Forthcoming in Topoi
https://link.springer.com/article/10.1007/s11245-024-10149-1
Please refer to published version

Beauty Filters in Self-Perception: The Distorted Mirror Gazing Hypothesis

Gloria Andrada (Universidade NOVA de Lisboa/ gandrada@fcsh.unl.pt)

Abstract

Beauty filters are automated photo editing tools that use artificial intelligence and computer vision to detect facial features and modify them, allegedly improving a face's physical appearance and attractiveness. Widespread use of these filters has raised concern due to their potentially damaging psychological effects. In this paper, I offer an account that examines the effect that interacting with such filters has on self-perception. I argue that when looking at digitally-beautified versions of themselves, individuals are looking at AI-curated distorted mirrors. This helps identify two potential cognitive effects of this behavior. First, it can elicit affective attitudes that change how individuals feel when looking at their unfiltered self-images. Second, exposure to filtered self-images might cause a perceptual normalization of such images. Finally, I argue that this form of distorted mirror gazing is a novel cultural practice for self-perception, and I highlight some ways in which this practice could be critically evaluated and ultimately changed.

Keywords: self-perception; cognitive development; filters; augmented reality; artificial intelligence; beautification; cognitive enculturation.

1. Introduction

Picture the following scenario. You hear a notification on your phone, and it's a message from a friend. You open it and see that she has sent you a selfie, i.e., an image of herself taken by herself using a digital camera. You want to reply with one of your own, but you haven't slept well, or you simply want to look better than the way you see yourself in the mirror. You take a picture of yourself and apply a *beautifying* filter. Your skin looks smoother, your eyes brighter, your lips a bit plumper. You send that. This scenario might be far from some readers' experience, but your phone might be doing this for you, even without you knowing it. In fact, some popular smartphones have a feature called "beauty" or "beauty mode" that is automatically activated when the front camera is used to take a selfie, seamlessly making you look better in your pictures.

My goal in this paper is to look into the effect that interacting with so-called *beauty filters* has on self-perception. To this end, I will offer an explanation of how self-perception, in particular face recognition, can be affected by looking at these digitally beautified self-images. Given the centrality of face perception to many psychological phenomena (such as self-esteem and body appreciation), looking carefully at the effects of this increasingly widespread human behavior seems relevant to understanding human cognition and human cognitive development in our current technological environment. Moreover, while increasing attention is being paid to some of the cognitive biases, attention changes, and

growing polarization that social media use and AI applications cause,[1] the effects of AI on self-perception and experience are yet to be sufficiently well understood. So this paper contributes to filling this gap. In doing so, it explores some other types of cognitive transformations that interacting with AI systems elicits in human cognition.[2]

My plan is the following. I begin by looking into how face filters work and some data regarding filters and social media use (§2). I then proceed to look at the cognitive ability of face recognition via mirror gazing and some of its implications in the development of self-perception (§3). This is followed by an explanation of what I take to be a central cognitive process behind the behavior of looking at *beautified* self-images, which I call the Distorted Mirror Gazing Hypothesis (§4), and which has two components: one concerning the affective attitudes that looking at altered and unaltered self-images causes (§4.1), and the other explaining the process of perceptual normalization of filtered self-images (§4.2). I conclude by characterizing the behavior of distorted mirror gazing as a new cognitive practice, one which is part of the process of cognitive enculturation (§5). This explains some of the risks of this behavior, but also illuminates the space left for agency and change.

2. Face filters and beautified self-images

Face filters are automated photo editing tools that use artificial intelligence and computer vision to detect facial features and modify them. What these filters do is analyze an image's shapes, colors and features, and then generate a new, supposedly appealing image. Some of these filters can be fantastical, for example, they can add cat-like features to a human face, even during online video calls. But others transform certain facial features, especially by making one's eyes and eyelashes bigger, reducing nose size, making one's lips bigger and plumper, and, in a racist tendency, making skin lighter in color, as well as smoother (i.e., reducing wrinkles, spots, and skin tone variation).[3] There are also recent face filters, such as the *Beauty Scanner*[4] filter, which allegedly "measure" your beauty using facial symmetry, facial structure, and the "golden ratio technique" used in facial cos-

---

[1] See for example Gillet and Heersmink 2019; Sparrow and Wegner 2011; Williams 2018; Vicente and Matute 2023.

[2] This connects with has been called *transformational transparency* (Andrada, Clowes and Smart 2021). This form of AI transparency (as opposed to opacity) is concerned with the neural and bodily transformations that pervasive use of AI technologies elicits.

[3] As a reviewer points out, making the skin smoother and wrinkle-free displays an ageist bias.

[4] See: https://play.google.com/store/apps/details?id=com.golden.ratio.face&hl=en_GB Last accessed: May 15, 2024

metics.[5] These filters give users a beauty score and suggest a beauty ratio for faces. Other apps, like *FaceTune*,[6] let their users directly apply filters with different features (e.g., a specific hairstyle, makeup, or changes in facial features).

Face filters use a combination of computer vision, machine learning, and image processing techniques. What follows is a simple reconstruction of what face filters do to a facial image when they are applied (Hedman et al. 2022). First of all, in computer vision there is a *face detection feature* in which the presence of a face in the image is detected using algorithms that can identify facial features such as eyes, nose, mouth, and overall face shape. This is followed by *landmark identification.* Basically, once a face is detected, the system maps out key points on the face (landmarks) such as the corners of the eyes, the edges of the lips, and the jawline. This helps in accurately applying the filter. The next step is *feature enhancement.* This is where the so-called beautification in beauty face filters takes place. This includes features such as changes in skin tone, eye and tooth brightness, and facial proportions and structure (e.g., the jawline). Depending on the filter, this may include augmented reality (AR) effects such as adding virtual makeup, or 3D elements like hats or glasses. These effects are mapped onto the corresponding facial landmarks, and can also adjust dynamically as the face moves by tracking its movements in real time. This process is afforded by sophisticated machine learning. The algorithms behind these filters are often trained on large datasets of faces to understand how different features vary between individuals. Convolutional neural networks (CNNs) are commonly used for the tasks of facial recognition and feature mapping, and generative adversarial networks (GANs) are used to create realistic enhancements and alterations (Hedman et al. 2022; Pandy et al. 2023). Modern smartphones and computers often have specialized hardware, such as GPUs and dedicated AI chips, to process the computation required by these filters.

We can see that face beauty filters are automated technologies that allow smartphone and social media users to curate and manipulate images of themselves via a specific type of augmented reality that “beautifies” their facial features, where the process of beautification is meant to improve their faces’ physical appearance and aesthetic attractiveness. These filters can be used for self-portraits (or “selfies”), or for videos on various social media platforms (see fig 1). They are also available in video-conferencing platforms (such as Zoom) and in some video call apps (sometimes directly and sometimes via plug-

[5] See for example Hayduke (2020).

[6] See https://www.facetuneapp.com/. Last accessed: May 14, 2024.

ins or additional apps that process the live camera). Take, for example, TikTok's *Bold Glamour* beauty filter. This TikTok filter was used over 16 million times during the first month after its release in February 2023.[7] It emphasizes the contour of one's cheekbones and jawline in a sharp but subtle way. It also highlights the tip of one's nose, the area under one's eyebrows, and one's cheeks. In addition, it lifts eyebrows, illuminates eyelids, and makes eyelashes thicker and longer. The use of this latter type of Augmented Reality beauty filters on self-images will be the focus of this work.

Fig.1 Self-Images with Instagram Beauty filters

Now, let us briefly look at some statistics. A recent UK report[8] found that in a sample of 175 participants with an average age of 20 years, 90 percent of young women reported using filters to edit their photos. Participants claimed to use filters on social media to smooth and brighten their skin tone, whiten their teeth, reshape their jaws or noses, and make their lips plumper and their eyes bigger. Research so far shows this behavior to be more common among young women, therefore it seems to be gendered.[9] However, the fact that social media use is increasingly common among young people and older adults, as well as the fact that most social media platforms offer beautifying filters, suggests that this practice will become more widespread in the coming years.

---

[7] See https://www.technologyreview.com/2023/03/13/1069649/hyper-realistic-beauty-filters-bold-glamour/ Last accessed: May 26, 2024.

[8] See https://www.city.ac.uk/__data/assets/pdf_file/0005/597209/Parliament-Report-web.pdf. Last Accessed: April 25, 2024

[9] In section 5 I come back to the gendered aspect of this behavior.

For example, in a recent survey conducted by Pew Research Center,[10] most adult U.S. citizens (68%) reported using social media platforms.[11] In particular, roughly half of U.S. adults (47%) reported using Instagram. This was followed in popularity by Pinterest, TikTok, LinkedIn, WhatsApp and Snapchat (with users ranging from 35% to 27% of U.S. adults). Interestingly, TikTok's adult user base has grown exponentially since 2021. And, importantly for our purposes, three of these widely used social media platforms (Snapchat, TikTok and Instagram) have incorporated beautifying filters. Moreover, one out of five U.S. teens described their social media use as "almost constant". In particular, the majority of teens, aged 13 to 17, said they use TikTok (63%), Snapchat (60%) and Instagram (59%) regularly.

This shows the need for more research on the effects of filter use on cognition in various populations across the globe. In fact, the use of AR beauty filters has already caught the attention of mental health professionals and aesthetic physicians (Habib, Nazir and Mahfooz 2022). Also, a growing number of cosmetic surgery professionals are reporting an increase in the number of patients seeking procedures to improve the way they look in selfies (Rajanala et al. 2018). More precisely, these procedures are usually done to obtain facial features that resemble those that have been enhanced and modified by automated face filters (Rajanala et al. 2018; Tremblay, Tremblay and Poirier 2020).

Summing up, AR-beautified self-images are a phenomenon that is already widespread in certain social groups, and although there is still evidence to be gathered concerning the use of filters across different populations, understanding their effect on cognition calls for attention. In the next section, I will look into the ability of face recognition via mirror gazing, in order to begin to understand some of the cognitive effects that looking at these digitally beautified self-images might have on human cognition, and in particular, on self-perception.

3. Facial self-recognition via mirror gazing

Developmental and phenomenological accounts of psychological development have traditionally considered facial self-recognition to be a decisive stage in the development of self-

[10] The survey was conducted on Sept. 26-Oct. 23, 2023, among 1,453 13- to 17-year-olds, and it covered social media, internet use and device ownership among teens. https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/ Last accessed: March 25, 2024.

[11] These findings come from a Pew Research Center survey of 5,733 U.S. adults conducted May 19-Sept. 5, 2023. Last accessed: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/. Last accessed: March 25, 2024.

consciousness (Gallup 1970; Merleau Ponty 1964). We find reference to the *mirror test* (i.e., being able to recognize the image in the mirror as your own image) as a test for self-awareness (Gallup 1970). And to this day, there is an ongoing debate over the extent to which nonhuman animal species are capable of recognizing themselves in mirrors, and the implications of this for their cognition (Gallup et al. 2002; Wittek et al. 2021; Schilhab 2004).

Although visual facial self-recognition is a significant achievement in development, it is important to point out that it should not be taken to be the central or fundamental form of self-awareness. As Zahavi (2014) writes:

> Although facial self-recognition might testify to the existence of a form of self-awareness, the failure to recognize one's own face certainly does not prove the absence of every form of self-awareness. (p. 201)

In other words, visually recognizing your face in the mirror as your own might be an indicator of self-awareness, but it is clearly not the only one, as there are other perceptual modalities. Moreover, as Zahavi (2014, pp. 197-207) states, in order to recognize themselves in the mirror, human infants must have a previous sense of their own bodies as environmentally embedded and embodied entities. For example, they should have an embodied sense of themselves in perception and action, so that they can detect the cross-modal and temporal match between their own bodily movements and the movements of the mirror image. And this seems to require proprioception of their own bodily movements and posture, and a kind of kinesthetic-visual matching ability.

This connects with an ongoing controversy over the mechanisms that give rise to face recognition via mirror gazing. Traditionally, self-recognition in humans, and in our closest primate relatives, has been understood as a cognitive product of primate evolution, which stems from more recent neural mechanisms that develop through experience-independent mechanisms during ontogenesis (Gallup and Anderson 2020). But recent research shows that the development of mirror self-recognition in infants is a perception-action achievement that builds on the infant's ability to haptically localize and reach to targets on their body (Chinn et al. 2024). This is based on the fact that infants with prior experience of reaching to tactile targets on their bodies, in the months before they could recognize themselves in the mirror, achieved mirror self-recognition earlier than infants who were not exposed to those experiences.

In any case, despite the controversy over what facial self-recognition via mirror gazing requires, and despite the fact that there are other, perhaps more fundamental forms of

self-awareness, mirror face recognition has been considered a highly relevant stage from a developmental perspective. Merleau Ponty (1964)'s analysis has been especially influential. What Merleau Ponty (1964) argued is that a mirror does not provide redundant information to the infant, but rather gives them a visual representation of their own body that is different from the one they can obtain on their own, in particular regarding their face. In other words, mirror gazing affords children the ability to perceive their own facial features, but also gives them a different apprehension of their bodily features that is not available via interoception or proprioception. By being exposed to mirrors (or, similarly, to other reflective surfaces) human infants become aware of how others visually perceive them. The mirror experience is thus a key developmental stage, since this is when infants learn how they are visually perceived by others (Merleau Ponty 1964). In doing so, infants are progressively exposed to a world where the visual appearance of one's face has a high social valence (Weiss 1999). This progressively elicits attitudes about themselves and their physical appearance that are then incorporated into what is traditionally known as the *body image*.

Although the definition of a body image is still contested in the literature,[12] for simplicity we can here follow a standard characterization put forward by Gallagher (1985, 2005). On this view, a body image is a conscious representation that serves body monitoring. Phenomenologically, it is experienced as an object of awareness, in the form of a reflective (i.e., explicit) bodily self-awareness. It involves personal-level processes such as perceptual bodily experience, a conceptual understanding of embodiment, and affective attitudes directed toward one's body. The body image is often contrasted with the body schema, which is understood as a (mostly) nonconscious representation that serves body performance, and which involves subpersonal processes that enable posture and movement (Gallagher 1986, 2005).

Despite the fact that this dyadic distinction has been challenged (see, e.g., de Vignemont 2009), what matters for us now is that the body image comprises a series of perceptions and attitudes toward our own body, which include how we see ourselves and how we value our physical appearance. And it is here where the social dimension of facial self-recognition is relevant. This is due to the fact that how we see ourselves in the mirror does not solely depend on visually recognizing our face, but also has an evaluative dimension (or dimensions) that largely depends on our cultural values. For example, recent body image literature reviews have revealed significant ethnic differences in weight status and body size perception (see Gramaglia, Delicato, and Zeppegno 2018). The same goes for

[12] For different perspectives on this see for instance de Vignemont 2009, 2018, and Weiss 1999.

the perception of beauty and beauty standards (Taylor 2015). Therefore facial self-recognition via mirror gazing contributes to self-perception in ways that go beyond visually recognizing the face in the mirror as you own, also including different aspects having to do with our place in the social world, such as how we perceive and value our physical appearance given our social environment. This connects with important psychological dimensions such as body appreciation (Quittkat et al. 2019). With this in mind, we are now better equipped to look into the effects that digitally beautified self-images can have on self-perception.

4. The Distorted Mirror Gazing Hypothesis

I want to propose that an important way in which interacting with so-called *beauty* filters affects cognition, and in particular, self-perception, is illuminated if we think of this interaction as a form of distorted mirror gazing. As I will proceed to show, this can explain some of the phenomena that have been associated with this practice, including some forms of body image dissatisfaction and body dysmorphia (Rajanala et al. 2018). This will be captured by what I call the *Distorted Mirror Gazing Hypothesis*, which has two components: (i) *affective distorted mirror gazing*, and (ii) *perceptual normalization of the AR-filtered face*. Let me introduce each of these components in turn.

4.1 Affective distorted mirror gazing

If we think of interacting with AR beauty filters as a form of distorted mirror gazing, a first clear cognitive dimension that we need to take into account is how seeing your face with virtually augmented modifications might change the way you perceive your own face. A first direct worry is that looking at subtly altered self-images might dangerously cause a mistaken identification of one's actual facial features with the AI-beautified features. This could cause a blurring of fantasy and reality, given our exposure to distorted self-reflections.

This identification with an unrealistic image of one's face might be a factor that contributes to negative body image issues, e.g., negative appreciation of one's physical ap-

pearance.[13] But how could this happen? What could explain the underlying cognitive mechanisms?

The Distorted Mirror Gazing Hypothesis starts with the idea that exposure to AI-beautified self-images gives rise to various affective attitudes that have an effect on the individual's mirror experience. This claim is grounded in recent work developed by Tramacere (2022). Tramacere argues that self-feelings can affect mirror gazing, in the sense that a previously negative or positive acquired feeling toward oneself can affect how one perceives oneself in the mirror. Her argument is simple and sound. It is based on two facts. First, it has been shown that our perception of others' faces is affected by our feelings toward them. Findings in social psychology show that when we perceive others, our affective attitudes toward them modulate our responses to their face. For example, consciously or unconsciously appreciating others affects whether, and to what extent, we respond with positive or negative emotions and corresponding facial mimicry when looking at them (Tramacere 2022; McIntosh 2006; Bourgeois and Hess 2008; van Baaren et al. 2009). In this respect, it has been claimed that individuals recognize happy faces faster and more accurately than negative faces if the former facial expressions had previously been associated with positive personality features. The same has been discovered the other way around; that is, in the case of faces that had previously been associated with negative personality features, individuals tend to be more accurate in categorizing negative expressions such as anger and sadness (Bijlstra et al. 2014; Albohn and Adams Jr. 2016).

Second, social neuroscience and social psychology show that we use similar neurocognitive mechanisms in our perception of others' and our own faces (Tramacere 2022; Decety and Sommerville 2003; Uddin et al. 2005; Bretas et al. 2021; Gallese 2003). This process comprises specific populations of neurons engaged in what has been called the "face processing network", such as specific areas of the primary visual cortex, the occipital face area, the fusiform face areas, and the anterior and posterior regions of the superior temporal sulcus.[14]

---

[13] A reviewer points out that this blurring of the self-image could be characterized as a splitting or bifurcation where someone gets to have a "normal mirror me" and a "filtered mirror me", and comes to prefer one over the other. If that is the case, then the Distorted Mirror Gazing Hypothesis I am putting forward, specially its Affective Distorted Mirror Gazing component, would be an explanation of at least part of the process behind the formation of this preference. This connects with important work in feminist phenomenology concerning the psychological fragmentation and the splitting of selfhood caused by different forms of cultural aesthetic pressure, and that causes what some have characterized as a form of self-objectification (see for instance Young 1990, Bartky 1990 pp. 22-31 add pages, and Weiss 1999, pp 39- 64). Though I don't delve into these aspects in the present work, in section 5, I do come back to some of the gendered aspects of the use of beauty filters.

[14] See Tramacere 2022 p. 3 for a very illuminating model on how this might happen.

From these two facts, Tramacere (2022) concludes that it makes sense to expect the affective attitudes we have toward ourselves to affect facial self-perception, as well as our behavioral and psychological responses to our own mirror image. In other words, our affective attitudes toward ourselves might positively or negatively bias our behavioral and emotional responses to our own face in the mirror. In light of this, Tramacere (2022) writes:

> If my argument is correct, a negative way of representing oneself could produce negative emotions and corresponding facial expressions during mirror self-recognition. For example, an aversive self-image could (perhaps unconsciously) bias individuals' facial expressions toward certain emotions (sadness, disgust, and anger), and corresponding covert facial mimicry. (p. 6)

So the key point is that just like in the case of perceiving other faces, self-feelings affect our perception of our own face. It should be noted that perception here is understood as recruiting and involving emotional and affective responses. Tramacere (2022) is therefore concerned with "higher-order processes of perception, where the multimodal sensory coding of a percept (i.e., faces) overlaps and intermingles with motor and affective coding" (p. 7).

With this in mind, let us go back to the Distorted Mirror Gazing Hypothesis, and its first component, namely, *affective distorted mirror gazing*. As we know, in the case of AR-beautified self-images, the self-image is distorted, in the sense that it reflects something that is not present in the offline world, but only in the digitally modified version. And the key point is that looking at a beautified self-image might elicit in the viewer certain affective attitudes that then affect how they perceive their unfiltered mirror image (for example, by using simply their front camera). This means that, according to my hypothesis, there are two interrelated steps that happen when someone is exposed to AI-beautified images of their face:

(i) First, an AI-filtered image elicits positive or negative self-feelings or other affective attitudes directed towards themselves.
(ii) Second, these affective attitudes affect the viewer's perception of their (unfiltered) facial image, which in turn elicits other positive or negative self-feelings or other self-affective attitudes.

This is just an initial model that should be fleshed out empirically. For example, research is required to identify what affective attitudes are elicited, whether they change ac-

cording to one's social group and/or individual differences, and the time one needs to be exposed to such images in order for the relevant affective attitudes to be developed. But for now, and focusing on the cases that have caught the attention of mental health professionals and that connect use of so-called beautifying face filters and negative body image issues, it is reasonable to hypothesize that what happens is the following.

First, looking at an AI-beautified self-image elicits positive feelings or other positive affective attitudes (see fig. 2). This could be explained by the fact that the alteration or distortion of the face is not random, since these changes are made according to specific societal beauty standards and ideals. The viewer thus perceives subtle changes in their own face that make it more in tune with current beauty ideals (e.g., less wrinkles or smoother skin). Then, this elicits a conscious or unconscious positive affective response, such as happiness. Obviously this is just a reconstruction that, as I have just indicated, needs to be backed up empirically, but research on body image disorders suggests that an intense desire for social conformity appears to be behind the overinternalization of societal standards of attractiveness (Lenny, Vartanian, and Hopkinson 2010). So perhaps it is precisely the desire to conform to societal standards that lies behind the development of positive attitudes toward the beautified version. Moreover, these positive affective attitudes might also interact with other previously acquired self-directed affective attitudes.

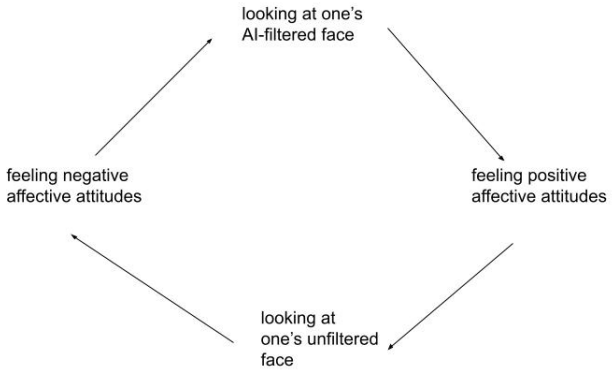

Fig. 1 Affective Distorted Mirror Gazing 1

Then the aforementioned positive affective attitudes affect the viewer's perception of their (unfiltered) facial image, for instance by making them feel inadequate, or as if something is missing. And these negative attitudes in turn contribute to strengthening their positive attitudes toward the beautified self-image. In other words, the mismatch between the beautified self-image and the unfiltered one might create a vicious cycle or feedback loop of increasing dissatisfaction with one's unmodified reflection. In the scarce empirical literature on the use of beauty filters, we can find some evidence in support of this, as some individuals have reported that after being exposed to AR-beautified self-images, they experienced "negative emotions when they switched back to their front camera" (Isakowitsch 2023, p. 245). So it makes sense that this kind of affective modulation is what is at stake in some cases, especially in those that lead certain individuals, especially young women and other vulnerable people, to seek aesthetic procedures that would make their actual face look more like its digital version ((Rajanala et al. 2018).

Of course, a situation in which an AR-beautified self-image elicits negative affective attitudes, and not positive ones, is indeed possible. And in fact, there is also some evidence in support of this. For example, in the previously mentioned research, three out of eight individuals reported that their emotional response when going back to the unfiltered front camera was more positive than negative (Isakowitsch 2023, p. 245). *Affective distorted mirror gazing* can accommodate different scenarios, including this one, as it simply states that AR-beautified self-images elicit different affective attitudes that affect unfiltered-mirror gazing. So, if we focus again on the cases that connect the use of beauty filters and negative body image issues, we could end up with a different form of distorted affective mirror gazing in which there is a different reinforcing loop of negative affective attitudes towards one's unfiltered self-image (see Fig. 3).

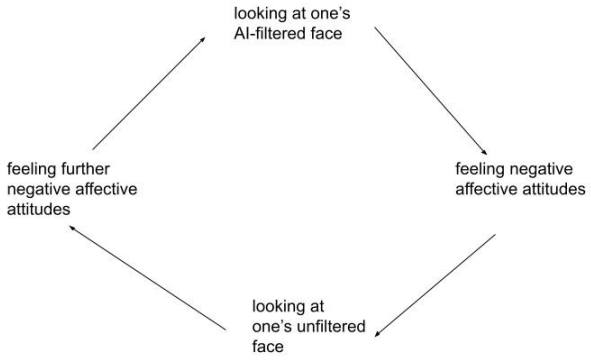

Fig. 3 Affective Distorted Mirror Gazing 2

This could be explained the following way. The beautified self-image is perceived as a set of beauty standards that individuals do not meet in reality, and that works as a sort of "carrot and the stick", an unattainable goal.[15] Looking at this unreachable digitally beautified self-image gives rise to feelings of inadequacy towards their altered self-image. And this, in turn, elicits further feelings of discomfort and inadequacy towards their unfiltered face. This form of distorted mirror gazing could thus become a reinforcing loop of negative affective attitudes, which could be behind the worrying cases that have caught the attention of mental health and other professionals.

Two final points need to be made here.

First, as we saw earlier, the perceptual process beyond mirror gazing is understood in a coarse-grained manner, that is, as recruiting and involving emotional and affective responses. Still, a key question that is left open concerns not only how beauty filters elicit affective responses that affect mirror gazing by modulating one's overall mirror experience, but also, in what sense, if any, these attitudes affect facial perception in a more restricted sense (i.e., as identifying visual properties of the observed object: one's own face in the cases that we are considering).

Second, in the previously mentioned study, most participants reported that the longer they looked at their filtered self-images, "the more real it felt" (Isakowitsch 2023, p. 244). This suggests that people can become used to their AR-beautified versions, and this

[15] Thanks to a reviewer for inviting me to explore this option.

is in fact in tune with what mental health professionals have been warning against, namely a risky identification with filtered self-images.

These two issues take us to the second component of the Distorted Mirror Gazing Hypothesis, namely *perceptual normalization of the AR-filtered face*.

4.2 Perceptual normalization of the AR-filtered face

As we just saw, AR-beautified self-images can elicit affective responses that modulate unfiltered mirror gazing, for instance by causing positive affective responses in individuals, and negative ones when looking at their unfiltered face (or vice versa). But beyond that, beautified self-images might also elicit other types of cognitive transformations. In particular, I will now explain how exposure to these beautified self-images might lead to a situation in which the altered or distorted mirror image is normalized, potentially leading to body image dissatisfaction and other related phenomena.

To address this issue, I will rely on an account developed by Tremblay and colleagues (2020) that relies on the predictive processing framework. A key idea behind this model of cognition is that perception is an active process that works by predicting bottom-up sensory cues drawn from its best models of what is likely to be causing them. Mathematically, this can be understood in Bayesian terms.[16]

Two core elements of this model are especially relevant for our purposes here. The first is the generative model, whose main task is the prediction of sensory signals, and the second is the precision-estimation mechanism. The generative model is a unified model of acquired knowledge based on previous experiences that are used to predict sensory signals, drawing from the statistically salient history of the agent. The precision-estimation mechanism assigns a probability to a sensory signal in accordance with its estimated degree of certainty or uncertainty. This is a way of minimizing the risks of the prediction task. Mismatches between priors and inputs, i.e., prediction errors, are propagated in the processing hierarchy in order to refine the top-down predictions. This results in an interplay that optimizes top-down priors and precision-estimations.

Self-perception can be understood following this model, namely as an interplay between expectations and sensory signals, both external (e.g., via vision or other perceptual modalities) and also internal (e.g., via proprioception). In other words, self-perception requires integration between a constantly updated self-model and lower-level information. Self-perception requires a constantly updated self-model in virtue of the interplay between top-down predictions (priors), which include beliefs and other attitudes about ourselves,

[16] See Clark 2015; Hohwy 2013.

and bottom-up (sensory) processing (Talsma 2015). For instance, as we age, our body image, which includes sensory evidence from our physical appearance, and other higher-level attitudes, needs to be updated, so that our self-model is as precise as possible. This includes an interplay between visuospatial and semantic aspects (including the body image), but also between sensorimotor and somatosensory aspects (including the body schema).

Tremblay and colleagues (2022) rely on this view of self-perception as self-modeling (which they articulate in depth) to explain the specific types of body image dissatisfaction that interacting with face filters has been said to cause. These authors claim that one's perceptual expectations or priors about one's own facial features is altered by constant exposure to filtered self-images. What happens is that AR-filtered features are normalized through the repeated presentation of such stimuli (filtered facial images), which progressively makes them less surprising, therefore lowering prediction errors. In vulnerable populations (such as teenagers), where there is a lot of pressure concerning social fitting and physical appearance, a distorted facial image can become progressively isolated from corrective sensory evidence. If this process goes on, then the self-model becomes rigid, and what this entails is that the distorted expected self-image is difficult to update, in order to maintain uncertainty minimization.

If we take this into account, and go back to the Distorted Mirror Gazing Hypothesis, we can see that this process of perceptual normalization of the filtered face is a good complement to the phenomenon of affective distorted mirror gazing. This is due to the fact that it explains how an omnipresence of distorted reflections of one's self-image might affect visual face perception, in a way that goes beyond specific forms of affective modulation of experience of mirror gazing. If we add to this the different affective attitudes that influence one's overall mirror experience, we can then see the types of risky cognitive transformations that interacting with this type of filter can give rise to.

There are, of course, some questions left open in the Distorted Mirror Gazing Hypothesis. These include, for instance, the sort of timeframe that the perceptual normalization of the filtered face requires, and how it might vary depending on individual, social, and cultural factors. Also, in order to further understand the cognitive dynamics of this growing phenomenon, it would be important to contrast the relevant time periods for each of the two parts of the hypothesis, and to further explore the relationship between them. This would also be relevant for further understanding the process of affective mirror gazing, as well as the connection between the affective modulation of mirror gazing and the perceptual normalization of self-images. In any case, this hypothesis is a first step towards a

structure that advances our understanding of the cognitive dynamics behind this emerging behavior.

5. Distorted mirror gazing and cognitive enculturation

So far, I have argued that we can think of the perceptual experience of looking at AR-beautified self-images as a form of distorted mirror gazing. The Distorted Mirror Gazing Hypothesis is an attempt to offer a general explanation of the cognitive transformations that interacting with AR beauty filters cause in self-perception. In this last section, I will zoom out and characterize this form of distorted mirror gazing as a particular way in which culture affects cognition.

Let me begin this section by saying that the cultural environment is key to the development of human cognitive abilities. To accept this, we need to deny a strong form of nativism, according to which cognitive development develops largely independently of the environment in which the child develops (Spelke 2022). However, nowadays this position is not widely accepted in developmental psychology, and it is now more and more common to recognize the central role of social learning in cognitive development (Heyes 2017; Menary 2007; Nelson 2007; Sterelny 2012; XXX XXX).

In particular, I want to suggest that looking at AR-beautified self-images, and more precisely, my hypothesis of understanding this as a form of distorted mirror gazing, is a behavior that is part of the process of *cognitive enculturation.* Cognitive enculturation is the process in virtue of which a structured cultural environment transforms human cognition (Menary 2007, 2018). The cultural environment is structured in the sense that learning is organized by cultural practices and social interactions. And cultural practices are distributed patterns of action, which are transmitted both across members of a community and across generations, and they can include techniques, tools and artefacts for different ends.

As Menary (2007, 2013, 2018) has argued, the repetition of cultural practices leads to their embodiment and enactment in behavior. Individuals embody such practices by being subject to neural and bodily transformations. These transformations take place through an extended developmental history, allowing individuals to perform an array of tasks. The most studied cases of neural transformations include the transformation of body schemas and the acquisition of various motor programs that allow cognitive agents to perform different cognitive tasks. For instance, a typical case is the reuse of neural circuitry for finger gnosis for numerical cognition. Here it has been shown that the neural regions for body-schema mapping of finger position overlap with the capacity for identifying numerical quan-

tities (Pinel et al. 2004; Venkatraman, Ansari, and Chee 2005; Anderson and Penner-Wilger 2012).

Given this, I find it reasonable to characterize the act of mirror gazing as a cultural practice for self-perception: It is something young humans learn to do in a structured environment, and which has a transformative effect on their self-awareness, as we saw earlier. Accordingly, I want to invite us to think of AR-distorted mirror gazing as a novel form of the enculturation of self-perception. In this regard, the Distorted Mirror Gazing Hypothesis is a first attempt to analyze the cognitive transformations that this recent cultural practice elicits. But there are more issues that are illuminated once we think of this behavior in the context of cognitive enculturation. To conclude, I will mention three of them.

The first has to do with the specific cultural values that this practice promotes. AR-beautified self-images are altered in ways that are not random, but rather follow specific beauty standards which are homogenizing and exclusionary.[17] For instance, as mentioned earlier, most of these filters whiten one's skin.[18] Beauty filters also make one's nose thinner. Just like in the history of rhinoplasty, this tendency can be said to reveal a strong Anglophilia (Pitts-Taylor 2007, p. 83).[19] So, if AI beautifying filters digitally modify physical features in line with oppressive cultural values, then there is a discussion to be had regarding the extent to which we approve, individually and collectively, of this cultural influence on self-perception. As Medina (2018) argues in relation to racism in visual culture, we should be aware of the subtexts embedded in visual communication, and cultivate a sense of critical responsibility in our production, consumption, and recirculation of images. And, this includes, I want to add, digitally modified self-images.[20]

Though I have tried to give, through the Distorted Mirror Gazing Hypothesis, a general account of the effects that looking at these digitally-distorted self-images can have in self-perception, the use of AR Beauty Filters, is, at least so far, a practice that seems to be directly targeted at women. For example, the names of some of these filters are "La Belle" or "Boss Babe", and it has caught the attention of many specially concerning its potential

---

[17] See https://www.technologyreview.com/2021/08/15/1031804/digital-beauty-filters-photoshop-photo-editing-colorism-racism/ Last Accessed: May 15, 2024.

[18] In this regard, current beautifying filters could qualify as *oppressive things*, namely artifacts that are in congruence with an oppressive system (Liao and Huebner 2020).

[19] This also connects with debates within feminist theory (e.g., Bartky 1990) and Black aesthetics (Taylor 2015)

[20] We might want to abandon or criticize face filters for other reasons, for instance, privacy, biometrics, and the very material production and maintenance of AI. But here, my focus has been exclusively on self-perception.

psychological damages in young girls (Ryan-Mosley, T. 2021).[21] This could entail that interacting with AR beauty filters is a gendered cultural practice for self-perception, and as such, the gendered dynamic and values of this growing cultural practice should be an important part of its critical assessment. For example, it would be helpful to investigate the extent to which AR beauty filters are part of what has been called a culturally imposed form of "feminine narcisissm" (Bartky 1990) that oppresses women through a form of self-objectification, infantilization, and extreme aesthetic pressure. This behavior could thus be considered a new expression of already existing gendered oppression. However, it is also important to notice that there is very limited data concerning what specific filters are popular, how filters are used and by whom (Miller 2024). And, using AR face filters for self-perception can affect, in principle, all individuals regardless of gender.

Relatedly, we can question the extent to which AI beautifying filters are part of a cultural environment that promotes an adequate cognitive development. As we have seen, there are reasons for concern, especially given their connection with body image dissatisfaction and disorders. This could, in some cases, lead to what has been called the cultural *warping* of cognitive abilities (XXXXXX), namely, a detrimental transformation of cognitive abilities due to an unfair cultural environment. The challenge we are left with in this regard is, first, acquiring a deeper understanding of the effect that these technologies might have on cognitive development, and then contrasting this effect with what we take to be an adequate self-perception and body image development.

Finally, seeing the interaction with AI beautifying filters as a new cultural practice for self-perception also illuminates the space left for change and innovation, as changes can be made both in the relevant human behavior and in the technological design. For instance, changes could be made to the AR filters. For example, they could be trained on data that reflects more diverse cultural values and leaves more space for change, thus making them less homogenizing. Second, once we know that interacting with them has a transformative effect on cognition, we could reflect on the very patterns of interaction they promote. For example, instead of using filters simply for beautification, other types of interactions could be promoted, such as self-exploration and discovery.[22] This could boost self-

[21] Thanks to both reviewers in TOPOI for encouraging me to explore the gendered nature of this practice a bit further.

[22] There are already applications like *FaceTune* that allow users to change specific features of their face such as hairstyle and hair color. It is described as designed to "bring out your best self in your photos and videos", however, it's easy to imagine apps that promote behaviors that aim at other goal besides compliance with specific norms of attractiveness.

knowledge, as it might be an interesting way for young individuals to discover things about themselves, such as what they like or dislike, for example.

Let me conclude this final section by drawing readers' attention to Weiss's (1990) analysis of some of the concerns that caregivers and educators previously raised about kids playing with *Power Rangers* (p. 74). Apparently, there were some concerns about the dangers that these toys could pose for young children, and the most pressing concerns had to do with the danger of children identifying with the Rangers' ability to morph from average humans to cyborgs with superpowers. The worry was that the human-like features of Power Rangers, or in other words, the fact that they were not "pure fantasy" creatures (p. 73), dangerously blurred the distinction between fantasy and reality, potentially causing kids to identify with this capacity to radically morph. Weiss's response to these concerns is interesting, and important for our purposes. She writes:

> While educators and parents might be justifiably concerned about the ideological effects of this particular type of transformation, the belief in transformation itself should not be targeted as the problem. (p. 74)

Weiss (1990) makes this illuminating remark to avoid the idea that embodiment (and in our case, cognition) is something given, rather than something open that develops and can change. And as in the case of Power Rangers, I want to suggest that the correct response here is not to directly object to AR beauty filters because of their transformative effect on self-perception, but to point out what is wrong about the particular cognitive transformations that they elicit. As shown in this last section, I believe that characterizing this distorted form of mirror gazing as a novel cultural practice for self-perception (among other possibilities) gives us a well-equipped framework to engage in this evaluative and critical task.

6. Conclusion

Face filters are automated photo editing tools that use artificial intelligence and computer vision to detect human facial features and modify them. In this paper I have focused on so-called *beautifying* filters that allegedly improve a face's physical appearance and attractiveness. The widespread use of these filters has raised concerns due to their potentially damaging psychological effect, especially in relation to their negative effect on one's body image and body appreciation. My aim here has been to unravel the potential cognitive transformations that interacting with AI beautifying filters might elicit.

First, I looked into a fundamental cognitive ability that lies behind looking at beautified versions of oneself, namely face recognition via mirror gazing. I then introduced the Distorted Mirror Gazing Hypothesis, according to which, when looking at beautified versions of themselves, individuals are looking at curated distorted mirrors. This hypothesis has two components. First, this behavior can elicit affective attitudes that change how individuals feel when looking at their unfiltered images (I have called this *affective distorted mirror gazing*). Second, there is a process of perceptual normalization by which the filtered image becomes expected and normalized (I have called this *perceptual normalization of the filtered self-image*). Finally, I have characterized the practice of interacting with face filters as a new cultural practice for self-perception that is part of the process of cognitive enculturation. I have also highlighted some ways in which this practice could be critically evaluated, and some alternative ways in which this type of filter could be used beyond beautification.

References

Andrada, G.; Clowes, R. W., and Smart, P. (2022). Varieties of transparency: exploring agency within AI systems. *AI and Society* 38 (4): 1321-1331.

Bartky, S. L. (1990). *Femininity and domination: Studies in the phenomenology of oppression*. New York, NY: Routledge.

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press USA.

Chinn, L. K., Noonan, C. F. , Patton, K. S., Lockman, J. J., (2024). Tactile localization promotes infant self-recognition in the mirror-mark test,
Current Biology 34 (6): 1370-1375, https://doi.org/10.1016/j.cub.2024.02.028.

De Vignemont, F. (2009). Body schema and body image—pros and cons. Neuropsychologia 48(3): 669–680.

De Vignemont F (2018). Mind the body: an exploration of bodily self-awareness. Oxford University Press, Oxford.

Gallup Jr, Gordon G. ; Anderson, James R. & Shillito, Daniel J. (2002). The mirror test. In Marc Bekoff, Colin Allen & Gordon M. Burghardt (eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. MIT Press.

Gillett, A. J., & Heersmink, R. (2019). How navigation systems transform epistemic virtues: Knowledge, issues and solutions. *Cognitive Systems Research*, 56: 36–49.

Habib A, Ali T, Nazir Z, Mahfooz A. (2022). Snapchat filters changing young women's attitudes. Ann Med Surg (Lond). 2022 Sep 17;82:104668. doi: 10.1016/j.amsu.2022.104668. PMID: 36268310; PMCID: PMC9577667.

Hedman, P., Skepetzis, V., Hernandez-Diaz, K., Bigun, J., Alonso-Hernandez, F., (2022), "On the effect of selfie beautification filters on face detection and recognition",Pattern Recognition Letters 163, pp. 104-111, _____.

Hayduke, A. (2020). The Golden Ratio Within the Human Face and Breast, Ivory Crown Press.

Heyes, C. (2018). *Cognitive gadgets*. Oxford, England: Oxford University Press.

Hohwy, Jakob (2013). *The Predictive Mind*. Oxford, GB: Oxford University Press UK.

Isakowitsch, C. (2023). How Augmented Reality Beauty Filters Can Affect Self-perception. In: Longo, L., O'Reilly, R. (eds) Artificial Intelligence and Cognitive Science. AICS 2022. Communications in Computer and Information Science, vol 1662. Springer, Cham. https://doi.org/10.1007/978-3-031-26438-2_19

Krachun, Carla ; Lurz, Robert ; Russell, Jamie L. & Hopkins, William D. (2016). Smoke and mirrors: Testing the scope of chimpanzees' appearance–reality understanding. *Cognition* 150 (C): 53-67.

Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded*. Hampshire: Palgrave Macmillan.

Menary, R. (2018). Cognitive integration: How culture transforms us and extends our cognitive capabilities. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford handbook of 4E Cognition.* Oxford, England: Oxford University Press: 187–215.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. New York: Routledge.

Miller, L.A. (2024): Preserving the ephemeral: A visual typology of augmented reality filters on Instagram, *Visual Studies*, DOI: 10.1080/1472586X.2024.2341296

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory*. Harvard University Press.

Pandey, A., Prasad, D., Kushwanth Reddy, K., Venkatesh, K., Chand, A., Nath, V. (2023). Face Detection Using Convolutional Neural Network. In: Nath, V., Mandal, J.K. (eds) Microelectronics, Communication Systems, Machine Learning and Internet of Things. Lecture Notes in Electrical Engineering, vol 887. Springer, Singapore. https://doi.org/10.1007/978-981-19-1906-0_55

Quittkat, H.L., Hartmann A.S., Düsing, R., Buhlmann U., Vocks S. (2019). Body Dissatisfaction, Importance of Appearance, and Body Appreciation in Men and Women Over the Lifespan. *Frontiers in Psychiatry*, doi: 10.3389/fpsyt.2019.00864.

Rajanala S, Maymone MBC, Vashi NA (2018) Selfies living in the era of filtered photographs. *JAMA Facial Plastic Surgery* 20(6): 443–444.

Ryan-Mosley, T. (2021). Beauty filters are changing the way young girls see themselves. https://www.technologyreview.com/2021/04/02/1021635/beauty-filters-young-girls-augmented-reality-social-media/. Accessed 16 April 2024.

Schilhab, Theresa S. S. (2004). What mirror self-recognition in nonhumans can tell us about aspects of self. *Biology and Philosophy* 19 (1):111-126.

Sterelny, K. (2012). *The evolved apprentice: How evolution made humans unique*. Cambridge, MA: MIT Press.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333(6043), 776–778. https://doi.org/10.1126/science.1207745

Spelke, E. S. (2022). *What Babies Know: Core Knowledge and Composition Volume 1*. United States: Oxford University Press.

Talsma D (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. Frontiers in Integrative Neuroscience 9: 19.

Taylor, Paul C. (2015). *Black is Beautiful: A Philosophy of Black Aesthetics*. Hoboken: Wiley-Blackwell.

Tramacere, A. (2022). Face yourself: The social neuroscience of mirror gazing. *Frontiers in Psychology. 13: 949211. doi: 10.3389/fpsyg.2022.949211*

Tremblay, S. C.; Tremblay, E., Safae and Poirier, P. (2021). From filters to fillers: an active inference approach to body image distortion in the selfie era. *AI and Society* (1): 33-48.

Vartanian, L. R., Hopkinson, M. M. (2010). Social connectedness, conformity, and internalization of societal standards of attractiveness, Body Image 7 (1): 86-89, https://doi.org/10.1016/j.bodyim.2009.10.001.

Vicente, L., Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports* 13: 15737. https://doi.org/10.1038/s41598-023-42384-8

Weiss, G. (1999). *Body Images: Embodiment as Intercorporeality*. Routledge.

Williams, J. (2018). *Stand out of our light: Freedom and resistance in the attention economy*. Cambridge University Press.

Wittek, N.; Matsui, H.; Kessel, N. ; Oeksuez, F ; Güntürkün, O. and Anselme, P. (2021). Mirror Self-Recognition in Pigeons: Beyond the Pass-or-Fail Criterion. *Frontiers in Psychology* 12.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

Total word count: 8841