made it possible to alter the familiarity of faces. Morphing a familiar face into an unfamiliar one creates a composite that has familiar features but may still be unrecognizable as a whole. Previous research has shown that participants fail to consciously recognize themselves in face composites created by morphing their own face with an unknown face [12,4]. This happens when the unknown face contributes a larger proportion of the composite (e.g., 60%) while the self-face contributes a smaller proportion (e.g., 40%). Despite this lack of conscious recognition, participants tend to prefer self-face composites. Other researchers have studied face composites created with a family member's or a friend's face [6,24].

As has been noted, a substantial amount of prior work has studied the effect of self-morphs on individuals' attitudes. DeBruine found that participants tend to trust self-morphs more than non self-morphs [12]. DeBruine also studied the attractiveness of self vs. non-self-morphs and found that participants find self-morphs more attractive [13]. Bailenson et al., created composites of participants with electoral candidates (with the participant's face contributing the smaller proportion and the candidate's face contributing the

larger proportion) and found that participants report higher intentions to vote for self-like candidates than for non-self like candidates [4]. Tanner and Maeng morphed Tiger Woods's face with a stock model's face. They collected data on willingness to buy from this composite versus a control composite before and at the peak of the famous Tiger Woods scandal [28]. They found a significant decline in reported levels of willingness to buy from the Tiger-morph after the scandal. These results have been explained through a "familiarity based valence accessibility" account. This hypothesis assumes that implicit recognition of a familiar individual in a morphed face is sufficient to enable an underlying (and pre-existing) valence judgment of the familiar individual to be automatically perceived [28].

2. THE CURRENT RESEARCH

Although various studies have examined the effects of face composites on various dependent variables, a number of unrequited issues require additional research. From a methodological perspective, most (if not all) of the previous studies have used pictures that were explicitly solicited from the participants – thus, participants may have been (perhaps subconsciously) aware of the research questions or objectives, making a demand effect possible. Furthermore, most previous research on self-morphs used artificial lab environments, as is customary and warranted for basic cognitive and perceptual psychological research. However, the use of such strict settings limits the generalizability of the research findings to actual real-world scenarios, reducing the potential implications of these findings for the HCI and privacy communities, as well as for every day users of online technologies.

In our research, we focused on more realistic and privacy-sensitive settings: we used pictures taken from participants' online social network profiles (specifically, from their Facebook profiles), without their explicit *ex ante* knowledge or awareness (while still ensuring proper experimental consent; all studies were conducted with IRB approval of our institution, and all studies secured informed consent of participants) in order to rule out the possibility that previous findings were, to some degree, confounded by expectation effects. Furthermore, this novel use of pictures from online social networks data also allows us to focus our examination to domains that are of interest to human-computer interactions. Namely, while previous research on self-morphs focuses on people's judgments and attitudes such as trust (e.g. [12,13]), in our studies we mostly focused on behaviors that directly pertain to online consumer behavior (such as purchasing intentions and self-disclosure behaviors) and highlight how consumers' personal data may be not merely accessed, but also used, in manners that are hard for end-users to predict or prevent.

In their 2009 staff report, the Federal Trade Commission (FTC) defines the term 'behavioral advertising' as "the tracking of a consumer's online activities over time – including the searches the consumer has conducted, the web pages visited, and the content viewed – in order to deliver advertising targeted to the individual consumer's interests" [16]. The industry greatly favors the use of such targeted ads because, in comparison to non-targeted ads, targeted ads generate higher click-through rates [15] and higher sales [5]. While personalization of ads can benefit consumers by exposing them to relevant products, the extensive collection and use of personal information also raises consumers' privacy risks and concerns. In fact, consumer surveys about perceptions of targeted advertising suggest that, by and large, people do not like being tracked and do not wish to receive targeted ads [29,21,25]. In this paper, we investigate the effect of individuals' facial images,

in the form of self-morphs, on online consumer behaviors (such as purchasing intentions and self-disclosure behaviors).

2.1 Overview of Studies

The design of our studies builds upon prior research on self-morphs. The set of studies covers an array of experimental setups, participants' pools, and dependent variables. Two studies (Studies 1-2) were conducted online, using pictures obtained from the participants' online social network (Facebook) profiles (thus, we used these pictures without explicit, ex ante participants' awareness, in order to ensure that any observed effects could only be attributed to the implicit exposure to self-morphs); one replication study (Study 3), instead, was conducted in a lab, using photos captured in the lab at the onset of the experiment. For technical reasons (explained further below), Study 1 only included Caucasian males and Study 2 focused on Caucasian females. Study 3 included participants from both genders. Study 1 focused on purchasing or hiring intentions (in addition to measures traditionally captured in morph studies, such as perceived trustworthiness); Study 2 focused on self-disclosure—a variable common in online privacy research, but novel in the context of morph studies; in Study 3, we only focused on replicating previous studies' results using a trustworthiness dependent variable. Studies 1 and 2 were conducted online; although more ecologically appropriate for testing online visceral marketing strategies, Studies 1 and 2 relied on a two-step design (discussed at length below), and therefore required significant per-participant recruitment and retention efforts; in Study 3, we conducted a large-scale laboratory experiment with a larger sample and higher power.

2.2 Morph Preparation

All three studies (and the pilots we ran to test our experimental infrastructure) relied on a two-step design: in a first phase, participants' facial images were collected (either from their publicly available Facebook profiles in Studies 1-2, or by taking a photo of them in the lab in Study 3). The second phase took place either several weeks after (Studies 1-2) or a few minutes after (Study 3) the first phase. Before phase 2, we created morphed images for the experimental and control conditions for each participant using Abrosoft's FantaMorph (www.fantamorph.com, see examples in Figures 1 and 2). During the second phase, morphed images were shown to the subjects as part of the studies' respective experimental designs. In the rest of this section, we describe the process through which we collected images for making the morphs.

Participants in the online studies were invited on Amazon Mechanical Turk to take part in a survey about Facebook activity. The survey took less than 5 minutes and participants were paid 50 cents for their participation. The survey included various questions about Facebook (such as how often and for what purposes participants use Facebook) to establish the study's legitimacy. The last question in the survey was the question of interest to us: participants were told that we were interested in collecting data from their Facebook profile in order to validate whether they would be eligible for future studies, and for this reason we asked them to provide a link to their Facebook profile page. We assured participants that we would only collect publicly available data, and that this question was optional – participants were informed that they could skip the question and still receive full payment. This enabled us to get access to Facebook profiles of our MTurk participants and collect their publicly shared facial images. These images were then used to create morphs to be used in the second phase. The morph-creation process replicated the methodology

used in prior research published in this area (specifically, [4]; see also Sections 2.3 and 5.1). Using this approach, we surveyed over 10,000 participants from MTurk and about 50% of them gave us links to their Facebook profiles. About 20% of those had publicly shared facial images which could be used in morphs (images that are well illuminated, good resolution, and where the participant's face is front-facing with neutral expression). These participants comprised our sampling population from which we recruited participants for the second phase of Studies 1 and 2, taking into account participants' ethnicity and gender (which they reported in the first phase survey).

Study 3 followed a similar two-step approach. However, Study 3's participants were invited to a lab, where their photo was taken and used to make morphs that were immediately shown to them. As noted, we also conducted two online pilots to test and hone our technical and experimental two-step procedure.

2.3 A Note on Replication

Our research endeavor did not begin as a replication attempt. Initially we focused on capturing the impact of face-morphs on new dependent variables that had not been the focus of prior research (Studies 1 and 2 and their pilots). We attempted to replicate existing results on previously used dependent variables (in Study 3) only after failing to find effects for our behavioral dependent variables. That noted, across all three studies presented here, we did try to follow as closely as possible all the technical steps in designing face-morphs and in presenting them to participants. While exact replication of methods was made harder by the fact that not all previously published papers comprehensively disclosed their methods, and not all authors were responsive to our requests for their materials, we were able to follow most closely the method used in [4], whose authors were the most responsive to our questions regarding their experimental material. The authors of [4] were responsive to questions and shared with us details of their morphing software (Magic Morph). Furthermore, given the large sample size used in their study, their clear description of the methodology employed, and the magnitude of effects reported, [4] seemed like one of most rigorous approaches and methods to follow. Thus, our morphing strategy was based on [4], although we utilized a different morphing software, Fanta Morph (after [4] was published we found in rounds of tests that new software Fanta Morph produced more realistic morphs). We also followed [4] in criteria for picking suitable images (for instance, images where participants were not wearing glasses).

3. STUDY 1

The scenario and setting chosen for this study was searching and hiring a private instructor online. We aimed to explore whether instructors whose images would be made of self-morphs would be regarded more favorably, giving them an advantage in the hiring process. This scenario illustrates one of the many ways sophisticated online entities could exploit individual's self-images that are publicly available on social network websites.

3.1 Method

A review of the previous studies on the effects of face composites revealed that the effect sizes (Cohen's *d*) ranged from 0.39 [28] to 0.70 SDs [6]. Based on that, we estimated an effect size of about 0.4 SD and aimed at a sample that would provide about 80% power to detect such an effect. We were able to recruit 118 Caucasian males ($M_{\text{age}} = 28.3$, $SD = 7.5$) through our pool of MTurk participants, which completed the study for \$1.5 in an average duration of about five minutes. This sample had a power of about

71% to detect the estimated effect size with a two-sided test, or 82% power for a one-sided test (i.e., to show that self-morphs are more attractive than control-morphs).

After reading and agreeing to the consent form, participants were given a list of musical instruments (e.g., guitar, violin, piano, etc.) and were asked to choose one instrument they would most like to learn to play. Then, participants were asked to imagine they are looking to hire an instructor who can teach them how to play the instrument they have chosen. They were then shown two images of two private instructors that they, supposedly, found in their online searches. Instructors were called “A” and “B”, both were reported to have had 10 years of experience in playing this musical instrument and both reportedly charged \$10 for a lesson. Participants were asked to indicate which instructor (A or B) they would personally choose to hire. One of the instructors’ images (randomly selected) was a self-morph, while the other was a morph of two unfamiliar persons. One of these two unfamiliar persons was a randomly selected other participant’s face (used at 40% in the morph) and the other face was a second spokesperson, different from the spokesperson used in the self-morph (unfamiliar to participants; used at 60% in the morph). An example is given in Figure 1. We used two spokespersons because using the same spokesperson would result in two very similar looking morphs where the differences would entirely be because of the 40% face used in making the morph. This could prompt the participants to specifically look for subtle differences between the faces and perhaps interpret the goal of the study. We randomly varied whether the self-morph was created with one stock-model or the other and whether it appeared on the left or the right in a split-panel.

For participants randomly assigned to the treatment (or “self”) condition, the morph was created by combining the participant’s face (obtained from publicly shared images on his Facebook profile) with the stock model’s face. The participants randomly assigned to the control (or “other”) condition viewed the same ads, but the face shown to them was a morph created by combining a randomly chosen other participant’s face from among the participants in the treatment condition with the same stock model’s face. This procedure ensured that the participants in the control condition viewed (in aggregate) the same images as the participants in the treatment (i.e., self-morph) condition, and that the only difference between the conditions was that for participants in the treatment condition the morphed image included their own face, whereas for participants in the control condition it did not. Allocation to treatment vs. control condition was done before participants started the study, using a computerized randomizer that assigned each invited participant to either be in the control or treatment condition.

Afterwards, participants were asked to rate each instructor (separately) on how trustworthy, attractive and knowledgeable he seemed to them, how much they liked him, how similar they thought he was to themselves and how strongly they identified with him. They also indicated if they found anything strange or unusual, or familiar, in either of the instructors’ images, and if they said they did, then we asked them to elaborate further (six participants said they recognized themselves in the image and were thus dropped from the analysis). In the next section of the study, participants were presented with five facial images, two of them had instructor A and B, and were asked to identify who was instructor A, and who was B. This was a check question included to ensure that participants remembered which face was which, because the questions on whether participants found anything unusual or

familiar with the morph were asked on a screen that did not show participants the morphs. Then, participants completed the Narcissistic Personality Inventory (NPI, [3]), which we included to examine whether the impact of a self-morph could be restricted to people who hold a higher, more self-loving, perception of themselves.

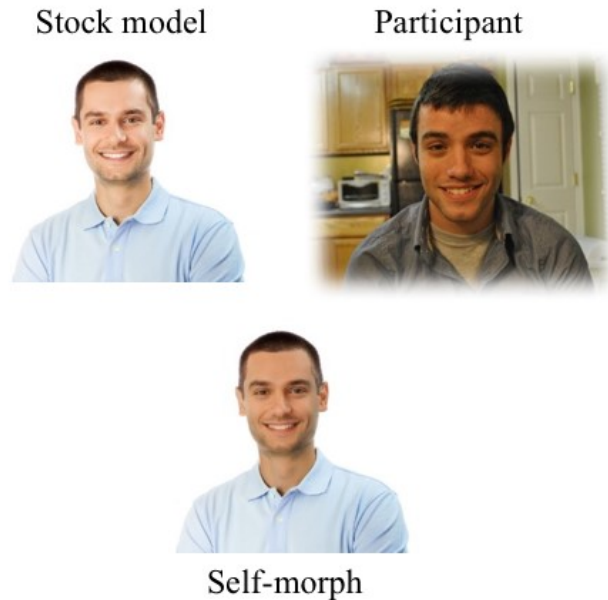


Figure 1. Example stimuli used in Study 1.

Lastly, participants completed the demographic questions and were fully debriefed. In this and in all the following studies, participants were explained that the study was connected to a Facebook survey they had previously completed, and that the researchers may have used publicly available pictures from their Facebook profile for the purposes of the current study. We explained that we did so only for research purposes and that all images collected for this study were kept secure and confidential. We also asked participants to refrain from sharing the details of this study with anyone else in any manner, until the study is completed. Participants’ responses to all questions in all studies (including responses to sensitive questions) were always kept separate from their personal or identifiable information. In all studies, participants were given contact details of the researchers and the IRB, and could also leave comments, concerns or complaints in the survey form itself. Moreover, we provided participants with the option to withdraw their responses from the survey by clicking on a link to a withdrawal form. All participants were thanked and paid, regardless of their final consent. (The option to withdraw responses after being debriefed was provided in all studies, but only two participants chose to use it.) Our IRB approved the procedure of this and all studies reported in the paper. Experimental materials used in Study 1 (as well as Studies 2 and 3) can be found online at <https://www.heinz.cmu.edu/~acquisti/SOUPS2018/Study1-2-3.pdf>.

3.2 Results and Discussion

Even though our pilot tests confirmed that the two stock-model images we used in creating the morphs were perceived to be equally attractive, in this study the morphs created with these images were not perceived to be equivalent. One was hired more often than the

other (60% vs. 40%; $p = 0.029$). Still, we examined the percent of participants who chose to hire the two different instructors (self vs. other) and found no difference in proportions: 49% chose to hire the self-morph whereas 51% for the other ($p = 0.84$).

Whether people had taken lessons in the instrument before had a significant effect on their decision to hire their own self-morph. Thirty-one participants had taken lessons to learn the musical instrument before, and they were significantly more likely to hire the self-morph (67.74%, $p = 0.048$). However, this result does not hold after we account for multiple comparisons. Eighty participants had never taken lessons to learn the musical instrument before and there were no significant differences there (45%, $p = 0.3173$). One person reported to be currently taking classes to learn the musical instrument. Paired t-tests on participants' ratings of self vs. other morph for trustworthiness, attractiveness, knowledgeable, liking, identifying with self and similarity to self were not statistically significant, as detailed in Table 1. Although the measures showed high internal reliability (Cronbach's alpha = 0.827) overall mean judgments were also not statistically different between the conditions (see Table 1). There were also no significant differences between the percent of participants who found self and other morphs familiar (9% vs. 6.3%) or between the percent of participants who found self and other morphs unusual looking (24% vs. 25%), $p > 0.9$. We ran mixed model analyses of NPI on all DVs. None of the interactions was significant ($p > 0.14$ before correcting for multiple comparisons, $p > 0.7$ after). (Descriptive statistics are presented in Table 1.)

To summarize, we did not find evidence for an effect of a self-morph on individuals' attitudes, intentions or judgments. Given the sampling constraints, the study was fairly powered, with more than 50 participants in each cell yielding a power of about 70% to detect a 0.4 effect size in a two-sided test (or 82% for a one-sided test). Even though the task was novel to most of the participants, and preferences between the two options could (or should) have only depended on the physical appearance of the two prospective service providers, having one of them include subtle aspects of the participant's own face did not seem to impact their decisions.

4. STUDY 2

In our next study, we decided to shift our focus to a different domain, and employ a different type of dependent measure – namely, self-disclosure. We opted for this choice for several reasons. First, in the current era of explosion of information on social networking websites and proliferation of personal information being harvested by online companies, users' propensity to disclose personal information is an important privacy issue [1]. The second reason is that previous research has reportedly found an effect of self-morphs on level of trust [12] so a logical extension of this effect might be that individuals will be more willing to disclose personal information to someone they trust more. Lastly, we thought that perhaps self-disclosure could be a more indirect way of measuring reactions to a self-morph.

In Study 2, we used a method for eliciting self-disclosure that relies on asking participants to respond to sensitive and personal questions, and that had been used successfully in previous self-disclosure studies (e.g., [22]). Participants were asked to imagine they are talking to a therapist who asks them several questions about themselves. The therapist was either a self-morph created by using a stock model's face and the participant's face, or a morph of the same stock model and an unfamiliar person in a between-participants design. We predicted that participants would be more likely to divulge personal, sensitive information when the

therapist's image was a self-morph. In this study, we used female participants only to expand our inquiry beyond males.

4.1 Method

Based on the null results of Study 1, we adjusted our estimated effect size to $d = 0.25$ and aimed at obtaining a larger sample. We calculated that a sample of about 300 would yield a 70% power for a two-sided test, and about 82% for a one-sided test. We thus recruited 310 Caucasian female participants ($M_{age} = 30.76$, $SD = 8.7$) from MTurk who completed the study for a payment of \$1.5 each. Participants were asked to imagine that they are looking for a therapist to discuss something going on in their life and they are referred to a specific therapist whose image is displayed. The image of the therapist was, for half of the participants, a self-morph of their own picture with a female stock model's picture (see Figure 2) and for the other half a morph of one of the other participants' image with the same female model's picture.



Figure 2. Example stimuli used in Study 2.

Participants were asked to imagine that during their meeting with the therapist the therapist asks them several questions about themselves. They could choose to answer these questions or indicate that they would prefer not to answer on a per question basis. The questions referred to engaging in unethical or socially undesirable behaviors that have been used in previous research about online self-disclosure [18]. Participants were asked to indicate, on a scale from 1 (never) to 5 (frequently) have they ever: *Had sex with the current husband, wife, or partner of a friend? Masturbated at work or in a public restroom? Had a fantasy of doing something terrible (e.g., torturing) to someone? Fantasized about having violent non-consensual sex with someone? While an adult, had sexual desires for a minor? Neglected to tell a partner about a sexually transmitted disease from which you were suffering? Had sex with someone who was too drunk to know what they were doing? Stolen anything that did not belong to you? Tried to gain access to someone else's (e.g., a partner, friend, or colleague's) email account? Looked at pornographic material?* Participants could also mark "prefer not to answer" for any of the questions.

Next, participants rated the therapist on how attractive, trustworthy, and knowledgeable they thought she was, how much they liked her, how good they thought she was at her job, how similar to themselves they thought she was and how much they identified with her. Then, we asked participants to rate how intrusive they found the questions asked by the therapist and whether she looked familiar or unusual to them. Participants then completed the NPI scale [3], and entered their demographics. They also indicated whether they found anything strange or unusual, or familiar, in either of the images, and if they said they did, then we asked them to elaborate further. Participants were debriefed as in the previous study.

4.2 Results and Discussion

We examined whether participants disclosed more to the self vs. the control-morph by examining participants *Active Affirmative Responses* (or AARs) that are the instances when participants indicated that they engage in the listed unethical or socially undesirable actions irrespective of the frequency with which they reported engaging in them [18]. In other words, AARs measure the amount of times participants indicated a response that was not “never” or “prefer not to say” to the listed unethical behaviors. Comparing AARs between conditions, we found no statistically significant differences between self vs. other conditions ($M = 2.91$ vs. 3.08 , $SD = 1.63, 1.59$, $t(313) = 0.944$, $p = 0.346$, Cohen’s $d = 0.11$). Similarly, there were no statistically significant differences in perceived ratings (see Table 1; Cronbach’s alpha for the measures was 0.889). The NPI scale showed a high internal reliability (Cronbach’s alpha = 0.746). Thus, we averaged the items

effect and NPI, $p > 0.37$. (Descriptive statistics are presented in Table 1.)

To summarize, it appears that in the domain of self-disclosure as well we could not find evidence for the self-morph effect, as a self-morph did not seem to lead participants to disclose more personal information when compared to a non-self-morph. This study was highly powered. Thus, we feel more confident that this null finding does not represent a sampling problem.

One remaining difference we could see between our studies and previously published ones was the fact that we obtained participants’ images from real-life services (their profiles on online social networks), whereas the previous researchers either took participants’ pictures at the beginning of the study [12,13] or asked participants to submit a high-resolution image of themselves [4]. Therefore, previous research had the advantage of high quality

Table 1. Comparisons between “self” and “other” condition on all measures in Studies 1-3.

DV	<u>Mean (SD)</u>						<u>t (p)</u>			<u>Cohen's d</u>		
	Study 1		Study 2		Study 3		Study			Study		
	Other	Self	Other	Self	Other	Self	1	2	3	1	2	3
Trustworthy	5.07 (1)	5.01 (1.1)	5.29 (1.1)	5.25 (1.2)	4.46 (1.2)	4.59 (1.2)	-0.47 (0.64)	0.3 (0.76)	1.202 (0.23)	-0.09	0.03	0.11
Attractive	5.44 (1)	5.35 (1.1)	5.27 (1.2)	5.28 (1.1)	4.16 (1.3)	4.22 (1.4)	-0.93 (0.36)	-0.1 (0.92)	0.457 (0.65)	-0.17	-0.01	0.04
Knowledgeable	4.32 (1.3)	4.55 (1.2)	4.81 (1.2)	4.69 (1.1)	4.40 (1.0)	4.46 (1)	1.74 (0.09)	0.89 (0.37)	0.65 (0.52)	0.32	0.10	0.06
Like	4.81 (1)	4.70 (1)	5.28 (1.2)	5.11 (1.1)	3.96 (1.2)	4.17 (1.3)	-1.12 (0.27)	1.31 (0.19)	1.972 (0.05)	-0.21	0.15	0.18
Identify	4.12 (1.3)	4.14 (1.3)	4.10 (1.4)	4.10 (1.4)	3.16 (1.4)	3.26 (1.5)	0.18 (0.86)	-0.05 (0.96)	0.804 (0.42)	0.03	-0.01	0.07
Similar	4.28 (1.2)	4.19 (1.3)	4.20 (1.3)	4.10 (1.4)	3.41 (1.5)	3.51 (1.4)	-0.65 (0.52)	0.61 (0.54)	0.835 (0.41)	-0.12	0.07	0.08
Good (Study 2 only)			5.22 (1.2)	5.04 (1.3)				1.29 (0.20)			0.15	
Overall judgments	4.66 (0.8)	4.67 (0.8)	4.88 (0.9)	4.80 (0.9)	3.92 (1.0)	4.03 (1.0)	-0.21 (0.83)	0.77 (0.44)	1.279 (0.2)	-0.04	0.09	0.12

to compute an overall NPI score for each participant and then used that average measure to examine whether an effect of the conditions could be different for different levels of NPI. We found a significant effect of NPI on self-disclosure (AARs): the higher the NPI score the more participants disclosed ($beta = 0.163$, $SE = 0.642$, $p = 0.03$). However, the effect of self vs. other morph was not statistically significant nor was the interaction between this

pictures that could ensure high quality morphs, thereby reducing a possible source of noise relative to our experimental design (but with the disadvantage of potentially adding demand effects). In our experiments, pictures were typically of lower resolution and poorer illumination than photos captured in a lab; furthermore, profile photos included several different poses and expressions, whereas

photos taken in the lab were always taken frontally and with neutral expressions.

5. STUDY 3

Thus, for our third study, we decided to employ a design similar to the previous researchers' by having participants come to a lab where their pictures would be taken and used to create self-morphs. This allowed us to ensure that our participants' images were of high resolution and well illuminated, with consistent expressions and poses across images, thus maximizing the possibility of detecting the self-morph's effect if such one truly exists. Also, we increased our sample size considerably, aiming to get at least 200 respondents in each condition, that would ensure a minimum of 80% power to detect a $d = 0.25$ effect size. We preceded this study with other pilot studies that pre-tested the stimuli and questions used in this research. Lastly, Study 3 only measured trustworthiness (a measure that had produced significant results in prior research [4]) without involving additional measures of hiring intentions, self-disclosure, or others. At this point, we predicted that a self-morph would *not* be judged as more trustworthy compared to a control condition's morph. Participants were recruited either using an online participants pool at our university, which included both students and non-students, and also using a mobile "Data Truck" that was parked at several common intersections during rush hours of a large U.S. city. Participants were invited for a "study about images" that took approximately 15-20 minutes and were paid \$10 for their participation. The sample included 495 Caucasian participants, 250 of them males, with an average age of 30.92 years ($SD = 14.57$). Study 3 took place in a lab. As participants arrived to the lab, the experimenter explained that this was a study about images and in order to take part in the study we need to take their picture, which may be used in future studies for future participants. One experimenter took a picture of the participant and uploaded it to a shared folder, while another experimenter sat in an adjacent room and prepared the morph by accessing the shared file, so participants could not see that their picture was actually being used at that time.

Participants were then seated at a computer and asked to complete an "Image task." In this task they were shown three images and asked to describe, in an open-ended manner, their thoughts and feelings about what they saw in the images. The three images were a scenery picture, a picture of several team members working together and a stock photo of a person. Next, participants completed a "Video task" that involved viewing a short video and answering some questions about it. The purpose of these tasks was both to give the experimenter time to create and insert the morph into the survey and to convince participants that the study was about image perceptions.

The third task was also called an "Image task" in which participants were asked to look at the picture of a person. This image was either a self-morph, created by morphing the participant's own picture with a stock model's face or the morph of another participant with the same stock model's face (at a 40:60 percent ratio, as in [4]). The randomized assignment was done in the following way: the first participant in a session was assigned to the "self" condition, and the next participant received the same morph as the first participant did (putting the second participant in the control or "other" condition), and so on for the following participants.

After viewing the person's image, participants were asked to rate how trustworthy they thought that person was from 1 (not at all) to 7 (very much). On the next page, participants rated how attractive and knowledgeable they thought the person was, how much they liked him/her, how similar they thought he/she was to them, and

how strongly they identified with the person. Then, participants completed the Rosenberg's Self-Esteem Scale [26] followed by five questions about their own appearances by indicating how much they agree (from 1 – strongly disagree to 5 – strongly agree) with each of the following statements: *I think I am more attractive than the average person of my age; All in all, I like the way I look; I typically dislike my own pictures; I am very critical about my own looks; I like being photographed; I do not like some of my facial features.* Then, participants were asked if they found anything unusual or odd about the person whose image they just saw or if they thought this person looked familiar. Then they provided their demographics and were debriefed as in previous studies.

5.1 Replication Notes

As noted earlier, Study 3 was the closest to a replication attempt of [4]. We deviated from [4] in the following ways and for the following reasons: a) participants' pictures were shot on site (and not delivered ahead of time) to ensure high quality and standards; b) we only used a between-subjects design, as our pilot study showed no advantage to a within-subject design; c) we focused on a simple trustworthiness dependent variable, as we had already captured other variables in previous studies (and pilots), whereas for this final study we aimed at testing a straightforward and broadly applicable metrics of face-morph's impact (trustworthiness is one of original and most common metrics in self-morphs studies: see, e.g., [12] and [28]).

5.2 Results and Discussion

As detailed in Table 1, no statistically significant differences were found between the conditions on the ratings ($p > 0.05$, except for liking, $p = 0.049$) or the overall judgment score (Cronbach's alpha = 0.85).

We then examined whether self-esteem, or liking of personal appearances, could help detect an effect of the self-morph. The RSES showed high reliability (0.882), and so did the questions of "self-looks" (0.734). Thus, we averaged the questions to form two composite scores: self-esteem and self-looks. We then conducted a regression analysis on the overall judgments score with condition, self-esteem, self-looks, the interactions of condition with self-esteem and self-looks, as well as gender and age as independent variables. We found that self-looks significantly predicted overall judgments, $\beta = .73$, $SE = .35$, $.07$, $t = 2.09$, $p = .037$. However, the condition variable (self vs. other morph) was not statistically significant, nor were any of the interactions, $p > .12$. Even when we excluded participants who reported seeing something unusual or familiar in the morphed picture, there was no significant effect of condition or the interaction of condition with self-looks and self-esteem on overall judgments, $p > .25$. In all our analyses, we could not find any support for a significant effect of the self-morph on people's judgments.

6. DISCUSSION

Previous lab research has suggested that people evaluate self-morphs differently than they evaluate face-morphs of unfamiliar people, and that self-morphs are judged as more trustworthy and attractive [12,13]. In the real world, self-morphs may be created using, for example, people's photos on social networks, and then employed to covertly influence consumers and individuals in a form of highly personalized "visceral" targeting - thus raising potential yet significant privacy concerns. Whether they might still exert influence on people's attitudes, however, was an important and open question that warranted direct research. In this paper, we examined the potential effect of self-morphs on people's online

behavior, only to realize that we could not find evidence of an effect: if such an effect does exist, it could not be captured in our studies under a variety of different settings. In the online studies that tried to find the basic self-morph effect using participants' images shared on a social networking site, as well as a highly powered third lab study, we could not replicate the effects of self-morphs that were reported in the past: we found no evidence that self-morphs impact judgments or choices regarding the purchase of products or services.

As is the case with any null result, there may be various reasons why we did not discover an effect of self-morphs in our studies that do not necessarily disprove the existence of an effect. Although most of our samples (especially Study 3) were relatively large, a bigger sample could have provided the ability to test whether the effect might still occur under some specific moderating conditions that could have explained the discrepancy between our results to previous studies. Our results may have also been due to other factors that pertain to the design and procedure of the studies. For example, while Study 2 focused on actual disclosure behavior, Studies 1 and 3 used hypothetical measures of attitudes, judgments, and behavioral intentions. It is possible that self-morphs may not affect attitudes and intentions, but could still influence people's behavior in an implicit and covert manner. Indeed, past research has shown self-morphs to affect outcomes of trust games, for one [13]. However, we still expected self-morphs to show the effect, if it does exist. Furthermore, the fact that we could not even replicate the effect on the same measure – trustworthiness – that was used in prior studies (e.g., [12]), should be regarded as problematic as well. Restricting (due to technical limitation of the morphing process) Study 1 to males and Study 2 to females might have also played a role, although usually making the sample more homogenous should increase, rather than decrease, statistical power (and Study 3 used both genders). An additional possible concern is that survey-based scales may have low fidelity in measurement, making it harder to detect small effects. However, the vast majority of morph studies, to our knowledge, also used survey answers as their main dependent variables. Across our experiments, while Study 3 ended up using survey scales similar to those employed in previous morph research, Study 1 actually leveraged scales from a different stream of literature (privacy and self-disclosure research) and Study 1 used a behavioral intention dependent variable. Finally, the morphing procedure may also play a significant role in their likelihood of affecting consumers. Images in Study 1 and Study 2 came from social media profiles; thus, the quality of resulting morphed images may have been different from morphs based on photos taken under controlled conditions in a lab. However, and importantly, quality of images was *not* different between conditions; furthermore, lower photo quality may not necessarily mean lower effect size, and Study 3 did use lab photos (as in comparable prior studies). As noted, we closely followed the methods used by [4]. We also contacted that research team and verified that we are indeed following the same procedure. In fact, we followed previous studies to extent made possible by published information and details shared with us both in terms of experimental design and morphing technique, and deviated from those in narrow details for hard-thought reasons.

While no single study was an exact replication (Study 3, for instance, focused on a simple trustworthiness metrics rather than on voting intentions), if the effects of self-morphs disappear even with relatively minor design changes, this does suggest that the effects of self-morphs on individuals' behavior may not be robust. One of our contributions therefore is that influence of face-morphs may be restricted to stringent lab conditions: while we cannot refute

whether it has *internal* validity, we show it may not have significant *ecological* validity.

Future research may endeavor to chase the effect of self-morphs on individuals' perceptions and behavior - it is of course possible that, in the future, other and perhaps more sophisticated and advanced morphing procedures may overcome the limitations of our studies and discover that self-morphs can be effective at influencing individuals' judgment and behavior. More broadly, it is also possible that, outside the realm of face-morphs, other types of personal information (such as an individual's preferences for certain colors or sounds) may be used in covert manners to target and personalize messages, invitations, or suggestions. If such a visceral effect does exist, and if online firms were able to take advantage of these technologies to collect consumer information and use it to subtly and nearly undetectably target messages to influence people's behavior, it would raise important theoretical, practical, and legal issues. Policy makers would then have to consider whether current online safeguards meant to protect individuals' privacy and autonomy need to be re-evaluated in order to prevent covert third parties from exerting undue influence in such forms. Firms, on the other hand, may have to consider whether or not to engage in such strategies, given their ethical and legal implications. A broader implication arising from this manuscript, therefore, is to highlight how, due to the vast self-dissemination of personal information, public yet personal data might be used in interactions with individuals by both services and independent third parties surreptitiously and covertly – that is, without the individuals' awareness of the data being used, and of the effect it may have on their decision making.

7. ACKNOWLEDGMENTS

This work was supported by the NSF Grant Award Number 1012763 (Nudging Users Towards Privacy) and NSF Grant Award Number 1327992 (Societal, Economic, Technological, and Legal Implications of Personalized Face Composites). The authors thank the reviewers for the insightful comments, and Mr. Jeffrey Flagg for excellent research assistance.

REFERENCES

1. Acquisti, Alessandro, Brandimarte, Laura, and Loewenstein, George (2015), "Privacy and Human Behavior in The Age Of Information," *Science*, 347(6221), 509-514.
2. Acquisti, Alessandro and Gross, Ralph (2006), "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook," In *Privacy Enhancing Technologies: Lecture Notes in Computer Science*, 4258, 36-58.
3. Ames, Daniel R., Rose, Paul and Anderson, Cameron. P. (2006), "The NPI-16 as A Short Measure of Narcissism," *Journal of Research in Personality*, 40(4), 440-450.
4. Bailenson, Jeremy N., Iyengar, Shanto, Yee, Nick and Collins, Nathan A. (2008), "Facial Similarity Between Voters and Candidates Causes Influence," *Public Opinion Quarterly*, 72(5), 935-961.
5. Beales, Howard. (2010), "The Value of Behavioral Targeting," *Network Advertising Initiative*, 1.
6. Bressan, Paola and Zucchi, Guendalina (2009), "Human Kin Recognition is Self- Rather Than Family-Referential," *Biology Letters*, 5(3), 336-338.
7. Calo, Ryan (2014), "Digital Market Manipulation," *George Washington Law Review*, 82(4), 995-1051.
8. Chen, Pei-Yu, Wu, Shin-Yi, and Yoon, Jungsun (2004), "The Impact of Online Recommendations and Consumer Feedback on Sales," *ICIS 2004 Proceedings*, 58.

9. Chevalier, Judith A., and Mayzlin, Dina (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345-354.
10. Cohen, Julie E. (2012), "What is Privacy For," *Harvard Law Review*, 126, 1904.
11. Critchley, Hugo, Daly, Eileen, Phillips, Mary, Brammer, Michael, Bullmore, Edward, Williams, Steven, Van Amelsvoort, Therese, Robertson, Dene, David, Anthony and Murphy, Declan (2000), "Explicit and Implicit Neural Mechanisms For Processing of Social Information From Facial Expressions: A Functional Magnetic Resonance Imaging Study," *Human Brain Mapping*, 9(2), 93-105.
12. DeBruine, Lisa M. (2002), "Facial Resemblance Enhances Trust," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1498), 1307-1312.
13. DeBruine, Lisa M. (2004), "Facial Resemblance Increases the Attractiveness of Same-Sex Faces More Than Other-Sex Faces," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1552), 2085-2090.
14. Engell, Andrew D., Haxby, James V. and Todorov, Alexander (2007), "Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in Human Amygdala," *Journal of Cognitive Neuroscience*, 19(9), 1508-1519.
15. Farahat, Ayman and Bailey, Michael (2012), "How Effective is Targeted Advertising?," *Proceedings of the 21st International Conference on World Wide Web, ACM*, 111-120.
16. Federal Trade Commission (2009), "Self-Regulatory Principles for Online Behavioral Advertising," *Federal Trade Commission Staff Report*, Washington DC, Feb 12.
17. Greenwald, Anthony G. (1975), "Consequences of Prejudice Against The Null Hypothesis," *Psychological Bulletin*, 82(1), 1-20.
18. John, Leslie, Acquisti, Alessandro and Loewenstein, George (2009), "Strangers on a Plane: Context-Dependent Willingness to Divulge Personal Information," *Journal of Consumer Research*, 37(5), 858-873.
19. Kanwisher, Nancy and Yovel, Galit (2006), "The Fusiform Face Area: A Cortical Region Specialized for the Perception of Faces," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109-2128.
20. Kramer, Adam D., Guillory, Jamie E. and Hancock, Jeffrey T. (2014), "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks," *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
21. McDonald, Aleecia and Cranor, Lorrie (2010), "Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising."
22. Moon, Youngme (2000), "Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers," *Journal of Consumer Research*, 26(4), 323-339.
23. Open Science Collaboration (2015), "Estimating the Reproducibility of Psychological Science," *Science*, 349(6251), aac4716.
24. Platek, Steven M. and Kemp, Shelly M. (2009), "Is Family Special To The Brain? An Event-Related fMRI Study of Familiar, Familial, and Self-Face Recognition," *Neuropsychologia*, 47(3), 849-858.
25. Purcell, Kristen, Brenner, Joanna, and Rainie, Lee (2012), "Search Engine Use 2012," *The Pew Research Center's Internet & American Life Project*, Washington DC.
26. Rosenberg, Morris (1965), *Society and the Adolescent Self-Image*, Princeton, NJ: Princeton University Press.
27. Samat, Sonam, Acquisti, Alessandro and Babcock, Linda (2017), "Raise the Curtains: The Effect of Awareness About Targeting on Consumer Attitudes and Purchase Intentions," In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)* (pp. 299-319). USENIX Association.
28. Tanner, Robin J. and Maeng, Ahreum (2012), "A Tiger and a President: Imperceptible Celebrity Facial Cues Influence Trust and Preference," *Journal of Consumer Research*, 39(4), 769-783.
29. Turow, Joseph, King, Jennifer, Hoofnagle, Chris, Bleakley, Amy, and Hennessy, Michael (2009), "Americans Reject Tailored Advertising and Three Activities that Enable It," In *SSRN: <http://ssrn.com>*, Vol. 1478214.
30. Verosky, Sara C. and Todorov, Alexander (2010), "Generalization of Affective Learning About Faces to Perceptually Similar Faces," *Psychological Science*, 21(6), 779-785.
31. Winston, Joel S., Strange, Bryan A., O'Doherty, John and Dolan, Raymond J. (2002), "Automatic and Intentional Brain Responses During Evaluation of Trustworthiness Of Faces," *Nature Neuroscience*, 5(3), 277-283.
32. Yan, Jun, Liu, Ning, Wang, Gang, Zhang, Wen, Jiang, Yun, and Chen, Zheng (2009), "How Much Can Behavioral Targeting Help Online Advertising?" In *Proceedings of the 18th International Conference on World Wide Web* (pp. 261-270). ACM.