Transformative Gazes: Meeting the Other's Eyes in Virtual Reality

The topic of the gaze has received major attention in multiple disciplinary fields. One particular variation of this topic is the *direct* gaze, i.e. a gaze that targets someone else's eyes.

My work concerns the direct gaze in the specific context of the experience of virtual reality; its possibility to trigger transformative effects on its users; and thus its potential incorporation into professional practices that aim at transforming given aspects of the individuals' experience.

In this pre-paper, first I will provide some clarifications about my research field and methodology. Second, I will introduce the notion of direct gaze and frame it within the context of image experience in a broad sense. Then, I will narrow down my scope to the direct gaze in virtual reality. By doing so, I will come to illustrate how the direct gaze can become transformative. Lastly, I will present the research direction that I will explore in greater detail in my seminar presentation.

1. Research Field and Methodology // (Empirical) Aesthetics

When speaking of aesthetics, one can typically refer to two meanings of the word: in the first case, aesthetics roughly indicates the study of beauty and/or the arts; in the second case, in line with its etymological origin (the ancient Greek «aisthesis»), it designates more broadly the study of perception and whatever falls within its domain. Anything that can be accessed by means of the senses, be it beautiful and/or artistic or not, is a suitable object for aesthetics as the study of perception.

My research is framed within aesthetics in this latter sense, and it targets a class of images including, but not necessarily limited to, those produced by virtual reality. Therefore, I am focussing on the perceptual processes triggered by these images. However, as I reject a compartmentalized view of human experience, I am interested as well in any further effects of the perceptual encounter with virtual reality, comprising affective, cognitive, and behavioural ones. If, on the one hand, this orientation is compatible with a traditional philosophical approach, on the other hand it can clearly benefits from contaminations with the cognitive sciences.

On this grounding, when it comes to methodology, my research can follow two complementary paths. First, it can rely on theoretical reflection, thus developing claims that are based on previous works in aesthetics (but also – quite often – art history) and whose acceptability and strength depend on the accuracy of their logical structure. Second, it can seek support in empirical evidence, typically

collected by means of quantitative or qualitative methods from disciplines such as cognitive psychology and the neurosciences. In this latter case, aesthetics becomes *empirical*.

In my work, I often employ theoretical tools to build hypotheses, and empirical ones to test them. While this pre-paper is mostly concerned with theory, my presentation will move further into empirical practice.

2. Background // Being Looked at by Images

As anticipated, we speak of direct gaze when someone looks straight in the eyes of another. The resulting experience of *being looked at*, thus establishing and developing eye-contact, is the core object of my research.[1]

Obviously, this phenomenon is first to be found in real-life encounters. However, and not unfrequently, we can engage in eye-contact with someone who is inside an image: paintings, photographs, posters, movies, and many other types of images can feature characters or entities that look at spectators. What do these images do to us?

First of all, they catch our attention. One may even say, they *demand* our attention. In this regard, art historian Michael Fried (1980; but see also 1967) has proposed a useful distinction between «absorptive» and «theatrical» images. The first ones are characterized by the fact that their protagonists are immersed in activities or thoughts that make them oblivious of the spectators. Theatrical images, on the contrary, not only acknowledge the spectators' presence, but strive actively for being noticed by them.[2] Their characters basically want us to look at them, and one way they can get what they want is looking at us. Chardin's *Soap Bubbles* (1733/1734) and Manet's *Olympia* (1863) are examples – respectively – of absorption and theatricality:

[1] Therefore, I am focussing on the effects that the direct gaze produces on its recipient, rather than sender.

[2] The distinction between absorption and theatricality was first developed with reference to abstract art (1967), which indicates that these categories may have nothing to do with what they *depict*; however, Fried himself later transposed his distinction to figurative art (1980), thus shifting his focus – as I do here – on *characters*.

Once they have obtained our attention, the images that look at us can ask for more. For instance, they can ask that we perform specific actions. This is true, most blatantly, for propaganda and advertising. An illuminating example in this regard, discussed at length by W.J.T. Mitchell (2005), is «Uncle Sam», one of the most famous World War One recruiting posters (together with its endless – and often satirical – variations):

Uncle Sam does not look friendly; rather, he tends to appear as somehow unsettling. In fact, as this poster arguably suggests, images that look at us can become threatening. The issue has been addressed, among others, by Horst Bredekamp (2010), who has described the discomfort experienced by spectators when they realize that not only they are looking at an image, but the image is looking at them as well. Such discomfort is connected with the suspicion that the image may be able to subject the spectators to some kind of mysterious power.

This sort of «power» has been investigated thoroughly by David Freedberg (1989), who has expanded on the relationship between images and their beholders, with a focus on those reactions that are usually deemed as unacceptable by scholars: fear and anxiety are among them. In his chapter on iconoclasm, in particular, Freedberg has noted that many aggressions to images target (and destroy) the eyes of the figures portrayed in them. If iconoclasm can be understood as a reaction to negative feelings induced by certain images, then such focus on the eyes suggests that one key factor that makes these images threatening is precisely their gaze.

However, how can images that look at us actually be perceived as threatening? How can any threat come from something that, after all, is confined to a canvas, a billboard, or a wall?

That images and the figures inhabiting them are separate from us and “cannot touch us” may sound uncontroversial on a logical level; however, it is less obvious when it comes to our concrete

experience. In his essay of profile and frontal figures in visual arts, Meyer Shapiro (1973) has argued that while the former clearly belongs to the space of the image, the latter appear to occupy a space that is virtually continuous with our own. The frontal figure, by turning outwards and looking directly at beholders, may even appear to be about to leave the physical support of the image to join our world. The threshold between these two spaces seemingly become porous, passable. An impression that painters and engravers have sometimes taken advantage of in order to produce more or less playful threatening effects on the beholders:

[3]

At this point, however, if we want to move forward towards the object of my research, we need to abandon the general level of «images that look at us» to start a more medium-specific discussion. In fact, any image is characterized by an original separation from its spectators, which can at some point be challenged by means of a direct gaze. However, the relationship between image space and spectators space is not the same in all media. Hence, the way in which such relationship can be remodelled or even overthrown is not the same either.

3. Topic // Direct Gaze in Virtual Reality

In order to understand the precise relationship between image and spectators (or, better, users) in virtual reality, I propose to rely on the notion of diegetic space and to consider how such space is structured in virtual reality by means of a comparison with cinema.

The term «diegesis» – or rather its French version «diégèse» – was inaugurated by aesthetologist and filmologist Étienne Souriau (1951). Souriau, whose research object was cinema, proposed a basic

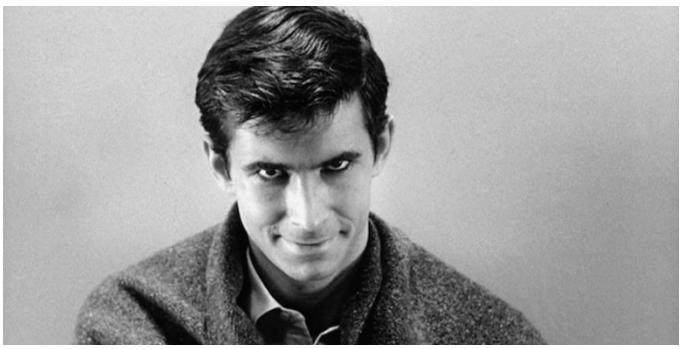
[3] Andries Stock, *Archer with Milkmaid*, 1610, engraving after Jacques de Gheyn. For the topic of the relationship between gaze and arrows, see Krüger (2020). I am thankful to Fabian here for pointing out to me this "archer" class of images that I have just started studying and that I am willing to explore much further!

distinction between two types of filmic space. The first is what he called screen space and it comprises anything that appears visually within the rectangular frame of the screen: light, colours, shapes, and so on. The second type of filmic space, instead, is that in which these flat appearances on screen become objects, events, and characters in the world created by the movie; in few words, where they become invested with *narrativ*e meaning. This, in Souriau's terminology, is the «diegetic space».

The diegetic space, then, is basically the world in which a given filmic narration takes place. For instance, it can be the world where a delusional man called Norman lives in a creepy place named Bates Motel, taking care of his stuffed animals (and mother) (*Psycho*, A. Hitchcock, 1960); or the world in which a problematic man called Royal pretends to be terminally ill in order to win back the love of his equally problematic family members (*The Royal Tenenbaums*, W. Anderson, 2001).

In all cases, and this is what crucial here, the diegetic world is supposed to be autonomous and distinct from the spectators' real world. This separation – which is ontological – is mirrored and reinforced by the spatial separation between screen and spectators' space. Normally, movie characters are very much like Shapiro's profile figures: they inhabit their own space, which they share with other movie characters, and not with us.

Self-enclosure is the norm in cinema, and it is precisely this norm that is challenged when a direct address occurs, that is when a character looks directly at spectators. In this case, the character breaks the «fourth wall» normally separating the diegetic space and spectators space and gives us the impression of temporarily stepping inside our world. This is what happens, for instance, in one of the last shots of *Psycho*, when Norman – completely overcome by his delusion and speaking to himself in his mother's voice – slowly lifts his head as to meet the spectators' eyes:



When a direct gaze occurs in cinema, therefore, two normally separated domains are bridged by the character's look. A variety of more specific effects can follow, depending mostly on the movie's narrative and stylistic context.[4]

[4] I have proposed a typology of direct looks in cinema in Cavaletti (2018).

Things change radically, however, with virtual reality. In fact, due to its technical properties and material usage conditions, virtual reality places the users right at the centre of the world it represents. When wearing a headset, users do not stand in front of the images, but rather they find themselves surrounded by them. Thus, virtual reality nullifies the spatial separation granted by the cinematic screen: spatially, diegetic space and users' space collapse onto each other. The two, however, remain distinct *ontologically*. As such, they also remain mutually exclusive. On this grounding, one of the two must disappear in the users' experience, and leave full room to the other.

The described situation, in my view, gives rise to two possible ways in which the relationship between diegetic space and users' space can be organized in virtual reality.[5] I call them respectively «absolute diegetisation» and «de-diegetisation».

In the first case, diegesis takes over the users' space, which means that users are encouraged to forget their own world in order to fully embrace the diegetic one. By doing so, they somehow become diegetic themselves, as they accept to consider themselves analogous to the characters in the virtual world and often even embody one of them, by means of an avatar. This option is frequently favoured by entertainment virtual reality, and most typically games.

By contrary, «de-diegetisation» occurs when the users' space takes over the diegetic one. Thus, even though the users are objectively surrounded by images, they are supposed to interpret them as if they were part of the real world, as if – more precisely – they were substitutes for objects or situations of the real world. The same holds for characters. So it is not the users to have to become diegetic, but it is the characters to have to be treated as if they were real persons. This set up is typical of "serious" virtual reality, that is all those applications that are employed in professional contexts.

How does all this help understand the effects of a direct gaze in virtual reality? Since, as I have observed, diegetic space and users space collapse onto each other in this medium, a virtual character looking at the users will not bridge any gap between different domains. On the contrary, it will reinforce the idea that virtual characters and users belong to one and the same space, and thus share one and the same nature. Whether characters and users will tend towards diegesis or reality depends on the structure of the virtual experience. However, either way will comprise some kind of *transformation*.

A direct gaze in conditions of absolute diegetisation will tend to strengthen the users' perception of temporarily in being in different shoes and temporarily inhabiting a different environment. Therefore, it will foster the users' deliberate illusion of having transformed into a diegetic character. A direct gaze in conditions of de-diegetisation will maintain the users' focus on their actual self and

[5] These two ways constitute the extremes of an ideal spectrum and they possibly present themselves in more hybrid and nuanced forms in practice. However, I believe that starting from a description of their "pure" form affords heuristically more useful tools when it comes to analysing practice from a theoretical point of view.

actual environment. This does not mean, however, that the users will not undergo any transformation; rather, it suggests that whatever they derive from the interaction will possibly affect their real life.

In my current research, I am focussing on the direct gaze as it occurs in conditions of de-diegetisation. More in particular, I am focussing on a specific class of "serious" virtual reality applications: namely, psychotherapeutic applications. It is in this contexts that, in my view, the direct gaze possibly expresses its transformative potential most fruitfully. However, in order to explain how this can be the case, I need to provide a more precise account of the eye-based interaction between characters and users in the context of the experience of virtual reality.

4. Focus // How the Gaze Becomes Transformative

Why should virtual eye-contact have any transformative effect? Why – in particular – should it have such effect in what I have called conditions of de-diegetisation, since the users basically remains themselves? In order to start answering these questions, I propose to leave aside aesthetics for a while and turn to the cognitive sciences.

Eye-contact is a basic form of interaction. The notion of interaction is at the core of Enactivism, a specific orientation within the cognitive sciences whose basic idea is that our worldly experience arises from continuous and dynamic exchange with our environment and other subjects in it (Thompson 2007). In simple words, we do not exist in isolation from our context. On the contrary, we are constantly and inevitably engaged in interactions with whatever surrounds us, be it another person or simply the surface of the pen we are holding in our hand. Enactivism seeks to account for interaction in a way that allows to understand more precisely how it collaborates to the way we perceive, interpret and perform actions in the world. Intersubjective dyadic interaction in particular has been conceptualized in the enactivist framework by means of the notion of «participatory sense-making», developed by Hanne de Jaegher and Ezequiel di Paolo (2007), with the later support of Thomas Fuchs (2009).

At the individual level, sense-making indicates the set of interactive processes mentioned above, through which subjects – precisely – "make sense" of their environment. Though sense-making is an ongoing activity, one can isolate specific episodes of it: for instance, searching for a book that I have lost in my office. Participatory sense-making indicates the same type of process, but as performed by two or more subjects together, collaboratively: for instance, asking Enrico to help me and searching for my book together. What is worth noting here is that, when this happens, individual experience is transformed: the interaction affords each subject a new understanding of the environment, of the others, and of themselves as well, by disclosing perspectives and meanings that

would not be available to single interactors on their own. This may be why finding my book with Enrico is often easier than finding it on my own.

In participatory sense-making, the gaze plays a key role. In fact, according to the enactivist standpoint, subjects interact by means of bodily gestures, actions, and signals, which include eye-based dynamics. As Fuchs and de Jaegher put it, gazes act as extensions of the body, and eye-contact in particular (or its absence) reveals one of the most intense forms of interaction (2009). Consider a different example: I am now rehearsing my presentation with my roommate Gustav, to see whether my arguments are understandable and solid. If Gustav looks away, I may get the impression that he is bored, and change my tone of voice as to make it more captivating. If he looks at me in the eye, I may assume he is interested in my speech, or try to decipher his expression if I perceive it as enigmatic.

We need, however, to describe the effects of the direct gaze in intersubjective interaction in a more precise way, and at the same time in a way that is not constrained by the hyper-specific context of single examples. In order to do so, we can rely on a dedicated model: the «watching eye effects» model put forward in recent years by Laurence Conty and colleagues (2016).

According to the authors, receiving a direct gaze leads us to engage in «self-referential processing», i.e. the experience that what is happening in a given context is strongly related to one's own person. When someone looks directly at us, our self-awareness is heightened, and we are more likely to refer what happens in the interaction to ourselves. Therefore, if participatory sense-making in general has a transformative potential on one's own experience, participatory sense-making that originates from eye-contact is likely to direct such transformative potential on one's own person.

As this short foray into the cognitive sciences suggests, interactions – and eye-based interactions more specifically – are inherently transformative. Putting oneself in a diegetic character's shoes, as we have seen in the previous paragraph, is certainly a patent form of transformation; however, interacting with a virtual agent while remaining in one's one shoes may induce no less significant (if less blatant) effects.

Are we sure, though, that participatory sense-making and the «watching eye effects» model are applicable to the experience of virtual reality as well? Do they remain valid in a context in which eye-contact does not happen between two human subjects, but between human subjects and virtual characters?

My tentative answer is affirmative and is grounded in the prominent and widely shared account of the experience of virtual reality originally provided by Mel Slater (2009). Slater proposes this experience to rely on two main forms of illusion: place illusion (PI) and plausibility illusion (Psi). The first constitutes the illusion of being in a place (i.e. the virtual space) in spite of knowing that one is actually not there. The second instead consists in the illusion that what is happening in the virtual

space is happening for real, in spite of knowing that this is not the case. When acting together, these two illusions basically make us behave in a virtual environment as we would do in real life.

In order to illustrate his point, Slater proposes an example that fits my purposes very well, as it revolves around an instance of direct gaze:

> Consider in a virtual reality that there is the appearance of a woman standing in front of you. Perceptually there is something there, in the same space as you […] This is PI. Now she smiles at you and asks you a question, and you automatically find yourself smiling back and responding to her question, even though you know that no one is there. [And this – I would add – is Psi] (Slater 2009, 3553).

The «automatism» of the reactions of Slater's hypothetical virtual reality user suggests that the illusionistic power of this medium enables to live experiences that – if not real – as least *work as if* they were real. Therefore, I believe it is acceptable to hypothesize that eye-based interactions in virtual reality are reasonably similar to their real-life counterparts. What remains to be explained is how this becomes relevant in professional contexts, and in the context of psychotherapy more specifically.

5. Research Direction // Direct Gaze, Virtual Reality, and Autism

Up to this point, I have been trying to demonstrate that virtual reality can have a transformative effect based on theoretical arguments. However, this has also been demonstrated in practice, based on the usage of virtual reality as a therapeutic tool.

In recent year, in fact, virtual reality has become a valuable resource in the treatment of several psychiatric disorders and in mental healthcare more broadly (Freeman et al. 2017, Park et al. 2019). Disorders that can be cured or mitigated with the aid of virtual reality include depression, anxiety, and post-traumatic stress disorder. I have discussed elsewhere the advantages and limitations of the usage of virtual reality in the treatment of these disorders (Cavaletti & Grossi 2020). In my current and future research, I intend to explore how virtual reality can be employed therapeutically in relation to a different disorder, in which the direct gaze and eye-contact are central issues: autism, or autism spectrum disorder (ASD).

ASD comprises a range of conditions commonly characterized – among other diagnostic criteria – by impairments in social communication and interaction. People with ASD can find it difficult to engage in conversation, show and reciprocate emotions, use and interpret non-verbal language. Importantly, they also often struggle in sustaining eye-contact (DSM 2015). The latter, however, is vital in social life: just like the rest of our body, and arguably even more, our eyes convey implicit

but crucial meanings, attitudes, and emotions. Looking at the others in the eyes allows us to express ourselves fully, and on the other hand it helps us understanding the others. Conversely, gaze aversion (i.e. avoiding the other's eyes) that is typical of people with ASD can obstruct communication robustly, thus reducing significantly the chances of a full and satisfactory social life (Elgarf et al. 2017).

In addition to psychotherapy, another field in which virtual reality have proved successful is skill-acquisition training. Virtual training is used today in surgery, in the military, in human resources departments. What if the therapeutic and training-oriented usages of virtual reality could be combined, in order to train people with ASD to sustain eye-contact and employ it fruitfully? What if the transformative power of the direct gaze in virtual reality could be used – somehow reflexively – to change precisely the way in which certain individuals manage the direct gaze in real life?

My seminar presentation is meant to discuss and develop this starting idea. By addressing critically some selected case studies, I will elaborate on how the transformative power of the direct gaze in virtual reality may be exploited to improve a specific yet key aspect of the life of people with ASD. More in detail, I will try to tackle and stimulate a discussion around the following main questions:

1. Do we have reasons to believe that virtual reality can be a suitable intervention tool for people with ASD?
2. What can we learn from (some of) the existing virtual reality applications for improving eye-contact in people with ASD, and from the empirical studies that have assessed their efficacy?
3. How is it possible to contribute to advance the state-of-the art in the field?

References

American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition, American Psychiatric Publishing, Arlington (VA).

Bredekamp, H. (2010), *Theorie des Bildakts*, Suhrkamp, Berlin.

Cavaletti, F. (2018), "A Transmedia Overturning: Direct Address from Theatre to Cinema", in *FactaFicta. Journal of Theory, Narrative, and Media*, 2(2), 135-154.

Cavaletti, F., & Grossi, G. (2020), "Take a DEEP Breath: Virtual Reality and Real Anxiety", in L. Malavasi & S. Tongiani (eds.) *Technophobia and Technophilia in the Media, Art and Visual Culture*, Aracne, Canterano.

Conty, L., George, N., & Hietanen, J.K. (2016), "Watching Eyes Effects: When Others Meet the Self", in *Consciousness and Cognition,* 45, 184-197.

De Jaegher, H., & Di Paolo, E. (2007), "Participatory Sense-Making: An Enactive Approach to Social Cognition", in *Phenomenology and the Cognitive Sciences,* 6(4), 485-507.

Elgarf, M., Abdennadher, S., & Elshahawy, M. (2017), "I-interact: A Virtual Reality Serious Game for Eye Contact Improvement for Children with Social Impairment", in *Joint International Conference on Serious Games*, Springer, Cham, 146-157.

Freedberg, D. (1989), *The Power of Images. Studies in the History and Theory of Response*, University of Chicago Press, Chicago & London.

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017), "Virtual Reality in the Assessment, Understanding, and Treatment of Mental Health Disorders", in *Psychological Medicine*, 47(14), 2393-2400.

Fried, M. (1967), "Art and Objecthood", in *Art and Objecthood. Essays and Reviews* (1998), University of Chicago Press, Chicago & London.

Fried, M. (1980), *Absorption and Theatricality. Painting and Beholder in the Age of Diderot*, University of Chicago Press, Chicago & London.

Fuchs, T., & De Jaegher, H. (2009), "Enactive Intersubjectivity: Participatory Sense-Making and Mutual Incorporation", in *Phenomenology and the Cognitive Sciences,* 8, 465-486.

Krüger, K. (2020), "Art is Aiming for the Eye", in Y. Hadjinicolaou (ed.), *Visual Engagements. Image Practices and Falconry,* De Gruyter, Berlin, 215-238.

Mitchell, W.J.T. (2005), *What do Pictures Want? The Lives and Loves of Images*, University of Chicago Press, Chicago & London.

Park, M.J., Kim, D.J., Lee, U., Na, E.J., & Jeon, H.J. (2019), "A Literature Overview of Virtual Reality (VR) in Treatment of Psychiatric Disorders: Recent Advances and Limitations", in *Frontiers in Psychiatry*, 10, 505.

Shapiro, M. (1973), *Words and Pictures. On the Literal and the Symbolic in the Illustration of a Text*, Mouton, The Hague & Paris.

Slater, M. (2009), "Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments", in *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549-3557.

Souriau, É. (1951), "La structure de l'univers filmique et le vocabulaire de la filmologie", in *Revue International de Filmologie*, 7-8, 231-240.

Thompson. E. (2007), *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*, The Belknap press of Harvard University Press, Cambridge (MA) & London.