
**The Inauthentic Online Self:
Perceptions of Naturalness Drive Judgments of Authenticity**

Matthias Uhl^a and Joshua Knobe^b

^a Faculty of Computer Science, Technische Hochschule Ingolstadt

^b Program in Cognitive Science and Department of Philosophy, Yale University

Running head:	The Inauthentic Online Self
Institution address:	Technische Hochschule Ingolstadt Esplanade 10 85049 Ingolstadt
	2 Hillhouse Avenue Yale University New Haven, CT, USA 06511
Email addresses:	Matthias Uhl (corresponding author) matthias.uhl@thi.de
	Joshua Knobe Joshua.knobe@yale.edu

Abstract

People sometimes behave differently depending on whether they are interacting online (by email, social media, etc.) vs. interacting in person. Four studies test the hypothesis that when an agent's behavior is different online vs. in person, people think that the online behavior is less reflective of who the agent truly is deep down. Study 1 found that the very same behavior is regarded as less reflective of the true self when it is performed online. Study 2 showed that this effect is not merely a matter of perceived impression management. Study 3 found that there is a general tendency such that behavior is seen as more reflective of the true self when it is performed in an environment regarded as "natural." In Study 4, a manipulation that led participants to see online behavior as more natural had a downstream effect on the degree to which this behavior was seen as reflecting the true self. Taken together, these results suggest that people's judgments about online behavior are not simply a reflection of idiosyncratic facts about the online sphere in particular but are instead driven in part by a far more general psychological process involving perceptions of naturalness.

Keywords: authenticity; naturalness; online behavior; online persona; social media; true self

1. Introduction

Imagine a person who behaves very differently when she is interacting with others online than when she is interacting with others in person. When you discuss something with her using email, social media, or text messages, she behaves one way, but when you actually meet her in person, she behaves in a different way. Which of these kinds of behavior would you think most reveals the person she truly is?

In this paper, we explore the possibility that people show a systematic tendency to see in-person behavior as more reflective of a person's self. Thus, if a professor is curt and businesslike when you interact with her by email but is always warm and friendly when you interact with her in person, you will tend to think that she truly is a warm and friendly person.

If this effect does emerge, we face a question about how to explain it. One obvious hypothesis would be that it is just a matter of people seeing online behavior as especially driven by impression-management concerns. Perhaps online behavior is seen as part of an attempt to put together a highly curated image designed to achieve certain social goals, while in-person behavior is seen as less driven by an attempt to make a particular sort of impression.

Our aim here is to explore the possibility that there is actually something deeper at work in this phenomenon. The core hypothesis is that people see interacting with others in person as the *natural* environment for human beings. People then show a general tendency to think that the things a person does in the natural environment are more reflective of who that person truly is.

1.1 The impression management hypothesis

If people do show a tendency to think that online behavior is less reflective of who a person really is, one possible explanation would be that this effect arises simply because people think that online behavior is more shaped by impression management concerns. For example, suppose that an agent creates social media posts that make herself seem happy and confident. In such a case, people might think that her social media posts fail to reflect her true self in a very straightforward sense. Specifically, they might think that she is not in fact happy or confident and that she is simply trying to make other people believe that she feels this way.

Existing research provides at least some support for this first way of conceptualizing the difference between online and in-person behavior. In particular, it has been argued that impression management plays an especially large role in people's behavior in the online sphere. Webpages are characterized by absent verbal or facial cues and give their developers almost full discretion about the image that they want to convey (Papacharissi, 2002). Similarly, performance on social media, although more complex (Marwick & Boyd, 2011), may be used for self-marketing and personal branding (Picone, 2015). The emergence of social media platforms like LinkedIn has facilitated identity construction through the possibility of actively shaping an individual's profile through the

selective visual and textual information provided to get professional or personal acclaim (Paliszkiewicz & Mądra-Sawicka, 2016). Drawing on this evidence, theorists have developed the concept of “online identity” (Marwick, 2013) or “online persona” (Moore et al., 2017). The key claim is that people establish a separate persona in the online sphere about which they have tighter control than about their “real world” identity which may be limited by a lack of self-esteem or other actual or imaginary shortcomings. Perhaps it is precisely for this reason that online behavior is seen as less reflective of the true self.

Yet, although this view has a certain plausibility, there is at least some reason to suspect that there might be more to the story. The reason is that the difference between online and in-person behavior may exist even in cases in which there is no greater amount of impression management in the online environment. For example, when people are interacting online through private emails or texts, one might think that their online behavior is not any more shaped by impression management than their in-person behavior is, but it is at least possible that the online behavior would still be seen as less reflective of the true self. For this reason, it might be helpful to consider alternative ways of understanding this phenomenon.

1.2. The naturalness hypothesis

One intriguing hypothesis would be that this effect arises as a result of people's intuitive notion of a *natural environment for humans*. The hypothesis would be that people think that interacting online is not a natural environment for humans, and that it is precisely for this reason that they think behaviors performed within this environment are not reflective of the true self.

To get an initial sense of how people might think about the “natural environments for humans,” it might be helpful to consider how people think about environments for other species. Take horses. People might think that the majority of horses now live in stables, but at the same time, they might think that living in a stable is not the natural environment for a horse. Instead, they might think that the natural environment for a horse is to roam free on a prairie. Then people might attach some special significance to the way that horses behave when they are in their natural environment. For example, if they see that horses always behave in a particular way when roaming free on a prairie, they might feel that this behavior reflects something very fundamental about the nature of horses.

The hypothesis we will be exploring here is that people apply that very same approach to thinking about human beings. People might think that many humans these days live in cities, drive in cars, and communicate with each other using phones or computers, but they might think that this is not the natural environment for humans. In particular, people may think that the natural environment for humans is not one that involves communicating with others online (by email, texts, etc.) but is instead one that involves interacting with others in person. They might then attach a special significance to how people behave when they are in this environment.

Although people might generally think that it is more natural for humans to live in an environment with less technology, we hypothesize that the ordinary notion of a natural environment

is not necessarily determined by its given level of technology. For example, independent of anything about technology, people might think that it is natural for children to grow up in a particular sort of familial arrangement (e.g., being raised by their own parents). As a result, it might even happen that certain technological advances make the environment *more* natural for humans. Suppose that technological advances allow you to do your banking from home, surrounded by your family, instead of going to a bank branch. People might think that your environment after these advances is more of a natural environment for human beings than it was before these advances.

A question now arises about the downstream effects on people's cognition of thinking that a particular environment is more natural. Within existing research, there has been a lot of evidence for the claim that things that are believed to be natural are seen as *good*. A natural-is-better bias and the belief that natural is safer is supported in empirical research for a wide variety of products (Meier et al., 2019a). It has been observed that the frequency with which the label "natural" is used for products and services from soda to cigarettes already suggests that it delivers a sales advantage. Participants rate identical products as healthier and lower in calories when they are labeled as natural (Skubisz, 2017). The preference for natural foods was found already in children as young as five years old (Wilks & Bloom, 2022). Adults consider cultured meat, for instance, as "unnatural" which seems to induce affective feelings like disgust and fear rather than analytical processes (Wilks et al., 2021). Novel food technologies like genetic modification meet strong resistance from consumers who see it as inherently positive if a food is produced with a minimum of obvious human intervention (Siegrist & Hartmann, 2020). Similarly, people preferred natural over synthetic drugs for minor and serious hypothetical medical conditions (Meier & Lappas, 2016). The naturalness bias for drugs was found across cultures (Ji et al., 2023) and replicated in a behavioral choice experiment (Meier et al., 2019b). It was recently even identified for physicians (Lappas et al., 2023).

The hypothesis we explore here is an importantly different one. We are not asking whether people think that there is something good about being in a natural environment but rather whether people think that being in a natural environment reveals the *true self*. Consider what it would be like to be on a battlefield, locked in combat with an enemy, and armed only with a sword and shield. People might think that this environment would be a horrible one, but at the same time, they might think that it is a natural environment for humans, and if they do, the hypothesis predicts that they will see what you do in this environment as especially reflective of your true self.

If this claim does turn out to be correct, it would contribute to the larger literature on the factors that influence judgments about whether a given behavior reflects an agent's true self (Christy et al., 2019; De Freitas, J., & Cikara 2018; Heiphetz et al., 2018; Strohminger et al., 2017). Existing studies suggest that morally good behaviors are seen as more reflective of the true self (Newman et al., 2014, 2015) and that this is also true for behaviors driven by emotion or instinct (Oktar & Lombrozo, 2022). The claim here would be that behaviors performed in a natural environment are also seen as more reflective of the true self and that it is this effect that drives the difference in perception between online and in person behavior.

1.3 Present studies

Four studies explored intuitions about online vs. in-person interaction and its relationship to the notion of a natural environment for human beings. Study 1 tested for a basic effect such that participants see behavior as less reflective of the true self when it is performed online than when it is performed in person. Study 2 asked whether this effect can be explained simply in terms of impression management. Study 3 looked at whether there is a general effect such that behaviors are seen as more reflective of the true self when they are performed in environments that are natural for human beings. Finally, in Study 4, we manipulated the degree to which interacting with others online was seen as a natural environment for human beings and checked for a downstream effect on judgments about whether behaviors performed online reflect the true self. The code and data files for the project are available at

https://osf.io/uqp5m/?view_only=0dc5588b6ba241b29f22923fffd9ec33.

2. Study 1

In the first study, we tested whether given kinds of behavior are thought to be less reflective of who a person really is deep down if people exhibit them online as opposed to in person. To this aim, we presented participants with vignettes that described a divergence in a fictitious person's way of behavior between an online environment and in person.

Method

All methods and analyses for this study were pre-registered (<https://aspre-dicted.org/v9dh3.pdf>).

Participants. Participants were recruited using Prime Panels from CloudResearch. We received 879 complete responses. Participants who failed the comprehension check ($n = 125$) were excluded. Analyses were conducted on the remaining 754 participants ($M_{\text{age}} = 37.84$, $SD_{\text{age}} = 13.75$; 42.8% men, 56.6% women, <1% other gender or prefer not to say; 95.0% stated English as first language; 4.4% stated otherwise, <1% prefer not to say).

Procedure and Measures. Participants were randomly assigned to one condition in a 2 (online vs. in person) \times 7 (behavior) \times 2 (order) design. Each participant received a vignette about an agent who tends to perform one behavior when interacting online but the exact opposite behavior when interacting in person.

For example, in the self-confident behavior condition, participants were told about an agent whose behavior is sometimes self-confident (vs. timid). In the condition where this behavior occurs online, the vignette was:

When Susan is interacting with people online, she tends to act really self-confidently. On social media, she regularly tells others what is going on in her life. In emails, she clearly expresses her opinions irrespective of whether the recipient might like them or not. When she catches her friends online saying or doing things that she does not like, she openly criticizes this.

The Inauthentic Online Self

When Susan is interacting with people in person, she tends to act really timidly. On the street, she is very reserved about her life. In personal conversations, she often hides her opinions if she thinks that her conversation partner might not like them. When she personally catches her friends saying or doing things that she does not like, she ignores this.

In the condition where this behavior occurs in person, the vignette was:

When Susan is interacting with people in person, she tends to act really self-confidently. On the street, she regularly tells others what is going on in her life. In personal conversations, she clearly expresses her opinions irrespective of whether her conversation partner might like them or not. When she personally catches her friends saying or doing things that she does not like, she openly criticizes this.

When Susan is interacting with people online, she tends to act really timidly. On social media, she is very reserved about her life. In emails, she often hides her opinions if she thinks that the recipient might not like them. When she catches her friends online saying or doing things that she does not like, she ignores this.

In the other behavior conditions, participants were told about an agent whose behavior is brave (vs. cowardly), hostile to foreigners (vs. open-minded with foreigners), romantic (vs. cynical), reclusive (vs. social), humble (vs. arrogant) and mean (vs. friendly). Within each condition, the order of presentation of the two paragraphs of the vignette was counterbalanced.

Participants then rated their agreement with a statement about whether the behavior reflected the agent's true self. For example, in the self-confident behavior conditions, the statement was "Susan sometimes acts self-confident and sometimes acts timid. But ultimately, deep down, Susan is self-confident" (1 = "fully disagree", 7 = "fully agree").

A comprehension check question appeared on a separate page (giving participants no opportunity to go back). Participants were asked: "In what ways does [agent] sometimes act?" Below the question, four different numbered answers were given of which only one was correct.

Results

Results are displayed in Figure 1. Analyses for all studies in this paper were conducted in R using lme4 and lmerTest. In this study, the data were analyzed using linear mixed effects models, with true self judgment as the dependent variable, environment (online vs. in person) and order as fixed effects, and vignette as a random effect (with random slopes for environment).

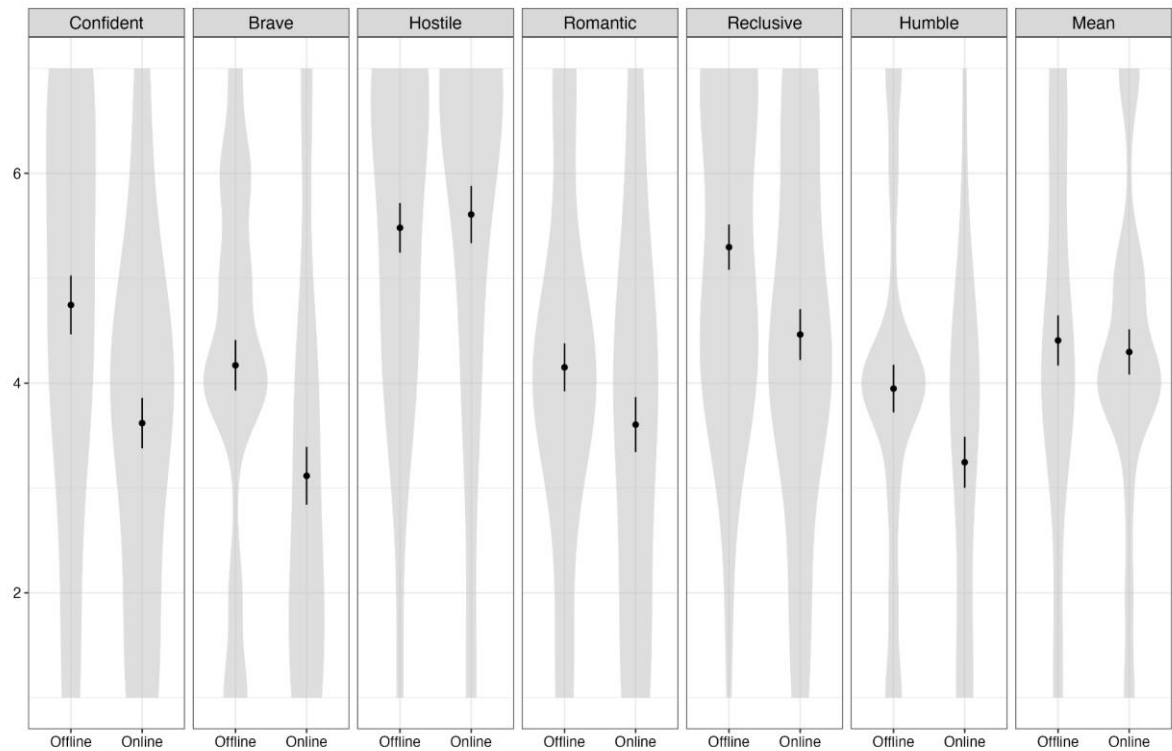


Figure 1. Violin plot showing true self judgments by environment and behavior in Study 1. Error bars show standard error.

There was a significant effect of environment, such that behavior was seen as more reflective of the true self when it was performed in person ($M = 4.58$, $SD = 1.84$) than when it was performed online ($M = 3.99$, $SD = 1.97$), $b = -.60$, $SE = .17$, $t = -3.6$, $p < .01$. There was no significant effect of order.

Discussion

Behavior was seen as less reflective of who a person really is when performed in the online sphere than when performed in person across a variety of different pairs of behavior. The results therefore provide empirical support for the hypothesis that behavior exhibited online is considered less revealing of who an agent really is, deep down, than behavior exhibited in person. The question that remains open is what drives this result.

3. Study 2

One might suspect that participants in Study 1 saw online behavior as less reflective of the true self merely because they assumed that when being online an individual would be more likely to manage others' impressions of its identity. On social media, for instance, people might select certain pictures that are less reflecting of their real life but rather idealistic representations of who these people long to be. Study 2 was set up to test this explanation.

Method

All methods and analyses for this study were pre-registered (<https://aspre-dicted.org/vw6n8.pdf>).

Participants. Participants were recruited using Prime Panels from CloudResearch. We received 912 complete responses. Participants who failed the comprehension check ($n = 164$) were excluded. Analyses were conducted on the remaining 748 participants ($M_{\text{age}} = 41.27$, $SD_{\text{age}} = 14.28$; 28.6% men, 70.7% women, <1% other gender or prefer not to say; 97.1% stated English as first language; 2.8% stated otherwise, <1% prefer not to say).

Procedure and Measures. Participants were randomly assigned to one condition in a 2 (behavior present online vs. in person) \times 7 (behavior) \times 2 (order) design. As in Study 1, each participant received a vignette about an agent who tends to perform one behavior when interacting online but the exact opposite behavior when interacting in person. The behaviors and vignettes used were identical to those of Study 1. Within each condition, the order of presentation of the two paragraphs of the vignette was counterbalanced.

Participants then answered a question about whether the behavior was driven by the agent's concern for the impression that the agent makes on others. For example, in the condition where the person tended to behave self-confident online, the question was "When Susan interacts with others online, to what extent is her behavior driven by her concern about the impression that she makes on them?" (1 = "not at all", 7 = "fully"). The comprehension check was exactly the same as in Study 1.

Results

Results are displayed in Figure 2. Data were analyzed using linear mixed effects models, with impression management concern as the dependent variable, environment (online vs. in person) and order as fixed effects, and vignette as a random effect (with random slopes for environment). There was a significant effect of order, $b = -.31$, $SE = .16$, $t = -1.98$, $p = .048$. However, the difference between in-person ($M = 3.87$, $SD = 2.47$) and online ($M = 3.54$, $SD = 2.44$) fell short of significance and was, if anything, in the opposite direction of what would be needed to explain the Study 1 effect, $b = -.36$, $SE = .18$, $t = -2.30$, $p = .06$.

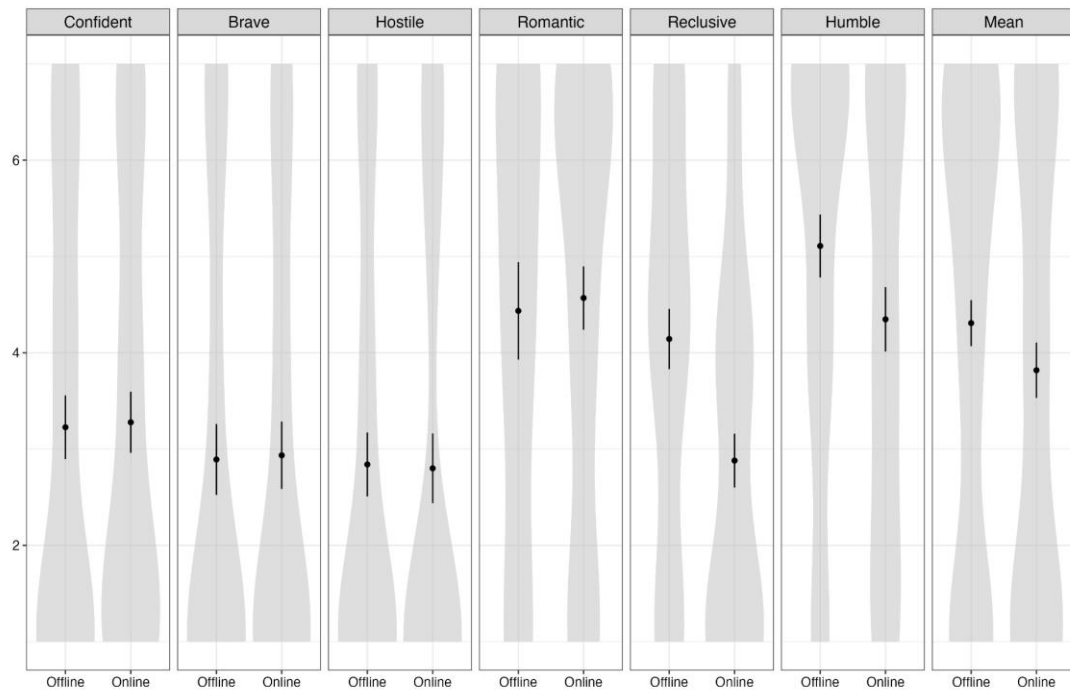


Figure 2. Violin plot showing impression management concern by environment and behavior in Study 2. Error bars show standard error.

Discussion

The distinction between in person and online that clearly predicted true self judgments in Study 1 did not predict ascribed concerns for impression management across the very same contexts of behaviors. Thus, the results of Study 2 suggest that the effect observed in Study 1 is not driven by the agent's supposed concern for impression management.

Of course, these results do not necessarily speak against the general claim that people see online behavior as especially shaped by impression management concerns. We do not find a significant effect on perceived impression management using these specific stimuli, but it might be thought that we would find an effect if we used different stimuli (e.g., stimuli that were randomly sampled from people's actual online vs. in-person behaviors). Yet, despite this, the results do provide evidence against the claim that the impact of environment on true self judgments is driven by impression management judgments. We found an effect of environment on true self judgments using certain specific stimuli, and we are now finding that those very same stimuli do not yield an effect on impression management judgments.

The question of what positively explains the difference in assessments of how revealing a behavior is for an agent's true self therefore remains open. We next turn toward a deeper explanation. It posits that the difference is driven by how natural for humans people find the environments in which a given behavior occurs.

4. Study 3

We propose an alternative explanation for the phenomenon that people see behavior as less reflective of the true self if it is exhibited in online environments as opposed to in person. This explanation posits that people consider online environments as less natural for humans and that behavior exhibited in less natural environments is perceived as less revealing of who a person really is, deep down. This explanation is tested in Study 3. It consists of a pre-study to identify relatively more or less natural environments and a main study which tests whether true self ratings depend on the perceived naturalness of the environment in which this behavior is exhibited.

Pre-Study

Method

Before running the main study, we conducted a pre-study to generate pairs of environments such that participants see one environment within each pair as more natural than the other. All methods and analyses for this pre-study were pre-registered (<https://aspredicted.org/7ri3f.pdf>).

Participants. We recruited participants using the Prime Panels from CloudResearch. We received 75 complete responses. Participants who failed the comprehension check ($n = 3$) were excluded. Analyses were conducted on the remaining 72 participants ($M_{\text{age}} = 55.67$, $SD_{\text{age}} = 14.17$; 38.9% men, 61.1% women; 97.2% stated English as first language; 2.8% stated otherwise).

Procedure and Measures. In the pre-study, participants received seven vignettes about different contrasting pairs of environments for humans to be in. For example, in the wilderness vs. prison condition, the vignette was:

In the wilderness, people hear the chirping of birds and the rustling of leaves as the wind sweeps through the branches. They feel the moss under their feet. They see vast steppes and grazing bisons in the distance of the untouched wilderness landscape.

In prison, people hear the muffled sound of heavy cell doors falling into the lock. They smell a mixture of floor wax and stew, cold cigarette smoke and unventilated beds. They see the single tree in the middle of the prison yard and the gate door in the high wall.

In the other environment conditions, the environments that participants read about were battlefield vs. drone operating room, offline vs. online, living with dying parents at home vs. visiting dying parents in hospital, workshop vs. factory, farmhouse vs. palace and monastery vs. casino. Within each condition, the order of presentation of the two paragraphs of the vignette was counterbalanced.

Participants then answered four questions for each of the seven pairs of environments they read about. The first two questions were about how common it was for humans to be in the pair of environments. The second two questions were about how natural it was for humans to be in the pair of environments. For example, in the wilderness vs. prison condition, the questions on commonness were “How common is it nowadays for humans to be in the wilderness?” (1 = “not common at all”, 7 = “very common”) and “How common is it nowadays for humans to be in prison?” (1 = “not common at all”, 7 = “very common”). The questions on naturalness were “How natural is it for humans to be in the wilderness?” (1 = “not natural at all”, 7 = “very natural”) and “How natural is it for humans to be in prison?” (1 = “not natural at all”, 7 = “very natural”).

A comprehension check question appeared on a separate page (giving participants no opportunity to go back). Participants were asked: “In each of the texts you just read, we described two different environments. Which of the following pairs of environments was NOT among the ones you just read about?” Below the question, four different numbered answers were given of which only one was correct.

Results

We conducted a separate paired *t*-test on each pair to test for differences in naturalness judgment and differences in commonness judgment. Results are shown in Table 1. We found a significant difference in naturalness for five of the pairs: wilderness vs. prison, battlefield vs. drone room, offline vs. online, workshop vs. factory and farmhouse vs. palace. These five pairs were therefore used in the main study.

As Table 1 shows, there were significant differences in commonness for three of these five items. Interestingly, for two of these items, the differences were in the opposite direction from the one observed for naturalness. That is, the environment that was judged to be more natural was also judged to be less common.

	Naturalness					Commonness				
	mean	s.d.	paired <i>t</i> test			mean	s.d.	paired <i>t</i> test		
			<i>t</i> value	<i>df</i>	<i>p</i> value			<i>t</i> value	<i>df</i>	<i>p</i> value
wilderness	5.06	1.82	8.11	1.54	<0.001	3.33	1.78	-1.48	-0.23	0.143
prison	2.32	1.73				3.76	1.98			
battlefield	3.46	2.00	2.17	0.30	0.033	3.43	2.07	-0.31	-0.04	0.759
drone room	2.89	1.76				3.51	2.02			
offline	5.65	1.72	3.44	0.66	<0.001	4.03	2.01	-8.54	-1.48	<0.001
online	4.40	2.08				6.38	1.03			
home	4.82	1.77	0.35	0.06	0.725	3.74	1.88	-8.11	-1.02	<0.001
hospital	4.71	1.84				5.58	1.74			
workshop	4.58	1.68	3.27	0.40	0.002	3.53	1.83	-3.43	-0.38	0.001
factory	3.92	1.70				4.22	1.86			
farmhouse	5.07	1.60	8.24	1.12	<0.001	4.03	1.82	7.15	0.92	<0.001
palace	3.14	1.83				2.39	1.72			
monastery	3.68	1.90	-1.60	-0.24	0.115	2.76	1.84	-9.35	-1.45	<0.001
casino	4.15	2.03				5.29	1.63			

Table 1. Descriptives and *t*-tests for environment pairs in the pretest of Study 3.

Main Study

Method

All methods and analyses for this study were pre-registered (<https://aspre-dicted.org/bz3s2.pdf>).

Participants. We recruited a new set of participants using the Prime Panels from CloudResearch. We received 677 complete responses. Participants who failed the comprehension check (*n* = 191) were excluded. Analyses were conducted on the remaining 486 participants (*M*_{age} = 50.52,

$SD_{age} = 15.90$; 34.6% men, 65.2% women, <1% other gender or prefer not to say; 94.7% stated English as first language; 5.1% stated otherwise, <1% prefer not to say).

Procedure and Measures. Participants were randomly assigned to one condition in a 2 (more vs. less natural environment) \times 5 (pairs of environments) design. We selected the five pairs of environments used in the pre-study for which we found significant differences in ratings of naturalness between the two environments of a pair. Each participant received a vignette about an agent who tends to perform one behavior when interacting in the more natural environment but the exact opposite behavior when interacting in the less natural environment.

For example, in the wilderness vs. prison condition, participants were told about an agent who behaves differently in the wilderness and in prison:

John spent part of his life in the wilderness and part of his life in prison.

In the wilderness, John heard the chirping of birds and the rustling of leaves as the wind swept through the branches. He felt the moss under his feet. He saw vast steppes and grazing bison in the distance of the untouched wilderness landscape.

In prison, John heard the muffled sound of heavy cell doors falling into the lock. He smelled a mixture of floor wax and stew, cold cigarette smoke and unventilated beds. He saw the single tree in the middle of the prison yard and the gate door in the high wall.

In the wilderness, John tended to behave in a certain way. In prison, John tended to behave in a different way.

In the other environment conditions, participants were told about an agent whose behavior differs between battlefield vs. drone operating room, offline vs. online, workshop vs. factory and farmhouse vs. palace. Within each condition, the order of presentation of the two paragraphs of the vignette was counterbalanced.

Participants then rated their agreement with a statement about whether the behavior in the first environment reflected the agent's true self. For example, in the wilderness vs. prison condition in which the wilderness was described first, the statement read "John behaved in a certain way in the wilderness and behaved in a different way in prison. But ultimately, deep down, the way he behaved in the *wilderness* is more revealing of who John really is." (1 = "fully disagree", 7 = "fully agree"). In the corresponding condition in which the prison was described first, the statement was "John behaved in a certain way in prison and behaved in a different way in the wilderness. But ultimately, deep down, the way he behaved in *prison* is more revealing of who John really is." (1 = "fully disagree", 7 = "fully agree").

A comprehension check question appeared on a separate page (giving participants no opportunity to go back). Participants were asked: "In the text you just read, we talked of different behaviors in a pair of environments. What specific behavior was described?" Below the question, four different numbered answers were given of which only one was correct.

Results

Results are displayed in Figure 3. Data were analyzed using linear mixed effects models, with true self judgment as the dependent variable, environment (natural vs. unnatural) as a fixed effect, and context as a random effect (with random slopes for environment). There was a significant effect of environment, such that behavior is performed in a more natural environment ($M =$

4.7, $SD = 1.6$) were seen as revealing of the true self more than behaviors performed in a less natural environment ($M = 3.6$, $SD = 1.7$), $b = 1.1$, $SE = .17$, $t = 6.1$, $p < .01$.

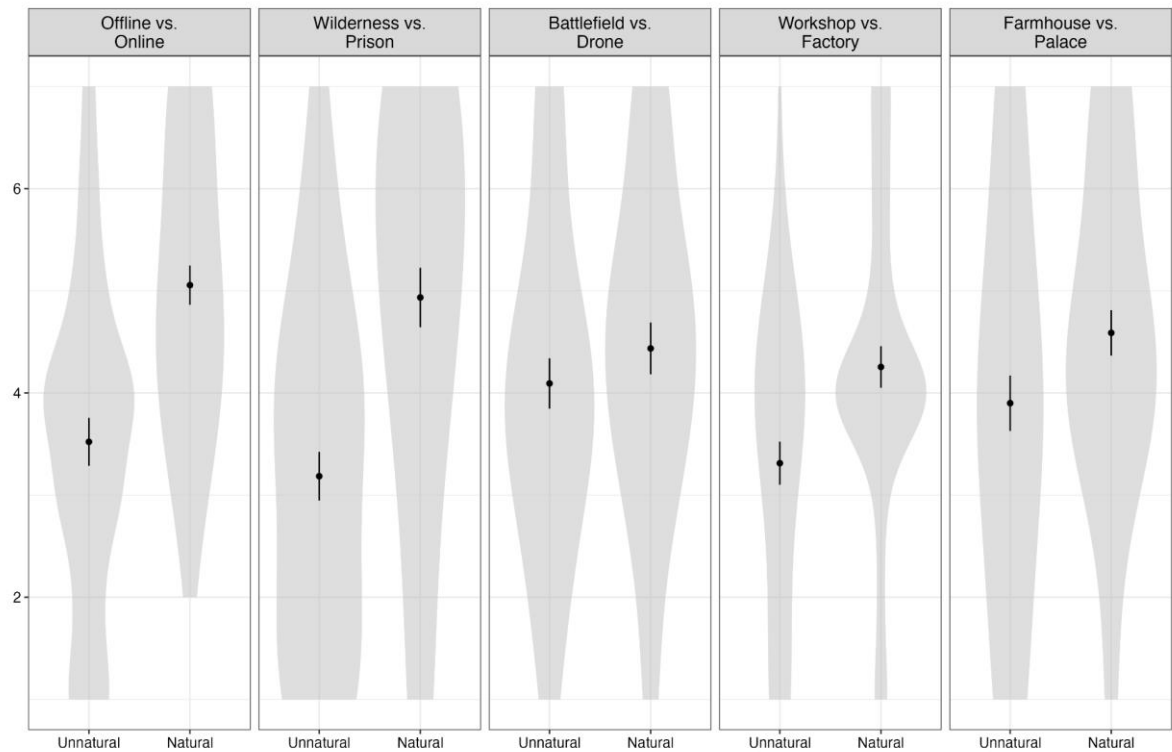


Figure 3. Violin plot showing true self judgment by environment and context in Study 3. Error bars show standard error.

Discussion

Across a range of different pairs of environments, we found a tendency such that people thought the behavior that most reflects the true self is the behavior in the environment seen as more natural for humans. Notice that this effect is not limited to the distinction between online and in-person but can also be observed for other pairs of environments whose characteristics seem to be orthogonal to the degree of technology penetrating these environments. Consider the example of the farmhouse and the palace where the former is perceived as more natural for humans to be in than the latter. In line with the general pattern, behavior in the palace is perceived as less revealing of an agent's true self than behavior in the farmhouse. It seems implausible that it is salient to participants that the palace is characterized by a higher degree of technology than the farmhouse.

Given the design of this study, we cannot be certain that the result arises because naturalness judgments are causing true self judgments. After all, there might be an unobserved confounding factor which in fact causes both naturalness judgments and true self judgments. Therefore, it is important to manipulate perceived naturalness directly to test whether the perception of an environment's naturalness for humans actually causes people's feeling that behavior exhibited in this environment is more revealing of who an agent really is, deep down.

5. Study 4

In this final study, we manipulate perceived naturalness and check for a downstream impact on true self judgments. More specifically, we use a manipulation that impacts the degree to which people see online vs. in-person environments as more natural. We then look at the downstream impact of this manipulation on true self judgments. When we use a manipulation designed to make people see the online environment as more natural for human beings, do people come to see behavior in the online environment as more reflective of the agent's true self?

To make this design feasible, we needed to turn to cases in which it would at least be plausible to think that the online environment was more natural than the in-person environment. Thus, we did not use cases like having a conversation with a close friend (where people might think that the in-person environment was obviously more natural) but instead used cases like doing banking or attending an academic conference (where people might be genuinely uncertain as to whether the in-person environment was more natural).

Method

All methods and analyses for this study were pre-registered (<https://aspre-dicted.org/nk3rc.pdf>).

Participants. We recruited participants using Prolific. We received 599 complete responses. Participants who failed the comprehension check ($n = 28$) were excluded. Analyses were conducted on the remaining 571 participants ($M_{\text{age}} = 38.99$, $SD_{\text{age}} = 12.22$; 40.8% men, 52.7% women; 6.5% other gender; 98.1% stated English as first language; 1.9% stated otherwise).

Procedure and Measures. Participants were randomly assigned to one condition in a 2 (reason for online vs. in person) \times 5 (pairs of environments) design. Each participant first received a vignette about a specific online environment with a juxtaposed in-person environment. They were then asked to provide arguments for why the first environment is more natural for humans to be in. We told them that 60 participants would try to make arguments for this claim and that we would pay the three participants providing the best arguments an additional bonus of \$5.

For example, in the banking condition, participants were told about the bank branch vs. banking online. In the condition where subjects had to provide reasons for why the bank branch is more natural for humans to be in, the vignette read as follows:

In what follows, we want you to consider two different environments that people can experience.

In the bank branch, people see other customers queuing up at the counter to do their banking. They see folding displays with advertisements for credit offers spread across the premises. They hear the scratching of their pen on paper as they confirm their banking transactions with their signature.

When banking online, people see the hot coffee on their desk steaming when they enter their access password. They see advertising banners with credit offers spread across their screen. They hear their spouse talking on the phone as they confirm their banking transactions by entering a transaction code.

The Inauthentic Online Self

Your task now is to give us an argument for the claim that **the environment people are in when they are in a bank branch is more of a natural environment for humans than the environment they are in when they are using an online banking system.**

Remember, we are going to pick the three arguments that we think people will find most convincing. These people will get a bonus of \$5 each.

Please include at least two sentences in the following textbox to make your argument.

In the condition where they had to provide reasons for why banking online is the more natural environment for humans to be in, the order of the paragraphs on the bank branch and banking online was reversed which put banking online first. Here, the task read as follows:

Your task now is to give us an argument for the claim that **the environment people are in when they are using an online banking system is more of a natural environment for humans than the environment they are in when they are in a bank branch.**

In the other environment conditions, the environments that participants read about were language school vs. language learning software, conference meetings vs. online conferences, casino building vs. online casino and coaching at clients' home vs. online coaching. Within each condition, the order of presentation of the two paragraphs of the vignette was counterbalanced.

Participants then answer the question of which environment they think is more reflecting of the agent's true self. For example, in the banking conditions, the question was "Imagine a person, John, who sometimes went to the bank branch and sometimes was banking online. In the bank branch, John tended to behave in a certain way. When banking online, John tended to behave in a different way. What do you think: In which environment is the way John behaved more revealing of who John really is, deep down?" (1 = "definitely the bank branch", 7 = "definitely banking online").

A comprehension check question appeared on a separate page (giving participants no opportunity to go back). Participants were asked: "You just read a description of a scenario in which a person was behaving very differently in two different environments. What kind of behavior was described in the scenario?" Below the question, four different numbered answers were given of which only one was correct.

Results

Results are displayed in Figure 4. Data were analyzed using linear mixed effects models, with true self judgment as the dependent variable, argument manipulation (offline natural vs. online natural) as a fixed effect, and environment pair as a random effect (with random slopes for argument manipulation). There was a significant effect of argument manipulation, such that participants were more inclined to say that the behavior performed in the online environment more reflected the true self when they were instructed to develop an argument that the online environment was more natural ($M = 5.26$, $SD = 1.85$) than when they were instructed to argue that the second behavior was more natural ($M = 3.66$, $SD = 1.85$), $b = 1.6$, $SE = .21$, $t = 7.6$, $p < .001$.

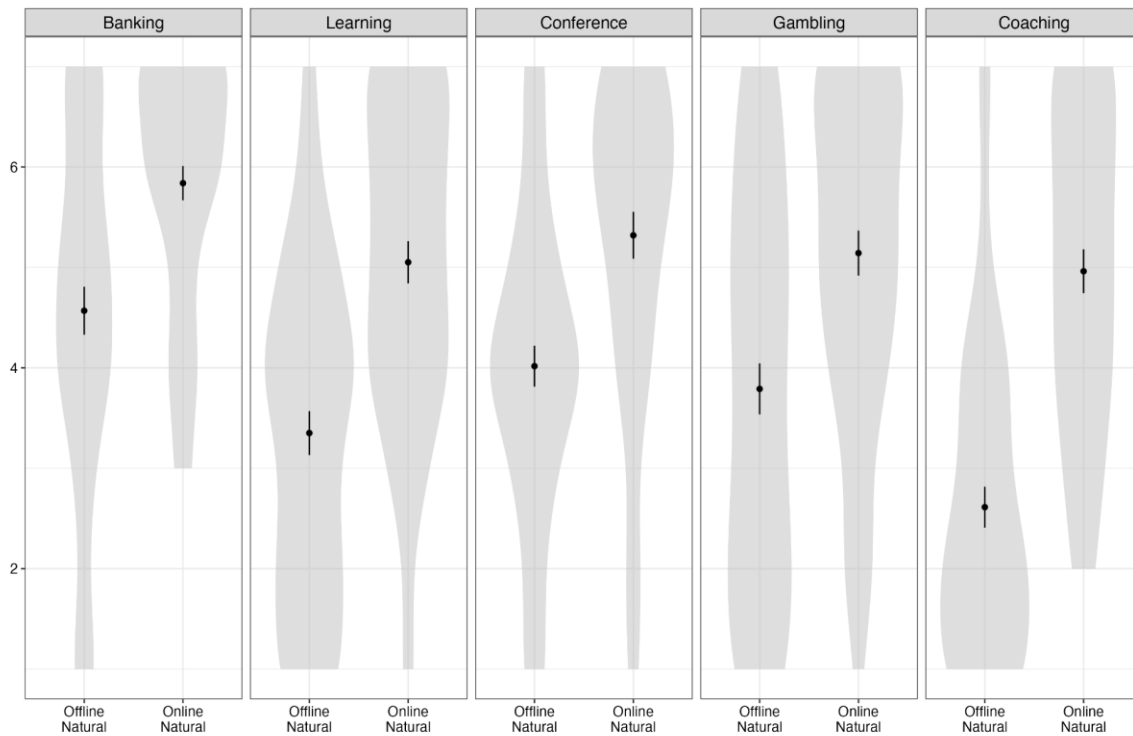


Figure 4. Violin plot showing true self judgment by argument manipulation and environment pair in Study 4. Error bars show standard error.

Discussion

Across five different contexts, participants' true self judgments were shifted toward seeing online behavior as more reflective of the true self when participants were asked to provide reasons for the higher naturalness of the online environment. The findings of Study 4 therefore confirm that the perceptions of an environment's naturalness actually cause whether behavior in this environment is perceived as more revealing of who an agent really is, deep down.

6. General Discussion

People sometimes behave differently depending on whether they are interacting online vs. in person. This phenomenon of multiple selves caused by systematically different behaviors raises questions about the authenticity of each self (Krügel and Uhl 2023). In this paper, we explored ordinary intuitions about such differences in behavior, focusing especially on intuitions about which behavior most reflects the person's true self. Studies 1 and 2 found that people see online behavior as less reflective of the true self and that this difference is not just a matter of perceived differences in impression management. Study 3 found that people show a more general tendency to think that the true self is revealed by behaviors performed in the sort of environment that is natural for human beings. Study 4 then found that when we manipulate the degree to which people

think interacting with others online is a natural environment for human beings, we get a downstream effect on the degree to which people see behaviors performed online as reflecting the true self.

Taken together, these studies provide evidence for the claim that online behavior is seen as less reflective of the true self and that this effect arises at least in part because interacting with others online is not seen as a natural environment for human beings. These findings have implications both for applied questions about online vs. in-person interaction and for more basic questions in cognitive science about how people understand the notion of a natural environment.

6.1. Implications for cultural attitudes toward online behavior

Our results tie in with a larger cultural discussion in which people increasingly express concern about the online sphere and call for a return to in-person interaction. A recent poll of US teenagers finds that more than one third think that they spend too much time on social media (Pew Research Center 2022). Similarly, a recent poll revealed that adult Britons are unhappy about their relationships with their smartphones, while a relative majority claimed to prefer to live in a world in which social media was never invented because they consider it a force for bad in society (More in Common 2024). The present results have two different implications for the study of this larger cultural discussion.

First, one obvious way to explain people's negative attitude toward online interaction would be to assume that people simply disapprove of certain specific behaviors that they believe are more performed online. For example, people might think that online behavior is more often aggressive and confrontational, or that online behavior is more often self-promotional and attention-seeking. Then they might have a negative attitude toward online interaction simply because they think it often leads to behaviors of this type. Although this sort of process may indeed contribute to people's negative attitudes, the present results suggest there is also something further occurring.

Specifically, we find that discomfort with behavior performed online is not just a matter of the behavior's specific content; there is also a discomfort that arises merely because it is happening online. Thus, in the studies reported here, even when we hold fixed the behavior performed, that behavior is considered less authentic if it is performed online than if it is performed in person. Someone who is romantic in personal online exchange but not in person is not a true romantic, while someone who is romantic in person but not in personal online exchange is. The negative attitudes people have toward online interaction might therefore be, at least in part, rooted in a concern about people acting inauthentically.

Second, these findings might shed light on recent discussions of paternalistic measures that are intended to fight online behavior. Some researchers suggest using nudges to enhance users' privacy awareness in social networks (Alemany et al. 2019). Others suggest digital nudges to fight social media addiction as part of a design for "digital detox" (Purohit et al. 2020). Warnings are also taken up by politics. In 2021, French president Macron set up a commission to provide recommendations for state regulations that, among other things, deter malicious online behavior and increase user vigilance (Bronner Commission 2022). Opinion polls also show that people are in favor of the government regulating and restricting social media (More in Common 2024). A question now arises about why people support paternalism in this case.

Again, an obvious first hypothesis would be that people observe that certain specific types of behaviors are more often performed online, conclude that these behaviors have various bad qualities, and then support interventions that decrease the prevalence of those behaviors. There might well be something correct about this obvious hypothesis, but the present findings indicate that there might also be something more complex at work. Perhaps the core issue is not just about people's judgments about which behaviors are good and which are bad, but also about people's judgments about whether those who perform these behaviors truly do want to perform them.

This hypothesis reflects an important strand within the academic debate on paternalism. One view within this debate is that people simply want different things in different situations and at different times (Sugden 2008), while another view is that paternalistic policies actually help people to perform the behaviors that they themselves most truly want to perform. On this latter view, libertarian paternalism "encourages both private and public institutions to steer people in directions that will promote their own welfare" (Sunstein and Thaler 2003, p. 1201), i.e., people's welfare as judged by *their own* standards. Support for paternalistic measures against online behavior might be best explained in precisely this way. That is, if government policies encourage citizens to interact with each other in person rather than online, people might think that these policies actually cause citizens to behave in a way that is more in keeping with what they themselves truly want to do. Future research might explore whether perceptions of authenticity help to explain paternalistic preferences against online behavior.

6.2. The Role of Natural Environments for Our Judgments

In our attempt to understand how people think about online vs. in-person environments, we drew on the hypothesis that people have an intuitive notion of "natural environments for human beings." Although we initially introduced this notion as a way to address one very specific question, future research could explore people's use of it more broadly.

As we noted in the introduction, the idea that people's behavior can be impacted by judgments of naturalness has already been explored in numerous studies, but the notion explored here is importantly different. Our focus has not been on judgments about whether certain things are natural in themselves but rather on judgments about whether certain environments are seen as natural environments *for a particular species*. People might think that the natural environment for penguins is the tundra of Antarctica and that the Sahara desert would be a very unnatural environment for a penguin, but this does not mean that people think there is anything intrinsically unnatural about the Sahara desert. In much the same way, people can make judgments about which environments are natural for human beings. These are not judgments about which environments are intrinsically most natural but judgments about which environments are natural for our species in particular.

One key task for further research will be to try to understand at a deeper level the relationship between judgments about whether an environment is natural for human beings and judgments about whether behaviors performed in that environment reflect the true self. Why exactly does the former have such an impact on the latter? One possibility would be that people associate naturalness with spontaneity (Levinovitz, 2020) and that they think that spontaneous behaviors most fully reveal the true self (Oktar & Lombrozo, 2022). Another possibility would be that there is a much more general tendency such that people think that the essence of an object tends to be most fully

revealed in the environment that is most natural for that object (e.g., that the true essence of a plant will be revealed in the environment that is most natural for that plant). Then, since people seem to think that the essence of a human being is his or her true self (Christy et al., 2019), people may think that the true self of a human being is most revealed in the environment that is natural for that human being.

In further research on this notion, one obvious topic would be which environments people tend to regard as most natural for human beings. The present studies focused especially on the role of technology, but future studies could look at other issues (e.g., which ways of being treated by one's parents or caregivers are seen as most natural). Taking this inquiry to a deeper level, one could ask what factors more generally influence people's judgments of naturalness (is it a matter of which environments people think humans inhabited in the distant past? are they influenced in any way by people's value judgments about which environments truly are best?). Finally, it might be helpful to look at the relationship between the notion of a natural environment and other folk-biological concepts, such as the concept of essence or the concept of teleology. Do people think, for example, that the essence of a species is most clearly manifest when it appears in the environment that is most natural for that species?

On a more general level, our studies suggest that folk concepts may play an important role in coming to a more profound understanding of our individual and cultural attitudes towards technology. Empirical folk concepts may be orthogonal to the analytical concepts that philosophers of technology have elaborated when reflecting on these issues. It therefore seems worthwhile to complement the normative reflection on technology with a systematic empirical investigation of people's underlying intuitions. This investigation may not only inform the academic discourse on the societal implications of technology but also the policy discourse on how to deal with these implications.

References

- Aleman, J., Del Val, E., Alberola, J., & García-Fornes, A. (2019). Enhancing the privacy risk awareness of teenagers in online social networks through soft-paternalism mechanisms. *International Journal of Human-Computer Studies*, 129, 27-40.
- Bronner Commission (2022). "Enlightenment in the Digital Age". Report from the Bronner Commission. <https://www.elysee.fr/front/pdf/elysee-module-19389-en.pdf>
- Christy, A. G., Schlegel, R. J., & Cimpian, A. (2019). Why do people believe in a "true self"? The role of essentialist reasoning about personal identity and the self. *Journal of Personality and Social Psychology*, 117(2), 386.
- De Freitas, J., & Cikara, M. (2018). Deep down my enemy is good: Thinking about the true self reduces intergroup bias. *Journal of Experimental Social Psychology*, 74, 307-316.
- Heiphetz, L., Strohminger, N., Gelman, S. A., & Young, L. L. (2018). Who am I? The role of moral beliefs in children's and adults' understanding of identity. *Journal of Experimental Social Psychology*, 78, 210–219.
- Ji, L. J., Lappas, C. M., Wang, X. Q., & Meier, B. P. (2023). The naturalness bias influences drug and vaccine decisions across cultures. *Medical Decision Making*, 43(2), 252-262.
- Krügel, S., & Uhl, M. (2023). Is only one of my selves authentic? An empirical approach. *Journal of Behavioral and Experimental Economics*, 102, 101971.
- Levinovitz, A. (2020). *Natural: How faith in nature's goodness leads to harmful fads, unjust laws, and flawed science*. Beacon Press.
- Lappas, C. M., Coyne, N., Dillard, A. J., & Meier, B. P. (2023). Do Physicians Prefer Natural Drugs?. *European Journal of Health Psychology*.
- LifeBonder (2023). Clean Internet: Fighting the Negative Impact of Social Media. <https://life-bonder.com/>
- Marwick, A. E. (2013). Online identity. *A companion to new media dynamics*, 355-364.
- Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133.
- Meier, B. P., Dillard, A. J., & Lappas, C. M. (2019a). Naturally better? A review of the natural-is-better bias. *Social and Personality Psychology Compass*, 13(8), e12494.
- Meier, B. P., Dillard, A. J., Osorio, E., & Lappas, C. M. (2019b). A behavioral confirmation and reduction of the natural versus synthetic drug bias. *Medical Decision Making*, 39(4), 360-370.
- Meier, B. P., & Lappas, C. M. (2016). The influence of safety, efficacy, and medical condition severity on natural versus synthetic drug preference. *Medical Decision Making*, 36(8), 1011-1019.
- Moore, C., Barbour, K., & Lee, K. (2017). Five dimensions of online persona. *Persona Studies*, 3(1), 1-11.

The Inauthentic Online Self

- More in Common (2024). *Phoning it in: Britons, Smartphones and Social Media*. May 2024. <https://www.moreincommon.org.uk/media/io5k3wjw/mic-phoning-it-in.pdf>
- Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, 40, 203-216.
- Newman, G. E., Freitas, J. D., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, 39, 96–125.
- Oktar, K., & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition*, 223, 105021.
- Paliszkiewicz, J., & Mądra-Sawicka, M. (2016). Impression Management in Social Media: The Example of LinkedIn. *Management (18544223)*, 11(3).
- Papacharissi, Z. (2002). The presentation of self in virtual life: Characteristics of personal home pages. *Journalism & Mass Communication Quarterly*, 79(3), 643-660.
- Picone, I. (2015). Impression management in social media. *The international encyclopedia of digital communication and society*, 36, 469-476.
- Purohit, A. K., Barclay, L., & Holzer, A. (2020, April). Designing for digital detox: Making social media less addictive with digital nudges. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-9).
- Siegrist, M., & Hartmann, C. (2020). Consumer acceptance of novel food technologies. *Nature Food*, 1(6), 343-350.
- Skubisz, C. (2017). Naturally good: Front-of-package claims as message cues. *Appetite*, 108, 506-511.
- Statista (2023). America's Top New Year's Resolutions for 2024. <https://www-statista.com/chart/29019/most-common-new-years-resolutions-us/>
- Strohming, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, 12(4), 551–560.
- Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, 19, 226-248.
- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago law review*, 1159-1202.
- Pew Research Center (2022). *Teens, Social Media and Technology 2022*. August 2022. https://www.pewresearch.org/wp-content/uploads/sites/20/2022/08/PI_2022.08.10_Teens-and-Tech_FINAL.pdf
- Wilks, M., & Bloom, P. (2022). Children prefer natural food, too. *Developmental Psychology*, 58(9), 1747.

The Inauthentic Online Self

Wilks, M., Hornsey, M., & Bloom, P. (2021). What does it mean to say that cultured meat is unnatural?. *Appetite*, 156, 104960.