**Abstract:** How does an aesthetic experience get processed? Building on the work of several other contemporary philosophers, I assume the aesthetic experience to be an instance of the larger category of perceptual experience. But an appeal to traditional perception theories that depend on the mutual exclusivity of perception and cognition won’t explain the aesthetic experience. This is both because they hierarchize perception and cognition, prioritizing one or the other, and because they begin with a static and antecedently defined object and thus struggle to explain the interaction between object and perceiver. The solution is found in an understanding of gist, which is the first approximately 200–300 msec of perception. Gist is whole-scene experience and not of discrete objects. It shares features with both perception and cognition, and can thus serve as the interface between the two faculties, allowing cognition to (at some later milliseconds) affect perception, while at other times (in the first ≤100 ms) giving the perceptual data more directly to the perceiver. Because different kinds of assessment are accomplished at different stages of the gist experience, it can be shown that gist guarantees both a stable, objective reality and leaves room for the flexibility that comes with cognitive penetration. The aesthetic experience is the paradigm example of gist experience that provides for both objective-world stability and socially generated originality.

.

Running Head Right-hand: Gist Experience

Running Head Left-hand: Dena Shottenkirk

15

Gist Experience

Dena Shottenkirk

I walk into The Metropolitan Museum of Art in New York City, make my way up the wide, stone front staircase and go into gallery 635. It is a large-seeming room, with lighting that is neither dim nor bright and about two dozen paintings hanging on the walls. And when I look to my left, I see a row of Velázquez paintings hanging on the wall. I walk over to the middle one.

What do I see? And *how* do I see it? What is the process of that experience?

This is, of course, a subset of a fundamental and reoccurring question in philosophy: How do we experience the world? How, in other words, does the external world come to reside in us?

The traditional view is that there is the act of perception and then there is the process of cognition and that the two procedures are mutually exclusive of one another. In this view, perception provides the basic fodder for thought, but thought *shouldn’t* assist in the construction of perception (e.g., to *think* you see something you don’t in fact see is a hallucination, called such because it is an instance of failed perception), and perception is simply *unable* to assist in thought—sensory experience is restricted to the embodied realm. Objects are to be seen in perception, and then thought is to use that data in abstraction and concept formation.

There have been some recent moves against this traditional view. I agree with this opposition and enter the dialogue by examining gist in perceptual experience, using the aesthetic experience as illustrative of perceptual experiences in general.

1. Gist: The Science

Gist is the first few hundred milliseconds of a perceptual experience, generally defined as perceptual data gained in the first 300 msec. There is wide agreement that what is gained at this moment is surprisingly full. It includes both low-level features and higher-level information. The former includes spatial relations and color, and the latter includes knowledge of the semantic category, e.g., whether it is a street scene or indoor scene and whether the scene is orderly and harmonious. Gist is, in essence, whole-scene recognition, and our brains seem interestingly designed to see this before we see discrete, individual objects. Gist is that part of the perceptual experience that is given immediately and often nonconsciously.

If a scene is presented for a second, it is easy to remember, but if it is shown for only 300 msec, about half of the scene is not remembered (Potter and Levy, 1969: 12). This notion of “remember” was tested by the subjects’ ability to recall the image. The question that arose at this point (e.g., 1970s) was whether those unremembered gist experiences at 300 msec were experienced at all. Using rapid serial visual presentation (RSVP), seminal research tested whether subjects seeing images at <300 msec were understanding them (e.g., recognizing the scene) but unable to remember, or whether they were simply not identifying them in the first place (Potter, 1976: 509). After being exposed to images at varying gist times, the groups were divided into those subjects who were given a test of recognition memory and, on the other hand, those who were asked to search for a target picture (e.g., “boat” as a word or as an image, which then would correspond to one of the RSVP images). In the latter, the subjects’ ability to detect the target was far superior to the recognition memory (518). Thus, we see and understand more than we are able to actively recall. Or, phrased differently, we have nonconscious perception.

The question I am interested in is how that data that is processed in those first 300 msec affects or guides the rest of our perceptual functions, including both the immediately following whole-scene experiences and later object recognition. I propose that some of that initial gist experience is essential to how we frame the rest of the perceptual and cognitive experience. I will show this drawing on the example of an aesthetic experience and demonstrating its reliance on the more general notion of gist in perception.

Let me return now to the immediate issue at hand and further explain some of the most important features of the gist experience. Some perceptual data is reliably obtained at the lower time frames. For example, at 50 msec, subjects were able to reliably discern *unity*or *order* after only a single glance (Schwabe et al., 2018: 18). I use the word “reliably” both because research has shown the near universal ability of participants to judge if a pictorial scene has unity within the first 50 msec and because subjects are unlikely to change their views regarding that particular kind of judgment upon longer or repeated gist experiences. It is a stable judgment, and research has repeatedly confirmed that our bodies are quite good at assessing this particular facet of the scene. In other words, order is a quality that is assessed for early. (Order would be recognized in a scene by the display of horizontals, verticals, grids, stable patterns, etc.) This is not quite true of other qualities. For example, observers who assessed an image as “orderly” or “harmonious” within the first 50 ms were able to assess a pictorial scene according to its *diversity* or *complexity* only after multiple glances at between 500 msec and 5000 msec (Cupchik and Berlyne, 1979: 93). It is important to note the differences in these times. We see and judge orderliness first. It is important to us. To value something on that that basis probably relates to basic safety considerations, e.g., an orderly environment is in contradistinction to a disordered one, which is often a source of threat.

At 100 ms, perceivers can tell whether it’s a street scene[[1]](#endnote-1) and if there are predictably certain kinds of objects, such as cars. This is not the same thing as actually picking out the object of a car, but it is recognizing the semantic category that the scene fits into and thus predicting what else might be in the scene. In other words, at 100 msec, an observer can say “street scene” as opposed to “indoor scene” and thus be willing to predict that the low-resolution shapes in the scene were probably cars. Thus, this early research showed that 100 msec gives the observer the time needed to identify a complex scene in the form of a picture. If one extends gist to include 300 msec, research has shown that the observer would be able fix that picture in long-term memory (Potter, 1976: 521).

As gist is that experience that happens in that first glance that we have of a situation and is preferential to place-recognition, showing that observers can recognize within the first 100 msec, whether it is an indoor scene or an outdoor scene (Fei-Fei et al., 2007: 14), it is useful to stop and ask ourselves “why?” Why would our bodies have this mechanism as an essential part of the process of perception, and what it is that it gives us?

Gist is activated in various situations: When one enters a room, one gets a gist of the room before one identifies specific objects. When one is presented a vista of an outdoor landscape, one sees the layout of the scene and whether it is calm or exciting, and when one is briefly shown a movie trailer, one can recognize the owners of the faces. And though this part of our perceptual process has only been significantly brought to light recently, much research has been done on the surprising amount of data that observers can gather in the 200–300 msec that constitute a “glance.” It is an important part of the perceptual process.

The brain physiology is important to review. But first a few words about the more general field itself. Though there was initially research done on pictorial scene recognition in gist in the late 1930s by Martin Brighouse (Locher, 2015: 78), it wasn’t until the late 1960s and the 1970s, when the seminal and highly influential research by Mary Potter was done, that interest in the field took off in earnest. Developments in technology have given that interest the tools it needed, and more is now known about what physiological parts of the brain are actually activated in the gist experience.

The first fMRI study on the brain locale of gist was published in a 1998 paper (Epstein and Kanwisher, 1998: 598). That research studied subjects who were shown pictures of scenes and pictures of non-scene objects. It showed that different parts of the brain accessed the two different categories: An area called parahippocampal place area (PPA), along with the retrosplenial complex (RSC), was responsible for the scene recognition but not for the object recognition.[[2]](#endnote-2) It is this area, which is on the ventral surface of the brain, that responds preferentially to places and scenes such as houses, outdoor spaces, etc., and is activated within the first 200 msec. There is thus general agreement by all researchers that gist experience is processed in a different part of the brain than is object recognition. The latter is accomplished in the inferior temporal cortex (DiCarlo et al., 2012: 420).

Using gist is thus a distinct way that we parse the data from the external world. It is important to stop for a moment and absorb the import of this. Philosophy has historically been focused on object recognition, particularly interpreted as an analysis of how objects become predicable for us. But if we examine the process of gist perception, a very different perspective opens up.

Gist is holistic perception, an efficient system that has evolved for obvious evolutionary reasons. Gist experience—sometimes called “preattentive perception”—shows great computational efficiency. We see basic scenes very quickly, we determine whether they are indoor or outdoor, we determine whether they are orderly or disordered and we register basic colors and line patterns. Color, for example, has been shown to be present in gist experience at 42 ms of an exposure (Castelhano and Henderson, 2008: 672). Object recognition, on the other hand, takes much longer and requires higher spatial frequency. Spatial frequency is the level of detail present in a visual stimulus per degree of visual angle. An image with lots of detail has a high spatial frequency, whereas a low spatial frequency has less detail. So, a very “blobby” and indistinct image would be low frequency. Thus, the shape of an object is seen at high frequency, while other global features are seen at low frequency. (Oliva and Torralba, 2006: 33) In sum, research over the past 30 years has pointed to the conclusion that gist perception is of the global characteristics of the scene and not of object recognition.[[3]](#endnote-3) Since gist experience is not about object recognition, it gives us a different way to view the problem of how a perceptual experience in general—and how an aesthetic experience in particular—comes to be known by us.

2. Gist in Art: The Experience

Let’s return to the Velázquez. I am going to give a comparative reporting of the experience of looking at a painting. This is, of course, a non-laboratory view of the experience, including both what I (subjectively) take to be the first brief gist experience of the work, followed by the many multiple takes of the perceptual experience as I repeatedly re-look at the work. It is in that way a reporting on what happens to us in ordinary everyday life as we experience the world, particularly when we look at an artwork. The point of this re-telling is to give the experience in as much of a slow-motion take as possible, thus allowing the telling to take into consideration those fundamental first gist moments and to track how the whole experience is controlled by those first few milliseconds. For it is in the viewing of an artwork where it is most evident how the initial gist of the scene takes us in or fails to take us in and how the subsequent repeated lookings follow in that trail. Looking involves a large number of eye saccades, and looking at a visual artwork requires allowing the temporal accretion of those saccades to register their meaning. Some instances of meaning register instantaneously, as in the 50 msec recognition of what we call “order.” Some take longer. What I am about to describe is a re-telling of the actual process of how we look: How the first glance directs us and how we confirm that first gist experience, particularly that of “order.” But it is also a telling of how in some important instances we can, at later stages, absorb different perceptual data that contradicts and denies *some* of the first judgments, making for a perceptual experience that yields more “complexity.” It is a re-telling of how the combination of those things make up the perceptual experience in the experience of an artwork and the role played by gist in that experience. I will argue that the original and first experience of order (vs. disorder) is the guiding one. The entire perceptual experience was timed at six minutes.

As a disclaimer, I should say that I was formerly an art critic, on staff at two magazines, *Artforum* and *Art in America*. I have looked at and written about a lot of art, so this is not the experience of a novice. (SeeSection 4 for a fuller discussion of the research relating to the relative unwillingness of trained observers to express an affective response in a gist experience.)

I walk to the middle painting on the wall. What do I see in my first micro-glances? The painting is slightly smaller than the surrounding paintings. It is more somber in color and more static in composition. The straight shape down the center of the canvas cuts it in half, but there is a small bit of white that firmly and solidly grounds the image by sitting right in the middle of the canvas. That white grounds the image and makes it solid, weighed and orderly but not particularly comforting.

Looking back on the experience and reviewing it, this was the first and instantaneous view that I had. I had, concomitant to this, an immediate “sense” of the picture. It was orderly, as it was primarily a vertical shape down the canvas, but the colors did not quite spell complete calm and reassurance. Gist faculties are a registering of what is frequently referred to by researchers as one’s “gut feelings.” To reiterate, they are also specific to the whole scene, responsible largely for identifying the kind of place we find ourselves in and are not identical with the part of the brain that is primarily responsible for facial recognition. Nevertheless, much gist research has been done on subjects’ ability to, for example, recognize faces, including the faces of famous people, and we’ve learned we are capable of seeing this kind of information very quickly and often very accurately. Even having had said this, though, I wouldn’t say that I knew within the first milliseconds that it was a portrait, but I did know my “gut feeling”: I knew I was drawn to it and that it warranted another look. And I knew that the general image was static, due to its compositional arrangement of the vertical stripe down the center of the canvas, which divided the canvas into a symmetrical and stable composition. It was very static. The composition was almost didactic in its orderliness—but slightly eerie. That was what I knew at the first brief glance.

I am back in front of the image now and in the present. When I glance again, I see it is a portrait. It is the figure that is cutting a vertical swatch through the center of the image. And I notice other things. I notice the diagonal lines that trace the shoulders and notice their symmetry. The image is now even more stable and orderly. It is almost perfectly symmetrical. But then I notice something that is a bit surprising: Despite the fact that he is slightly turned so that his face is at about a 45-degree angle to the viewer, with the left side receding, the shoulders appear evenly presented, the same size and cutting the canvas at the same degrees. This is odd, as perspectival drawing would normally require that the left-side shoulder be smaller than the right. But the shoulders are almost exactly the same. And I then get a reiteration of my initial gist experience of the work: These (wrongly) symmetrical diagonals serve to ground and weight the image even more. It is orderly, but in a slightly frightening way.

His facial expression confirms that. The eyes, focused on the viewer, stare unsympathetically. It is the face of a man about 40, dark-haired, white skin, with mustache and goatee covering part of the skin. The eyes seem at this point to be the main point. Perhaps that is because we always think the eyes are the main point. Perhaps this is my bias, I think. So, I switch to the hair, and it is here that I find the real surprise. The right side of the portrait has the earlobe-length hair painted in a way that we usually see in this time period: The dark undercoat of paint provides the ground for the interspersed lighter colors, giving an overall realist impression of the varying colors that are in most hair colors. It sits still on his head. It is as frozen and as tense as the eyes are.

Then I notice the left side of the hair. It is insane! It is as though it is in movement—as though he is tossing his head and the hair is moving in a blur. The painting of this is utterly bizarre: The hair closest to the face—the vertical strip of color that represents those strands of hair—is the darkest and stillest, so it is very similar to the hair on the right, which is treated in the more usual and expected manner. But next to the “normal” strand of hair there are about a half-dozen different gradations of vertical color as the hair extends out from the face, with the last striation of hair being a thin color almost indistinguishable from the background olive of the walls. The effect is almost like flipping through a color-chart, where brown becomes slightly lighter then slightly more light and slightly more light until it is almost white. The effect is one of both motion and dissolution.[[4]](#endnote-4)

An additionally odd thing: This left side of the hair, which sits on that side of the head that is turned away from the viewer, is dimensionally the same as the other side. In other words, in normal anatomy and the perspectival drawing that adheres to it, the side of the head that is turned away from the viewer should be volumetrically smaller to the viewer than the side that is closer and thus more on display. But in this case, they are the same. The effect is not subtle. The identicalness of their shapes adds to the symmetry and order, but that does not add up to a calming reassurance. It adds up to oddity.

There is more.

The shoulder that is farthest away from the viewer is painted such that there is a hard line between it and the background. This is again unusual, for generally the farther away part is painted so that the edges blur and so recede into the background. The common technique is to take a thin coat of the background color and paint it over just the very edge of the figure. It’s a common technique and generally adhered to. Velázquez has done the opposite. The hardness of the edge makes it jump out because it has come forward to the viewer. In addition, there is a slightly lighter hue to the wall-color right before it meets the shoulder, making it look like it is lit from behind, drawing the viewer’s eye even more to that dividing line. The result is stupendous: It makes the further away shoulder *torque* almost violently toward the viewer.

This is working now in conjunction with the blurred movement of the hair on that same side. The body, seemingly at first glance, stolid and weighted and ordered, actually has half of it moving and spinning.

Finally, I notice the left eye on that hidden side of the face that is in shadow. It is its own thing. It is not still, deadly focused and seemingly unsympathetic, as was the right eye. It seems less sure of itself. It droops slightly down at the outside corner; the pupil is less dark. It seems actually a bit worried. And now my eye returns to the white color in the middle of the canvas that I first noticed in my initial grasp/gist. It is hard, sticking straight out, unyielding. It is a white that is quite colder than the pink tones of the skin. Like a bulls-eye, my eye is drawn to it. I am both anchored and unhinged. And I then note the wall sign: Art historians postulate that this is a self-portrait.

I have given you my perceptual process, moment by moment, in its sequential order. Two things of note: What I have ended up with here is both a reconfirming of my original glance—my gist, as it were (e.g., *the image was ordered*)—and it is also a non-confirming of some of the other perceptual experiences that I had. In other words, I have also given you the ways that some of those initial readings were contradicted. This led to an appreciation of the complexity of the piece, but that complexity maintained itself *within the framework of my initial gist experience*: I knew it was ordered. And that gist I never disavowed, except to learn that the order was challenged: There was contradiction there, and it was a bit frightening. But that, too, was a confirmation of my original grasp of the painting—that was told by the somber qualities of the colors. The eeriness was confirmed. And thus, the original gist was confirmed.

This process of seeing some things at a glance, revisiting the perceptual process and being willing to both confirm and disconfirm various parts of the original gist is an important part of the aesthetic experience. This is even more true in contemporary art since the 1960s. The first glance will give us valuable information, most importantly in those bodily responses of order or harmony and in the general affective “gut feeling” assessment we have: Is it calm or threatening, welcoming or worrying? To fully experience the work, we must look over and over, seeing one thing and then comparing it with another. It is the back-and-forth of the seeing that makes for the experience. And all of it is based on the initial gist experience.

3. Perception: The Philosophy

There isn’t an account of the theoretical model that explains this process. Let me try to present the problem and thereby motivate the search for a solution.

Perception is always defined as the act of being aware of something through the senses, and generally it begins at the point where perception is seen as an openness to the world. In this, this beginning point of openness to the world commits one to describing objects and their properties, e.g., the red apple, and explaining how it is that we come to know that. This shouldn’t seem odd, as the goal—that is, the goal of the game of explaining perception—is to explain how it is that the external world, e.g., the world of things and matter, crosses the barrier of our bodies and becomes fodder for thought. To explain knowledge of objects is thus, presumably, to explain this.

Therefore, the task is usually attempted by following the trail, so to speak, of the objects as they traverse that barrier and trying to track, in that process, their identity: Are those sensorial experiences still material? Or are they now mental? And what does that mean?

But this point of view presents us with a standoff between the perceiver and the object perceived, with both entities antecedently and statically defined and both together presenting the conundrum of how to explain their relationship. There is a large disjunct between the perceiver and the thing perceived, e.g., the object of perception. How do the properties of that object become part of the experience of the perceiver?

The point of view that holds the faculty of perception in charge of absorbing the sensorial data from the world and the faculty of cognition as operating on that bottom-up data and that, in addition, sees these two faculties as mutually exclusive (I will call this the “ME” view), is, as I stated at the beginning of the chapter, been the usual view throughout history. This general point of view has various subspecies under it, but historically they largely divide between those who see perception as dominating the process, e.g., indirect perception theories, and those who see cognition as dominating, e.g., direct perception theories. The former argues à la Hume that it is our perceiving that we fundamentally experience, while the latter argues that our cognitive faculties directly apprehend the (untainted, so to speak) object qua object (see John Searle for an example of this). What they both share is the view that the relationship is hierarchical and is thus so on the basis of the fundamental ontological division between the two participants (despite promissory reductivist claims). It is therefore quite difficult on both accounts to explain how one participant (the object) comes to be in the other (the perceiver).

The middle of the twentieth century saw a change in this within epistemology, with both coherentism and constructionalism arguing for a more dynamic relationship between perception and cognition, though it was followed by computational theories of cognition[[5]](#endnote-5) that returned to a more rigid divide between perception and cognition. The latter saw, once more, perception functioning as bottom-up in support of cognition.

I retell this story in order to illustrate the problem. The problem is how to explain the process of perceptually experiencing the world, and to remember that that problem has been told in terms of processing objects. I also retell the story in order to locate where we are today and so to see what is at stake for us now, for many philosophers have recently approached the problem differently, arguing for a dynamic relation between perception and cognition. This is seen, to name only a few instances, in theories of perception that emphasize predictive processing[[6]](#endnote-6) and in others that discuss cognition’s effects of top-down processing on perceptual faculties in the form of implicit bias.[[7]](#endnote-7)

I would join in on the side that argues for a dynamic model based on gist experience. My argument for that is twofold and is as follows.

First, an epistemological point. I follow in the lines of many philosophers who argue that it is not the case that all of reality comes to us unbidden and on its own terms. Our faculties do not just mirror reality.[[8]](#endnote-8) A known entity sits next to many others that are not so well known or not known at all. There is selection. There is editing. Thus, *all* entities *could be* noticed, they *could be* classified. And they sometimes do get so treated, once they have been pointed out or attended to and now become a part of the individual’s phenomenal experience. We then can, in the future, *expect* the data and *recognize* it when we see it. But those entities that are so selected do leave behind others that are not. And this is an acknowledgment of the cognitive faculty’s effect on the perceptual. Perception is not passive. And it is not entirely conscious. The larger perceptual field includes both nonconscious and non-foveated items.

I would also like to point out that editing is also seen in gist experiences as well as the more conscious states of perception that exist within longer time intervals such as I have just discussed. For example, we recognize the faces of famous people as we flip through the selections of movies online. We know those faces and have come to expect them in the world, hence *we see* that data. We can do that in 200 ms and thus editing happens throughout various levels of perceptual experience, from the millisecond to the longer interactions. The main point that I’m emphasizing here is that we more readily notice things for which we already have a name and/or with which we have had prior experience. (There are those instances of phenomenal awareness that are non-predicative cases, for example, a car’s backfiring or a slap in the face. But I am not interested in those, as they are not relevant to the gist experience.)

That brings us to the second point. And this point is a logical one. What I am calling the ME view is unable to explain how objects come to be in the first place. It is important to trace how this comes to be.

There are two separate but related problems. The second one I take more seriously. Firstly, there is what I will call the development problem: An individual experiences and identifies objects on the basis of matching those experiences with prior ones. For example, this thing before me is an apple and I know that because I’ve seen apples before. But this iterative process must have a beginning, and on the ME model of thought it is hard to postulate what that might be. (It is not so hard on the predictive processing model, as that knowledge can be built through accretion.) Let me explain the problem. If perception brings in the data in its pure sensorial state and then cognition operates on that, it is hard to see how the process begins: How would cognition ever have been able to *begin* the process, if the process itself operates on matching to prior experience? How did the first cognitive experience get its foothold?

The second but more serious problem revolves around naming. This part of the logical problem is also a version of the same problem, e.g., how can the process be given a starting point? In this case, the problem revolves around starting the analysis with the object. The difficulty is that there is no way one can start with the object qua object. For the existence of an object assumes a substratum of past sensorial experiences by individuals who, through their cumulative and shared articulations of those experiences, have acted together to name that object.[[9]](#endnote-9) So, that object only exists *as an object* as *the consequence* of multiple and shared experiences. Those experiences then accrue together in the form of shared linguistic practices, which are multifactorial and can apply to various constituent parts of the object via the identification of general terms that bound their set of instances. For example, we identify the general term “apple” as having, potentially under its umbrella, individual instances that instantiate such predicates as varied as red, yellow, rotten, delicious, etc. That general term has come to exist as a result of the various linguistic practices that have identified the individual instances of “apple” and *those* individual instances exist *as a consequence* of being the source of the sensorial experiences of a perceiver. The particular object did not come first. The object qua object (read: named object) is consequent to—in order of temporal experience: (1) the individual sensorial experience, (2) the shared sensorial experiences and (3) the name attached to the general term. To do an epistemological analysis of how-we-know-the-world with the construct that one starts from the antecedently existing object and then traces how that object becomes an instance of experience to the perceiver is surely to misconceive the relationship. It is not the case that perceptual experience gives us the raw data that our cognition then operates on. The object is not there already. It is to adopt a static model when in fact a dynamic one is needed. What we need is not an analysis of how objects come to be predicable information for us but how it is that we take our experiences—primarily our gist experiences—and convert that into named entities. This is why a dynamic model is called for. It is a back-and-forth arrangement.

Perceptual data exists in a temporal and provisional order. This means that not only is data dependent on previous data but also that later perceptions—as they are dependent upon prior ones—only come into existence as a consequence of the prior ones. Thus, this is not a predetermined or a priori path. In any moment of experience there are many things that are missed. Not all is seen and recognized. This is true when we analyze the contents of gist experience, and it is also true when one looks at perceptual recognition on an object level. Some bits of perceptual data are consciously recognized by us, while others are experienced only nonconsciously, and other bits of data are completely not experienced at all. This triage is the consequence of habitual readings of external reality, social fact recognition and the biases that trail as a consequence of what we allow ourselves to see and recognize and what we filter out. There is a very important distinction between the entities that are perceived by us and the entities that *could* be perceived by us.

The problem as I see it is the point at which the analysis began. If we begin, as do those theorists who identify perception and cognition as mutually exclusive (the MEs), then we are stuck with trying to figure out how the act of perception becomes fodder for thought, a question which then leads us to wonder how the object comes to reside in our mental faculties. This is to ask for trouble. To begin with the object seems like a dead end.

I propose the opposite: *Tracking the experience as it becomes a public object.* Or to phrase it slightly differently, the trajectory I see is not measured in the usual way by starting with the object—ascertained by perception and then operated on by the independent faculty of cognition—and attempting to determine how it is that the object conveys predicable information to the subjective experience of the perceiver. I instead begin *with the gist and track how an individual’s gist can become constitutive of public objects. Gist is what we use to make objects.*

To reiterate the points. Editing happens. Perception is rich. Even in object recognition, very little is foveated, and more entities exist unattended in the maelstrom of un-ensembled perception. The having of a gist experience is the first stop our bodies make in the interaction with the external world. It is the first move in the editing process. I include in the process both the existence of nonconscious perception as well as nonconceptual perception. Gist gives evidence of both. As an essential move in the editing process, gist gives us the basis that we use when we move to the creation/naming of objects. We negotiate entities, we decide to call one set of data “an object”—e.g., a particular name, and we decide that politically and contentiously. We don’t see and/or experience all of reality; we edit, we select, we choose.

Thus, the argument for a dynamic model is that in order to understand how it is a perceptual experience happens (whether an aesthetic perceptual experience or otherwise), we need an understanding of how we name the objects in the first place.

But this creating/naming of objects is not entirely determined by our relativistic or pragmatic concerns; some naming is tethered by the constraints imposed by objective reality. And it is here that gist comes into play in an important way. This is the way that gist gives us what empiricism traditionally wanted from perception. Firstly, research has pointed to the fact that some experiences seem to be automatic for the perceiver and provide both the basis for future experiences and a way of rooting the perceiver in the objective world (Freidman, 1979: 323). Add to this the fact that the gist experience partially yet accurately locates the external world: It does this by asking in the first 50 msec, “Is it orderly?.” Straight vertical and horizontal lines qualify as orderly; horizontal lines alone do not. The latter, especially when they become slightly off-kilter, can seem unstable to us (a tree falling down! a lion darting across the plane!) and slightly threatening (although also often “exciting” and “interesting,” which is in contradistinction from orderly). The “orderly” criterion accurately assesses a scene’s basic characteristics. This we know with accuracy and objectivity. It is thus objective perception of the world.

Secondly, gist also assesses patterns that are social in nature. We recognize famous faces, for example; in that we see what others have told us is important and worth seeing. Thus, cognitive top-down pattern recognition is also seen in the gist experience. It gives us “interesting” and “complex,” though we are more apt to change our minds on these assessments than we are on orderly, or inside scene/outside scene assessments. And this makes sense, as things like “interesting,” “complex” and “a-face-worth-remembering” are largely socially determined. Gist is therefore partly cognitive, able to be influenced by social constructs.

And thirdly, gist is important because it registers our subjective reaction to the data, both the objective data and socially constructed choices. Gist gives one’s own bodily reaction to that fact: Like or dislike/yes or no/desire or aversion; these are all dichotomous choices that reveal our assessment of our relative safety to our surroundings. It is our bodies’ reading of the environment and is an essential ingredient to the emotional component of experience. Value is not “added on” to empirical experience; it is an integral part of it.

The gist experience therefore gives us both an accurate perceptual take on the objective world, allows us to yield to the decisions and preferences of others and to take those editing decisions into our initial experience of the world and finally, gives us our reaction of desire or aversion. It gives credence to those three things.

And thus, it serves as the basis of our experience, and that becomes constitutive of our construction of objects. For it is all three things: (1) the objective world, (2) our social editing of what counts in that objective world and (3) each of our own individual takes on the two former things; together they contribute to the ever-changing process of assigning names to publicly constructed objects.

4. Starting With the Gist, Not Objects

Art is one of those publicly constructed and thus fundamentally important objects .

When I look at that painting of Velázquez’s—that painting that is now thought to be his self-portrait—I see the world not through my eyes anymore, but through his. As close as is possible between humans, I share someone else’s view. I see—at least somewhat, at least a little bit—him, as he saw himself. The experience was a bit shattering.

And I knew that from this first few milliseconds of viewing. My body told me that scene was orderly, but the colors—also known in less than 100 msec—were not particularly soothing. So something was immediately off kilter. My body did not react with total comfort. And yet I was drawn.

We have seen that people are often not aware of their initial gist experiences, and yet they convey vital information to the rest of the perceptual and cognitive processes. While objects are things that can be moved and that we identify as things to be possessed or things to be passed over to others, the gist of a scene gives us not objects but the general layout of the situation. We name objects, but even in those gist moments that we are conscious of we don’t name them. It is important to stop for a moment and think about this.

Objects are things with names. And why do we name some things and not others? I don’t have any names for the various sensations of my digestive processes; this is the problem faced by doctors and it makes them resort to artificial attempts to quantify the experience, e.g., “On a scale from one to ten, how much pain do you feel?” There is an interesting divide between the things we name and the things we don’t name. It seems to me that a fundamental motivation for naming is possession: We name those things that we might want to possess or, more accurately, fight with others over its possession. You have to name something in order to effectively point to it and you have to point to it before you can fight over it.

But the gist of a scene is not one of those things. It is *my* gist, my first reading, my first assessment of *how my body is in relation to this scene.* I don’t have to fight with anyone about this, or even often tell them about it.

But it registers with me. And most quickly determined is: Is it orderly? How does my body stand in relation to this scene? *Am I safe?*

Why was I drawn to the Velázquez? For me, the painting was orderly, but the colors were off, and that combination was instantly attractive to me. I knew from past experience that the combination could be fruitful. There has been empirical research showing that trained viewers were more “unwilling to express an affective response in a gist experience,” more willing to modify their assessment about a work of art than non-trained viewers and more willing to explore the work’s complexities. (Locher, 2015: 78) A “trained” viewer, or what we would call perhaps a connoisseur or a critic, is willing to put to the test some of their previous views and to view and re-view the scene. But someone who has done a lot of looking also knows something that untrained viewers also know: You pretty much instantly know if you want to look again. I have heard and read some complaints recently about how short a time-duration contemporary museum/gallery visitors give over to the looking at an artwork. But the truth is that one or two moments are all that is needed. A second or two. There is a very good reason why we do this. The gist is our first entry into the scene, yes, but it is not only our first entry: It gives us *our reaction to that scene.* It is right that it is called one’s “gut feeling.” It is our gut telling us “yes” or “no.” Look some more or don’t.

The gist, by giving us our gut reaction to the scene, is our quick assessment of liking or not-liking that is captured in those first few milliseconds. It is the pull and push of perception: The desire and aversion of our bodies to the environment in which it finds itself. It is an important component of our emotional reaction. And that reaction of desire or aversion then influences the visual and cognitive processes that follow: When we initially like the gist of the scene, then we look harder—see more details—in discerning the experience and are more willing to accept what we see. We parse the reality differently from those situations that we don’t like.

We confront the world with our bodies. This is the every-second experience of our lives. And since what we confront is only partially absorbed by us, what we do when we confront the world with our bodies is pragmatically edit out the parts not relevant to us at that moment. Naming is part of this process.

Art is a kind of naming. The artist understands the world by selecting and editing out that data that each of us experience as embodied beings and by simultaneously presenting that data within a certain attitudinal framework (e.g., value it *like this!*) the artist hands that vision over to others so that they may similarly experience what the artist experienced. The process then begins anew as it is an experience of the viewer. The viewer comes to that artist’s editing of the world and that experience in turn becomes an instance of (the viewer’s) editing and parsing, selecting and caring.

Gist explains perception because it is not object recognition. Gist also serves to explain perception because it is an instance of a dynamic model. An analysis of perception must begin not with how objects become predicable for us as observers but with the broadly conceived sensed experience of the perceiver. This would include both conscious and nonconscious perception, and most importantly, it would include the gist experience. The emotional “mood” of the experience is the starting point. Objects are not entirely there in toto to be “discovered” in perception. And the view that sees perception and cognition as mutually exclusive is wrong. We do, in part, construct objects. And gist experience is instrumental in this.

The aesthetic experience shows how this is so. An aesthetic experience, like all experience, begins with gist. I take in the initial gist experience and then use that in my further and innumerable glances/views/eye saccades. I add them together, throwing some out, keeping others and doing the kind of bizarre summation that we do. And with that aesthetic experience, I construct the linguistic object—the description of that experience—that I then can add to others’ descriptions of their experience.[[10]](#endnote-10) And the sum of those experiences gets then to be labeled and to be an object of experience that is referred to.

Thus, the aesthetic experience can only be understood within the framework of gist. And while gist also serves to explain perception generally construed, it is in aesthetics where the role of gist is most evident. For it is in the aesthetic experience that we see the pull and push of our bodies when they interact with the world.

Art is a way of bridging the distance between the knowledge obtained by an individual (e.g., the individual artist) and the knowledge adopted by a group (e.g., those who view the art). The viewer has taken that particular selection of data given over by the artist and reexperiences the world in terms of that new data. I see the Velázquez. I see what he has selected out of his immediate world. And now it has become part of my immediate world, controlled largely by my initial bodily readings given by the gist experience. Thus, the aesthetic experience is composed of his initial editing of reality, which in turn is given to me in bodily experience and now is constituent of my experience. But it, of course, does not end here. Others stand in front of Velázquez and others have similar/differing experiences. We compare them. We talk about them. And then we attach names—it is a “solid” composition, it is “slightly unnerving,” etc. Thus, the aesthetic experience—starting with the experience of the artist—parses and edits the plethora of data around us and then, with the experience of the viewer, is constituent of that ever-changing reality that we anchor with names.

Aristotle noted that we get pleasure from learning. Others have noted this as well. But the greatest pleasure doesn’t necessarily come from the most abstract learning. Art does not lead us only into a realm of abstract generalities but rather maintains us in the realm of particular concrete, bodily experience and gives us an understanding of ourselves as embodied, living beings and of others so similarly construed. And it is by paying attention to gist that we come to better understand that.

Notes

References

Bar, Moshe and Aminoff, Elissa (2003) “Cortical Analysis of Visual Context” in *Neuron*, 38 (2), 347–358.

Castelhano, M. S. and Henderson, J. M. (2008) “The Influence of Color on the Perception of Scene Gist” in *Journal of Experimental Psychology: Human Perception and Performance*, 34 (3), 660–675.

Cupchik, Gerald C. and Berlyne, Daniel E. (1979) “The Perception of Collative Properties in Visual Stimuli” in *Scandinavian Journal of Psychology*, 20 (1), 93–104.

DiCarlo, James J., Zoccolan, Davide and Rust, Nicole C. (2012) “How Does the Brain Solve Visual Object Recognition?” in *Neuron*, 73 (3), 415–434.

Epstein, R. and Kanwisher, Nancy (1998) “A Cortical Representation of the Local Visual Environment” in *Nature*, 392 (6676), 598–601.

Fei-Fei, Li, Iyer, Asha, Koch, Christof and Perona, Pietro (2007) “What Do We Perceive in a Glance of a Real-world Scene?” in *Journal of Vision*, 7 (1), 1–29.

Friedman, Alinda (1979) “Framing Pictures: The Role of Knowledge in Automatized Encoding and Memory for Gist” in *Journal of Experimental Psychology: General*, 108 (3), 316–355.

Kanwisher, N. (2001) “Neural Events and Perceptual Awareness” in *Cognition*, 79 (1–2), 89–113.

Li, Li-Jia and Fei-Fei, Li (2007) “What, Where and Who? Classifying Events by Scene and Object Recognition” www.researchgate.net/profile/Li\_Jia\_Li/publication/4301709

Locher, P. J. (2015) “The Aesthetic Experience with Visual Art ‘At First Glance’” in *Investigations into the Phenomenology and the Ontology of the Work of Art* (ed. P. Bundgaard and F. Stjernfelt), Medford: Springer, Cham, 75–88.

MacEvoy, S. P. and Epstein, R. A. (2011) “Constructing Scenes from Objects in Human Occipitotemporal Cortex” in *Nature Neuroscience*, 14 (10), 1323–1332.

Oliva, Aude and Torralba, Antonio (2006) “Building the Gist of a Scene: The Role of Global Image Features in Recognition” in *Progress in Brain Research*, 155, 23–36.

Potter, Mary C. (1976) “Short-term Conceptual Memory for Pictures” in *Journal of Experimental Psychology: Human Learning and Memory*, 2 (5), 509–522.

Potter, Mary C. and Levy, Ellen I. (1969) “Recognition Memory for a Rapid Sequence of Pictures” in *Journal of Experimental Psychology*, 81 (1), 10–15.

Schwabe, K., Menzel, C., Mullin, C., Wagemans, J. and Redies, C. (2018) “Gist Perception of Image Composition in Abstract Artworks” in *I-Perception*, 9 (3), 1–25.

Schyns, P G. and Oliva, A. (1994) “From Blobs to Boundary Edges: Evidence for Time-and Spatial-scale-dependent Scene Recognition” in *Psychological Science*, 5 (4), 195–200.

1. See Aude Oliva and Antonio Torralba, 2006; Mary Potter, 1976. [↑](#endnote-ref-1)
2. See Kanwisher, 2001; Moshe Bar and Elissa Aminoff, 2003; MacEvoy and Epstein, 2011. [↑](#endnote-ref-2)
3. See Schyns and Oliva, 1994; Fei-Fei et al., 2007; Li-Jia Li and Li Fei-Fei, 2007. [↑](#endnote-ref-3)
4. The Metropolitan Museum, in their online entry on “Portrait of a Man,” points out that this painterly effect is not the result of the damage done to the painting before its extensive restoration in 2009, a point with which I agree, as it is consistent with his techniques. The restoration led art historians to view the work as important but to debate whether it is a self-portrait. The Met scholars seem to lean toward believing that it is. That also seems reasonable. [↑](#endnote-ref-4)
5. See Jerry Fodor 1975 *The Language of Thought.* Cambridge, MA: Harvard Press. [↑](#endnote-ref-5)
6. See Andy Clark 2015 *Surfing Uncertainty.* Oxford: Oxford University Press. [↑](#endnote-ref-6)
7. See Tamar Gendler “On the epistemic costs of implicit bias,” in *Philosophical Studies* 156:1 (2011), pp. 33–63. [↑](#endnote-ref-7)
8. See Nelson Goodman 1976 for discussions of this, particularly in *Languages of Art*, Indianapolis: Hackett, p. 8. [↑](#endnote-ref-8)
9. For an interesting discussion of this and related problems, see Bill Brewer 2017 “Consciousness and Content in Perception,” in *Philosophical Perspectives* 31:1, pp. 41–54. [↑](#endnote-ref-9)
10. For a further discussion of language and artists, see Dena Shottenkirk and Anjan Chatterjee 2010 “Those Dumb Artists! Amnesiacs, Artists, and Other Idiots” in *Structural Analysis* ed. Matthew L. Camilleri. Hauppauge, NY: Nova Science. [↑](#endnote-ref-10)