

Two Dynamical Themes in Husserl*

Jeff Yoshimi
UC Merced

Abstract: I describe and partially formalize two aspects of Edmund Husserl’s phenomenological philosophy, in a way that highlights their relevance to cognitive science. First, I describe “constitutive phenomenology”, the study of structures (what I call phenomenological “models”) that constitute a person’s sense of reality. These structures develop incrementally over the course of a person’s life, and serve a variety of functions, e.g. generating expectations relative to actions, and determining the contents of context awareness. Second, I describe “transcendental-eidetic phenomenology”, which posits a hierarchy of laws, each governing the way consciousness must be organized in order for a particular type of thing (a physical thing, a person, a social institution, etc.) to appear.

The German philosopher Edmund Husserl (1859-1938) developed a complex theory of consciousness, which has broad relevance for contemporary cognitive science, and in particular for approaches informed by dynamical systems theory. I reconstruct and partially formalize two broad and forbiddingly titled areas of Husserlian phenomenology: “constitutive phenomenology” and “transcendental-eidetic phenomenology.” Both can be thought of as forms of possibility analysis (analysis of structures in a space of possibilities), and thus both can be understood using concepts from dynamical systems theory, which focuses on possibility spaces or “state spaces”. Though I believe these aspects of Husserl can be naturalized in a specific way (Yoshimi, 2011), my main purpose here is simply to present Husserl’s ideas as clearly as possible, so that they can be drawn on more easily by empirical researchers.¹ In the concluding section I briefly describe how I think these ideas can inform empirical research.

* Penultimate draft of <https://benjamins.com/catalog/aicr.88.08yos>

¹ There are multiple ways one might “naturalize Husserl”, and I don’t want to bias this discussion by precluding any particular approach. Two particularly relevant lines of research in the recent literature are associated with (Balduzzi & Tononi, 2009) and (Fekete & Edelman, 2011). However I make an assumption

“Constitutive phenomenology” is Husserl’s theory of how a person’s overall sense of reality, their “world”, is built up or *constituted* in flowing streams of experience. Think of a kind of internal model developing in layers, as a person has different kinds of experiences in different domains. “Transcendental-eidetic phenomenology” is Husserl’s theory of how, in order to experience particular types of things, certain constraints must be met. These constraints can be thought of as rules governing the way possible experiences must be instantiated if particular types of thing are to appear. The theory is a development of themes that go back at least to Kant.

Given space limitations I will only sketch these ideas here. I will, however, try to be clear about what the open questions are and what work remains to be done.

1. Background

A basic construct, from which all other Husserlian constructs can arguably be derived, is a *phenomenological state space*, a space C of possible conscious states. Each point in C corresponds to a possible human conscious state.² *Conscious states* are understood to be what Gurwitsch (1964) called “fields of consciousness”, which encompass everything a person is aware of at a point in time—thoughts, feelings, the visual field, smells, itches, pains, etc. Various questions arise in relation to the conscious

both groups disagree with (namely, that conscious states at times are determined brain states at times; what is sometimes called “synchronic” mental-physical supervenience). So, were they to naturalize these Husserlian ideas, they would probably do so differently than I do.

² We will focus on conscious states possible for a particular person at a particular “stage of maturity” (more on this below). We can think of this as an accessible subset of C for that person, at that stage of maturity. For simplicity, references to accessible subsets are suppressed here, and I simply refer to C in general. Also note that these ideas can be generalized beyond human consciousness to all logically possible conscious states. For more on both points see (Yoshimi, 2011).

field. For example, it is not clear how extensive it is. Gurwitsch thought we have a fairly rich context awareness, encompassing our sense of our body, some sense of the immediate past and future, etc., while others have claimed that we have a surprisingly limited overall awareness at any time (O'Regan & Noë, 2001). There are also questions concerning the notion of an *instantaneous* conscious state. Perhaps there are no conscious states at times, but only over durations. In that case we can redefine “states” (and hence points in \mathbf{C}) to be brief durations of consciousness.

The question naturally arises of what the mathematical structure of \mathbf{C} is. Is it a metric space, so that we can say that some pairs of conscious states are more similar to one another than other pairs are? For example, two experiences of a tree while walking around it seem intuitively to be more similar to one another than either of those experiences and an experience of skydiving.³ Is \mathbf{C} a vector space, so that we can meaningfully speak of adding conscious states and multiplying them by scalars?^{4, 5} These are open questions, though it is typically assumed that \mathbf{C} is at least a metric space.

³ When I say two conscious states C_1 and C_2 are more “similar” to another than another pair of conscious states C_3 and C_4 are, I mean that the distance between C_1 and C_2 is less than the distance between C_3 and C_4 . “Similarity” in this sense is not to be the same as judged similarity, which cannot be captured by a metric space representation (Tversky, 1977). For example, distances are symmetric, but judged similarities are not. One might agree to “The father is like the son” but not “the son is like the father”. This suggests that \mathbf{C} is not a metric space. On the other hand, it could be that \mathbf{C} is a metric space, but that judgments of similarity are based on a higher order non-metrical process.

⁴ Some have explicitly argued that \mathbf{C} is a vector space or a structure in such a space. For example, (Churchland, 2005) uses vector addition in a color perception space to represent chromatic fatigue. (Stanley, 1999) describes \mathbf{C} as a structure in a vector space, where (modifying Stanley’s ideas a bit), scalar multiplication corresponds to intensification of a state (I see this same tree, but more vividly), and vector addition corresponds to a kind of experiential conjunction (an experience of seeing a tree plus an experience of smelling a rose corresponds to an experience of seeing a tree and smelling a rose). On the other hand, there are reasons to doubt whether \mathbf{C} is a structure in a vector space, but the issue is complex and I will not elaborate on it here.

⁵ Below when I speak of \mathbf{C} as a vector space, I mean to leave open the possibility that it is a subset of such a space (e.g. a hypercube or some other kind of solid). In that case the vector space operations would have to be suitably restricted.

We can represent what William James famously called the “stream of consciousness” as a path in \mathbf{C} , a time-ordered succession of conscious states, something like a trajectory for a phenomenological dynamical system. Whether these paths must be continuous is another open question. Can there be phenomenological discontinuities? Perhaps when we go to sleep or wake up there is discontinuity, though perhaps even in those cases we gradually “fade in and out” of a null state. Can a path in \mathbf{C} can ever cross itself—that is, can one ever have precisely the same conscious experience twice? Husserl denies this, since the temporal context of experience is always changing and so it seems we can’t have the same exact total experience twice.

The space \mathbf{C} is large and unwieldy (compare the state space of an entire brain, which tracks, even on a simple connectionist representation, the activity of billions of neurons).⁶ What is typically of more interest in particular phenomenological contexts are specific *types* of experience—color experiences, bodily experiences, cognitive experiences, emotions, etc. Husserl approaches the issue via his theory of parts and wholes or “mereology”.⁷ Husserl uses this theory to consider how conscious states can be broken down in to various kinds of parts. One part of my current conscious field is my visual field, and one part of that is my awareness of the laptop before me. I also have some sense of my body, some thoughts, etc. Each is in some sense a part of my total consciousness now. What we’d like is a way to talk about *all possible parts* of various

⁶ The dimensionality of a system can sometimes be reduced by observing that relevant behaviors of the system lie on a lower-dimensional subspace of its full space. So, the effective dimensionality of “brain space” could be lower than the connectionist representation suggests.

⁷ Mereology is the study of parts and wholes. Its modern formalization arguably begins with the third of Husserl’s *Logical Investigations* (Husserl, 1970). Husserl wanted to distinguish different kinds of part-structures in the phenomenal field. For example, my experience of the door is “part” of my experience of the whole house. My experience of the door’s color is also a part (but in a different sense) of that experience. For a review of contemporary mereology see (Varzi, 2011).

kinds—all possible visual experiences (all possible visual parts of total experiences in \mathbf{C}), all possible bodily experiences (all possible kinesthetic parts of total experiences in \mathbf{C}), etc. To do so, we can consider *subspaces* of \mathbf{C} .

What exactly is meant by a phenomenological “subspace” is yet another open question. One approach is to treat \mathbf{C} as an n -dimensional vector space, in which case the concept of a subspace is well-defined.⁸ On this view, \mathbf{C} is a product of finitely many subspaces, and we can think of any particular conscious state as having mereological parts which correspond to points in these subspaces. Thus, we can take the space of conscious states \mathbf{C} to be a product of a visual experience space \mathbf{V} , a bodily configuration space \mathbf{B} , a thought space \mathbf{T} , an emotion space \mathbf{E} , and finitely many other subspaces, so that $\mathbf{C} = \mathbf{V} \times \mathbf{B} \times \mathbf{T} \times \mathbf{E} \dots$ ⁹ The idea is that any particular conscious state has a visual component, a component corresponding to bodily experience, some thought component, some emotional coloring, and so forth.¹⁰ To handle cases where a given conscious state has no component of a given kind (for example, no visual component), we can allow that the zero vector of each subspace corresponds to absence of that type of experience. Presumably some of these subspaces can be further decomposed into lower dimensional subspaces (e.g. the visual experience space might be a product of multiple spaces corresponding to color experiences at each location in the visual field).

⁸ On the possibility of representing conscious states as vectors in a vector space see note 4. If \mathbf{C} is not a vector space it could still be thought of as a topological space, in which case subspaces are still well-defined. However, because topological spaces cannot in general be written as Cartesian products of subspaces (e.g. in the case of a Möbius strip), some details of the account given here would have to be modified.

⁹ I am assuming that an n -tuple (v_i, b_j, \dots) in such a product corresponds to a conscious state which contains v_i, b_j , etc. as mereological parts. Also recall from note 5 that I leave open the possibility that \mathbf{C} is actually a subset of the full vector space, which is useful to note here insofar as some combinations of phenomenological parts may not be possible.

¹⁰ This is also consistent with Husserl’s own approach to such problems, which emphasizes “spheres” of experience (e.g. the “sphere” of sensory experiences, of affective experiences, etc.). See (Yoshimi, 2010).

A path in \mathbf{C} —representing an interval of a stream of consciousness—can be projected to any of \mathbf{C} 's subspaces, so that we can consider “sub-streams” of the total stream corresponding to particular aspects of one's unfolding consciousness. For example, if we project a path from \mathbf{C} to \mathbf{V} , what results is the stream of specifically visual experiences within that stream of consciousness. Each point along a projected path in \mathbf{V} corresponds to a visual experience that is a mereological part of the corresponding point in \mathbf{C} . When a path is projected from a higher to a lower dimensional space (as in this case), its topological structure can change. For example, a projection can “fold” a path so that it crosses itself, even if it did not previously cross itself (see figure 1). This fact has phenomenological consequences. Even if Husserl is right that paths in \mathbf{C} don't cross themselves (we never have the same total experience twice because of ever-changing context awareness), this does not necessarily apply to paths once they are projected to subspaces of \mathbf{C} . For example, we can probably have the same visual experience twice by blinking before an unchanging scene, and we are all but guaranteed to have the same color sensations twice at a given location in the visual field (Yoshimi, 2011).

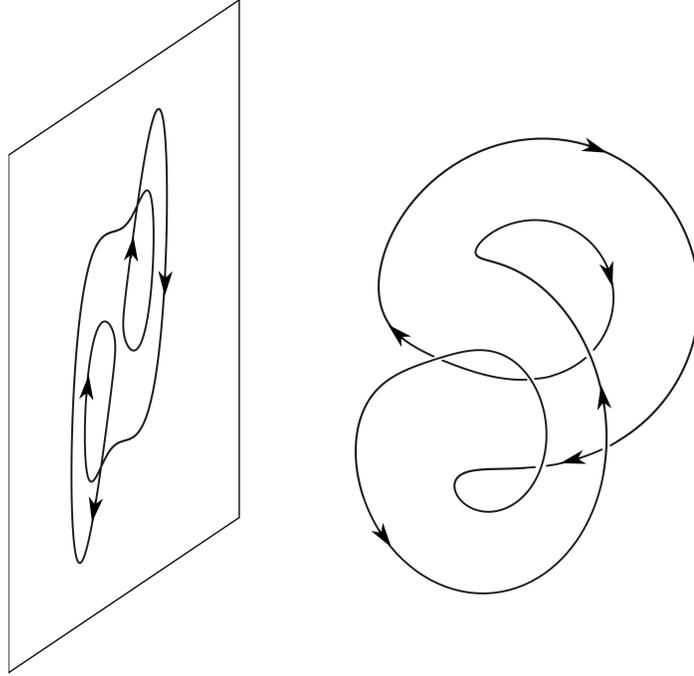


Figure 1: A path in a 3-d space projected to a 2-d subspace. The path in the 3-d space does not cross itself, but its projection to the 2-d subspace does cross itself.

In what follows, it will be important to distinguish \mathbf{C} and its subspaces—which are fixed and unchanging relative to a person’s overall stage of maturity¹¹—from various structures inside of those spaces, which change and morph as a person learns. \mathbf{V} contains all possible visual experiences I could possibly have—it is fixed, for the purposes of these analyses. On the other hand, within \mathbf{V} there are probability relations between these experiences, whereby, depending on where I am and what I am doing, I expect some visual experiences to occur more than I expect others to. These probability relations

¹¹ By a “stage of maturity” I mean a person’s overall level of development, which determines the range of possible experiences they could have. I am at the same stage of maturity I was last week, but I am at a different stage of maturity than I was when I was 1, 5, or 10 years old (the set of possible experiences for my 5-year-old self is different from the set of possible experiences for my current self). Presumably a person’s stage of maturity corresponds to broad anatomical properties of their brain. Thus, if I were to suffer serious head trauma or a massive stroke, what I am calling my “stage of maturity” would change. The purpose of this concept is to allow us to fix the phenomenological state space for a person. Of course, the relevant anatomical properties are themselves gradually changing, so this is ultimately a fiction, but it is a convenient fiction insofar as it facilitates dynamical analyses like these. Similarly, assumptions are often made in dynamical systems analyses, though they are rarely made explicit.

change over time and can be thought of in terms of a changing structure inside of an ambient space which does not change.

2. Horizon Theory / Constitution

Husserl was, like Kant, a transcendental idealist, who believed that reality could be understood as a kind of projection of consciousness. The exact nature of Husserl's position (or for that matter, Kant's) is controversial, but the interpretive controversies are irrelevant here. Our focus will be on certain formal constructs Husserl used in developing his position, that are useful whether or not one accepts Husserl's broader philosophical views, and that are particularly relevant to research in cognitive science. In particular, we will focus on Husserl's concept of the total "horizon" of our understanding of the world, which in some sense "constitutes" reality as we know it. In this section I develop Husserl's account and show how it can be formalized using the tools introduced above.

A basic feature of our phenomenology is that we assume a real world exists and that we are a part of it. The claim seems too obvious to merit stating. To motivate it, consider the contrast between Husserl's focus on our awareness of a *real world outside of us*, and the prevailing introspectionist psychology of Husserl's day, which emphasized inner awareness of one's own sensory states (e.g. an ability to discriminate tones in a chord or colors in an after-image). Husserl is, one might say, more "outwardly" oriented in his phenomenology than the introspectionists were, insofar as he emphasizes our awareness of the external world. Husserl begins by describing the "natural attitude" of everyday life, a kind of default mode of life which is contrasted with various forms of

theoretical reflection. The Husserlian goal is to describe these experiences in the natural attitude as accurately as possible, to capture everyday life *just as it is*. One feature of everyday life, says Husserl, is that we believe the things around us are real—we are naïve realists in the natural attitude. As I sit here writing this, I am at a desk in a room and I assume the desk and room are real things. You reading this no doubt assume that the screen or paper you read these words on is real. Only in the grip of extreme philosophical reflection will someone doubt the existence of the world.

So this is our basic datum: in everyday life we take it for granted that a real physical world exists. Moreover, we have a complex, largely tacit understanding of how this world works. We have a sense of how it is laid out, and where we are in it. We have a sense of how individual things work (a naïve physics). I believe that if I drop my laptop it will break. I believe that I have a physical body that moves in certain ways in this world, and that must be protected from damage. Husserl spent much of his career developing a theory of this implicit understanding we have of the world around us. Rather than describing it in terms of a list of tacit beliefs, he says that our understanding of the world can be understood in terms of a quasi-mathematical structure (Yoshimi, 2009), a “manifold” or “horizon”—what I will call a phenomenological “model”—which corresponds to a kind of picture of a person’s total assumed reality. The job of constitutive phenomenology is to describe this model and to explain how it “constitutes” a person’s overall sense of reality.

Before turning to a formal analysis of horizon theory, here is an intuitive sketch, in terms of a somewhat unusual and mixed set of metaphors. Think of a horizon model as an actual model of the world, spread out on a huge table, like the scale models used in

special effects and miniature wargames. Now think of your current visual experience as a translucent, flexible image—one view of the model—that fuses or morphs in to it, becoming part of it. Your visual experiences over time are successively fused in to this model, and over time the model as a whole becomes more detailed and articulated in familiar areas. In other areas it is sketchy but not empty—even though you’ve never been to North Dakota you still have some sense of what it is like there. If you were actually to visit North Dakota, you would begin to flesh out that part of the model in more detail. Historical maps of the world—detailed in Europe but sketchy in the American interior—come to mind. But this is no static map, it is a dynamical model that becomes more articulated as a person visits different parts of the world.

I now turn to a more formal analysis of Husserl’s horizon theory, whereby these models correspond to structures in subspaces of C . I will not attempt an exhaustive analysis of Husserl’s theory here but rather will focus on three main features. (1) A person’s horizon-model corresponds to a totality of possible experiences, including every experience a person has ever had. (2) This horizon is more or less “delineated” or “articulated” in different regions. (3) As a person has new experiences the horizon is updated. I will focus on visual and bodily experience—which, says Husserl, constitute the physical world as we know it (Husserl, 1997)—and will then briefly consider extensions to other domains of a person’s sense of reality.

(1) A manifold or horizon is a *totality of possible experiences* of a given kind for a person, what Husserl also refers to as “pre-knowledge” that informs our current experiences.¹² Husserl focuses on possible experiences of individual objects (what he

¹² The term “horizon” has several distinct meanings in Husserl, all of which I draw on here. (1) a substrate of background knowledge similar to what (Searle, 1983) calls the “background of intentionality”, which

sometimes calls an “internal horizon”), a “manifold of percepts belonging to the same external thing” (Husserl, 1970, p. 307), but he also generalizes this in various ways, e.g. to all possible experiences of a physical world. We will focus on our visual model of the physical world, which corresponds to a subset of \mathbf{V} , the space of possible visual experiences.¹³ This subset contains, at a minimum, all actual visual experiences you have had within your present stage of maturity. Imagine taking the stream of experiences you have had during this period, a winding path in \mathbf{C} , and projecting it to \mathbf{V} . This projected path doubles back on itself many times in some regions of \mathbf{V} , but has barely ventured in to other regions. The resulting set of points in \mathbf{V} does not, however, correspond to the complete set of possible visual experiences of a person’s world. In particular, it contains experiences a person has not actually had, but that are nonetheless consistent with that person’s overall understanding of the physical world. The total set of visual experiences a person could have of the physical world as they understand it, is their horizon-model of the world.¹⁴

(2) The horizon is more or less “articulated” or “delineated” in different regions. Husserl also describes this in terms of the expectations (what he calls “motivated” or “prefigured” possibilities) we have in different circumstances. When I am in my home town, I have very specific expectations about what I will see next when I make turns on

captures all of a person’s tacit knowledge of the world, (2) a structure in a possibility space, what I have called a phenomenological “model”, (3) a “felt horizon” or some more specific “feeling” of context. When I say “horizon” without qualification I mean sense (2).

¹³ It does not correspond to all of \mathbf{V} , because some visual experiences correspond to random sensory data—like “snow” on a tv—and thus do not present a world. Cf. the discussion of maelstroms in section 3.

¹⁴ The question arises of how to generate the full horizon for a person from those points in \mathbf{V} that a person has actually instantiated. One approach is to try to find a mathematical manifold (or finite set of manifolds) that these points lie on or near. Such a manifold would embed all actual points and contain other points “between” these as well. Finding such a manifold is non-trivial, but corresponds to an active area of research, sometimes referred to as “manifold learning” (Izenman, 2008).

various streets; but when I drive through an unfamiliar town I have much less specific expectations. These ideas can be formalized in terms of a function which associates my current visual experience and bodily movement with what visual experiences I expect to have next. Husserl himself describes such a function:

When I undertake a series of movements in the free system, “I move myself,” the appearances that are arriving are already prefigured. The appearances form dependent systems.... Only through this interplay of independent and dependent variables is what appears constituted as a.. perceptual object... (Husserl, 2001, pp. 51–52).

Notice that Husserl refers to the “appearances that are arriving” as dependent variables, which are a function of independent variables which include bodily movement. We can formalize this in terms of a three-argument “expectation function”, from (a) a person’s overall background knowledge (see note 12), (b) their current visual experience in \mathbf{V} , and (c) a felt bodily movement in \mathbf{B} , to a probability distribution on \mathbf{V} . The probabilities in this distribution can be interpreted as degrees of surprise, where lower probabilities correspond to greater surprise. If I am before a wall and believe there is a window behind me, and I turn around, some resulting visual experiences will surprise me; others will not. If I turn around and see the window I expected to see, I won’t be surprised. On the other hand, if I see no window at all, or a large gaping hole in the wall, I will be very surprised. These correspond to more or less probable visual experiences relative to the probability distribution produced by the expectation function. This in turn allows us to make sense of the notion of more or less “articulated” regions of the horizon-model. In “well

articulated” regions of the horizon, the expectation function will produce relatively sharp distributions, insofar as we have specific expectations relative to different movements; in less well-delineated regions, the function will produce flatter distributions, insofar as we have less specific expectations relative to different movements.

Finally, (3), the horizon is constantly being updated; as Husserl says, it is always in “flux.” Husserl describes this flux in several ways. On the one hand, he refers metaphorically to the acquisition of “sediments” or “precipitates”: as we interact with the world our understanding of it is enriched, in a kind of geological process whereby sediments of knowledge are deposited layer-by-layer on to our current understanding. Less metaphorically, he describes this flux in terms of changes to what I am calling the expectation function defined on a horizon. In particular, he describes a kind of associative learning rule, whereby expectations that are confirmed or “fulfilled” on the basis of our ongoing experiences are strengthened, and those that are “frustrated” are weakened (he also refers to incremental changes in the “weightings” of associative links). This can be formalized by viewing updates following expectation fulfillment or frustration as changes in background knowledge, which in turn alter the probability distributions produced in response to a given action-perception pair. Every time I see something I expect to see, background knowledge is updated in such a way that the probability distribution incrementally sharpens around that expectation. If I turn a familiar corner and see the restaurant that is always there, my confidence in seeing that restaurant is incrementally strengthened, so that next time I turn that corner I will be just a little more confident about seeing it again. Conversely, when my expectations are frustrated, the probability distribution is changed so that whatever I did in fact see becomes more probable. If I turn

the corner and am surprised to see the restaurant has been demolished, I will update my background knowledge, so that next time I turn that corner I will now expect to see the demolished restaurant.

Having outlined and partially formalized Husserl's horizon theory, let us briefly consider some of the work it does in his broader phenomenological theory.

A first function of horizon models is to generate feelings of familiarity. In areas of the horizon-model where the expectation function generates relatively sharp distributions relative to body movements, our surroundings feel familiar. By contrast, in areas where the expectation function generates flatter distributions, our surroundings feel less familiar.¹⁵

A second function of horizon models is associated with (Gurwitsch, 1964), who describes fields of consciousness (points in C) as complex structures involving a focus of attention (or *theme*), as well as a *thematic field*, a form of context awareness involving ideas related to the theme.¹⁶ Gurwitsch says that focal experiences are always surrounded by a thematic field.¹⁷ Items in the thematic field can subsequently be focused on, and are thus *potential themes*. For example, if I am looking straight ahead and see a tree, I have some context awareness of the things around the tree, which I could subsequently focus on. The contents of this context awareness seem to be determined by a person's horizon model of the world, and in particular by the most probable visual experiences generated

¹⁵ Relevant empirical literature includes the psychological literature on familiarity (Yonelinas, 2002), and on skill acquisition and expert performance (Ericsson & Lehmann, 1996). Some related philosophical discussions include Searle's account of the background of intentionality (Searle, 1983) and Dreyfus and Dreyfus' model of skill acquisition (Dreyfus, 1980).

¹⁶ He also describes a third structure, a "margin" of subjective data that are not at the focus of attention, and that are not related to the theme.

¹⁷ A similar, empirically grounded account, with what I think is a plausible story about neural correlates, is in (Koch, 2004).

by the expectation function relative to movements in various directions. What I am focally aware of visually corresponds to a current point in \mathbf{V} , and my context awareness of what surrounds what I am focally aware of corresponds to what I expect I would see were I to move my body in various directions.

I have thus far focused on a person's internal phenomenological model of *physical* reality, insofar as it is constituted by visual experiences and bodily movements. But one's sense of what exists is more complicated than that. Our overall sense of reality includes many kinds of things : our understanding of friends and social groups, of mathematics, of history, of various political controversies and current issues, of fictional characters living in fictional worlds, etc. In each case some version of the whole story told above obtains. In each case it seems we build up a kind of model of the relevant domain over time, and that this model contextualizes our experiences and expectations when we have experiences in that domain. Consider some examples: thinking through a mathematical problem, debating the nature of justice, planning a trip, gossiping, discussing a policy, or evaluating a film. In each case one's moment-to-moment thoughts are embedded in an explorable context of other thoughts and experiences one has had and could have in relation to the topic. In each case there is a distinction between more or less articulated regions of the model (the math I know well vs. the math I hardly know at all; the people I know well vs. people I hardly know), and in the more articulated regions one will feel more confident and familiar than in the less articulated regions. Moreover, just as context awareness when exploring the physical world is determined by a horizon model, so too, it seems, with context awareness when exploring non-physical domains. To take one of Gurwitsch's favorite examples, when engrossed by a mathematical

problem “we may experience references to the possible solutions of the problem... directions in which a solution might be found... [or] assumptions and theorems which might seem to assist in solving the problem” (Gurwitsch, 1964, p. 1). However, there are tricky issues here. When I am solving a math problem, debating a political issue, or thinking about my friends, what are the possible “actions” relative to which expectations are generated? How does emotional valence figure in to this story? Places where bad things happen seem to acquire a negative tinge, and similarly for places where good things happen, but it’s not clear that emotions have their own horizon structure independently of a kind of overlay on other constituted domains. These are, I think, deep and interesting topics, which I hope to explore in future studies.

3. Transcendental-Eidetic Phenomenology

A second form of possibility analysis in Husserl is associated with what he variously calls “eidetic,” “transcendental” or “transcendental-eidetic” phenomenology. The word “eidetic” is a modification of “*eidos*” or “essence”, an unchanging and universal truth (the reference is ultimately to Plato, whose theory of forms emphasized a realm of pure forms). “Transcendental” is a reference to (Kant, 1999), who famously emphasized conditions on the possibility of various kinds of experience. For example, some conditions on the possibility of judging *that a brown table is before me* are, according to Kant, that I be able to perceive things in space and time, and that I understand what extensive magnitudes, intensive magnitudes, and relations are. Husserl combined these ideas in his transcendental phenomenology. He aimed to describe a set of

pure, unchanging “laws” of consciousness, in the form of a set of essences, each of which corresponds to a condition on the possibility of a particular type of experience.

Husserl’s method is to identify essences or laws governing particular types of conscious experiences, using a method called “free variation.” To understand what these essences are and in what sense they “govern” types of experience, it helps to begin by asking what would happen if there were no essences or laws of consciousness at all. If we allowed that a person’s experiences could have any arbitrary character, then subjective life could simply be a procession of chaotic sensory data, what Husserl calls a “maelstrom.” In such a case we would have no experience of a *world*, no sense of an enduring reality, and perhaps no sense of being persons at all. As Husserl says, “it would be a maelstrom so meaningless that there would be no I and no thou, as well as no physical world—in short, no reality” (Husserl, 1997, pp. 249–250). If we are to experience a world—if we are to have a sense of an enduring reality—then subjective data must be *organized* in some way, they must hang together and unfold in certain coherent ways. Elaborating these rules—rules that must be observed if a world is to appear—is the project of transcendental-eidetic phenomenology.

Another way to think of these rules in terms of the movie, *The Matrix*. Recall that the people in the *Matrix* are in casings, and that electrical impulses are pumped in to these casings to produce the illusion of a world. The important point for our purposes is that someone had to program a computer to deliver these electrical impulses in an appropriate way, producing meaningful inputs and responses to motor outputs (for example, when an encased person sends out signals for moving right, an appropriate image should be sent back to her retina), etc. That is, someone had to program a

computer so that it would make the person in the casing believe she was in a world. Someone had to *program a world*. This person, we discover in the third installment of the *Matrix* series, is called “the architect.” The architect had to program rules to prevent people in the *Matrix* from experiencing chaos, he had to force sensory data to appear and to change in response to what people do in a coherent way. These rules are like the rules that Husserl was interested in discovering via his eidetic method.

In practice, Husserl did not focus on high-level, generic rules that prevent subjective chaos altogether (he considered these just to make the general point that having a sense of reality implies the existence of some rules). Rather, Husserl focused on particular types of consciousness, what I will call “domains” \mathbf{D}_i , and then asks, what is a condition on the possibility of having an experience or experiential process in \mathbf{D}_i ?¹⁸ That is, what structures are common to all members of a given domain \mathbf{D}_i ? Moreover, he believed these domains formed in to a hierarchy, each governed by its own eidetic laws. At the highest level are rules which prevent chaos altogether. At lower levels (some of which correspond to special domains Husserl called “regions”), there are rules governing particular types of experience, what Husserl calls “regional axioms” or “syntaxes” (Yoshimi, 2009). Thus we have conditions on the possibility of experiencing visual scenes, unchanging environments, rotations in unchanging environments, interactions with other humans, etc. Continuing the *Matrix* analogy, these laws are like modules of the *Architect’s* program, subsets of the code that govern particular types of appearances:

¹⁸ By a domain I mean some structure constructible from \mathbf{C} . In some cases \mathbf{D}_i is a subspace of \mathbf{C} ; in other cases, it is a subset of a subspace of \mathbf{C} ; in yet other cases it is something more complex, e.g. a set of possible paths in a subspace of \mathbf{C} .

the “physical reality” module, the “other person” module, the “form from motion” module, etc.

Husserl identifies these regional essences by the method of “free” or “eidetic variation.”¹⁹ Free variation can be broken down in to three steps. First, you specify some domain \mathbf{D}_i of interest, for example, possible perceptions, possible color experiences, possible conversations, and so forth. Second, you consider an arbitrary exemplar of that domain and then imagine permutations of that exemplar (being sure to remain within the bounds of the domain in question). In the third and final step you observe what properties $x_1...x_n$ remain constant or “invariant” while you permute the exemplar. These invariants $x_1...x_n$ are the “essences” or “*eide*” of \mathbf{D}_i , and they can be described in the form of phenomenological laws: for any experience in \mathbf{D}_i , it will have the properties $x_1...x_n$.²⁰

There are clearly problems with this approach. In particular, it is not clear how a necessary truth governing a domain of possibilities can be established when one can only consider a proper subset of those possibilities in a finite amount of time. One response to this and similar worries is to simply reject the stronger (and in my view, less plausible) aspects of Husserl’s epistemology and treat his results as fallible hypotheses. In fact, there is evidence that Husserl himself thinks of his project as being as fallible as any human endeavor.²¹ From this standpoint the eidetic method provides a way of framing conceptual / phenomenological hypotheses that are subject to revision or rejection based

¹⁹ For a critical review of the literature on free variation see (Kasmier, 2003). Also see (Yoshimi, 2007).

²⁰ In framing this method, Husserl was influenced by his previous work in mathematics. In fact, his dissertation was on the calculus of variations, a mathematical technique which has affinities with the method of free variation; see (Yoshimi, 2007).

²¹ As he put it in a lecture course: “phenomenology’s scientific articulation, like any theorizing... comes with its own wellspring of deceptions” (E. Husserl, 2006, sec. 22). For further discussion of Husserl and fallibilism see (Hopp, 2009).

on results in other areas.²² It should also be noted that phenomenology, understood in this way, is not much different from empirical psychology.²³ In both phenomenology and in empirical psychology (and the cognitive sciences more broadly), subjective data and other forms of data can jointly support fallible hypotheses.²⁴

Let us consider some examples to see better how Husserl's eidetic method worked in practice.

Let D_p be the set of perceived physical things. What are the essences of perceived things? What is invariant in this domain of C ? What is required in order to experience a physical thing? Take an arbitrary member of this set, say, a ball of wax, an armchair, or a sofa, and imagine variations on it. Let's begin with an armchair. Now change it in some way, for example, change its color—it is a red armchair, a blue armchair, etc. Now change its size, or even let it morph into something altogether different (Husserl is said to have once freely varied a seminar participant into a ball of string). You can permute the imagined object as much as you want, so long as you remain in the set of perceived physical things (so you cannot permute it into something abstract, like a number). No matter how much you change this exemplar, it will remain extended in space, and subject

²² Compare (Mangan, 1991) on “convergent phenomenology”: “The virtue of emphasizing a convergent approach to the study of consciousness is that even if each line of evidence is in itself just plausible, the total ensemble, when it converges on the same finding, can give that finding strong support. The principle is the same in certain kinds of radiation therapy: many relatively weak beams of radiation, converging from different angles, produce an intense dose of radiation” (83-83).

²³ Husserl himself did not think of phenomenology in this way. He clearly thought that phenomenological data has priority over (and is in fact foundational for) the empirical sciences (Yoshimi, 2010).

²⁴ Shimon Edelman (personal correspondence) gives the example of perception science: “All of perception science is rooted in phenomenology in the sense that subjects must rely on their phenomenal states when responding in an experiment... [for example], a subject may be faced with a split display, each half showing a color, the task being to adjust the color of one of the halves so as to match that of the other half.” So, in a sense phenomenology is just a kind of psychology. However, there are differences of emphasis and, moreover, Husserl's theory of consciousness and his methodological procedures are, *de facto*, novel. So there is plenty to mine from Husserl's philosophical output. From this perspective the goal of Husserlian neuro-phenomenology is to integrate his best insights in to the cognitive sciences.

to change.²⁵ Moreover, no matter how much you deform and change it, you will never be able to see more than one side of it at once. Let us call this the “one-sidedness” of perception. Finally, no matter how crazy a perceived physical thing you create in your imaginings, you will never come up with something which has an inside larger than its outside. Let us call this “containedness.” Thus, extension, flexibility, one-sidedness, and containedness all seem to be essences of the class of experienced physical things. We can state these as (again, fallible) laws: if you have an experience of a physical thing, that thing will be experienced as extended and flexible, as being perceivable from one side at a time only, and as having an outside larger than its inside.

In some cases we can frame essential laws that relate one class of experiences to another. For example, we can consider how the set of perceptual representations of physical things (\mathbf{D}_p) is related to the set of imaginative representations of physical things (\mathbf{D}_i).²⁶ Husserl suggests that the two classes are related by a 1-1 map. First, he argues that each to each perception in the set of perceptions \mathbf{D}_p there corresponds an imagination in the set of imaginations \mathbf{D}_i . As he puts it: “there exists an ideal linkage between perception and imagination which assures us *a priori* that to each percept a possible image corresponds” (Husserl, 1970, p. 327). He also claims that imaginations can be uniquely associated with perceptions: “every phantasy [imagination] has the value of a possibility, and it guarantees a perceptual possibility” (Husserl, 1997, p. 252). To show this, two variations can be performed. We first consider an exemplar in \mathbf{D}_p (possible perceptions), in which a thing is directly given, “in the flesh.” For example, my perception right now of

²⁵ Husserl borrowed these particular examples of essences from Descartes, who made them famous via his wax example.

²⁶ Both domains contain representational contents abstracted from variations in attention, intensity, emotional coloring, and so forth.

this computer. It is certainly the case that I can also close my eyes and imagine the same computer. Now I permute that perception so that the computer is larger or smaller, or viewed from different angles, and in every case I can imagine the same thing. So it is an essence of perception that to every perception some imagination corresponds. Now consider an exemplar of \mathbf{D}_I , the set of imaginations. Take any imagination you could have. Say, you imagine a dragon breathing fire. It would not be easy to produce, but in principle we could set things up so you actually perceived the dragon breathing fire. And as we change the dragon in imagination it is always logically possible that the imagined dragon could be perceived in the flesh. Similarly with any imagined physical thing. These two essences suggest a 1-1 map between imagination and perception.

A third example shows how Husserl's ideas are related to ideas independently developed in empirical psychology. Let \mathbf{D}_B be the class of perceived borders, and consider what is essential to such borders. First consider an arbitrary perceived border in \mathbf{D}_B , and then permute that border (being sure that it remains a border through the permutation). Throughout the variation we note that there is always a sudden increase in color change on either side of the border (otherwise we will perceive gradations of color but no sharp boundaries):

A border, indeed, a line results from the leap in quality that occurs... when both parts having been evenly colored... are set off in relief from one another through sharply contrasting colors. A "gradual transition" through a gradually changing qualitative gradation yields a border only if the transition first goes very slowly, and then proceeds very quickly, and then very slowly again (Husserl, 2001, sec. 31).

He also describes this in terms of the rate of color change as a function of space, which suggest a law to the effect that if a boundary is perceived, then the rate of color change at the boundary must be greater than the rate of color change on either side of the boundary. Thus it is part of the essence of perceived boundaries that the rate of color change on either side of the boundary is less than it is at the boundary. This is what allows us to perceive individual objects at all, and by extension, groups of objects, which may be why Aristotle said: “number is perceived by the negation of continuity” (Aristotle, 1961, bk. 3). Husserl is here stating, on phenomenological grounds and in an incipient way, ideas that have since been experimentally investigated and confirmed (Frome, Buck, & Boynton, 1981).

A fourth example involves a more complex domain. In his study of the constitution of space in *Thing and Space* (Husserl, 1997), Husserl considers various forms of bodily movement, which constitute different kinds of spatial structures. For example, he considers the “cyclical manifold” of turning, and says that, if we assume an unchanging environment, then if we rotate in place, the sequence of visual images will repeat (that is, we will have a topological circle in \mathbf{V}). So, our domain \mathbf{D}_T consists of paths in \mathbf{C} , which correspond to turning around once in a stationary environment. Now take an arbitrary exemplar of this domain, e.g. spinning in place in your current environment, while assuming that this environment does not change. If we imagine various changes to the rotation, e.g. in terms of its speed, or what objects are around us, we still get a repeating sequence of visual images. We might call this the “periodicity of

visual images” This periodicity is an essence of \mathbf{D}_T . A projection of any path in \mathbf{D}_T to \mathbf{V} is, on this account, a topological circle.

The eidetic laws governing expectation generation (discussed in section 2)—what Husserl also calls “motivational laws”—provide a final example. We already saw how these laws work above, but let us see how they correspond to essences discoverable via free variation. As an illustrative example, we can focus on a domain \mathbf{D}_S , of perceptions of a stationary physical environment after having looked left and right.²⁷ More specifically, we can take \mathbf{D}_S to contain paths in \mathbf{C} that involve a look to the left followed by a look to the right in what one takes to be a stationary environment. Take an arbitrary exemplar of this class: just look to your left and right and then look forward again. Suppose I see brown walls on both sides. At this point we have the following two “motivated possibilities”: if I look right again, I expect to see a brown wall; if I look left again, I expect to see a brown wall. The brown wall would not surprise me, relative to my current visual experience and background knowledge. Thus, there is an eidetic law that says something like this: any time you look left then right in a stationary environment, you will expect to see what you just saw to the right if you look right again, and similarly if you look to the left again. More formally: any continuation of a path in \mathbf{D}_S that involves a subsequent look left, will also involve an expectation of what was previously seen to the left, and similarly for a subsequent look right. It is impossible to look left and right in what one takes to be a stationary environment and *not* expect to see the same thing again

²⁷ The laws described here are instances of a more general set of laws governing arbitrary movements in various kinds of environments. Husserl himself considers movements of the eyes, head, and body, first in stationary, then in changing environments (Husserl, 1997).

if one looks left or right again. The point seems to hold if we consider different environments, different head rotation velocities, etc. So, it seems to be an essence of \mathbf{D}_S .

4. Conclusion

I will conclude by describing my approach to naturalizing constitutive and eidetic phenomenology, which is based on the assumption that brain states can be associated with conscious states via a “supervenience map”. This map can be used to study the structure of \mathbf{C} , by projecting back and forth between it and the space of brain states \mathbf{B} (Yoshimi, 2011). In this kind of a framework, horizon structures can be studied using neural network models of embodied brains. The state space of such a model is an approximation of the state space of a real brain. In provisional studies I have embedded these models in virtual environments and recorded all their states as they interact with an environment. In this way I can generate simulated “life paths” in an agent’s state spaces. I then try to find a manifold on which these paths are embedded, which can be taken to be the agent’s “horizon” model of its environment. I can also, at least in principle, compute the expectation function described above, and watch in real time as expectations are generated while the agent explores its environment. In these ways it is possible to visualize and mathematically analyze the kinds of horizon-structures Husserl only describes in a qualitative way.

Transcendental-eidetic phenomenology is also relevant to empirical research in cognitive science, as some (focusing on Kant) have recognized (Churchland, 2002; Palmer, 2008). In particular, if there are constraints on possible experience in a given

domain, those constraints should (assuming supervenience) have some analog in the corresponding domain of possible neural activity. That is, structures constraining activity in subspaces of **C** should be correlated with structures constraining activity in corresponding subspaces of **B**.²⁸ This in turn suggests that Husserl's and Kant's conditions on the possibility of coherent experience should be associated with constraints on neural processing, which themselves derive from the structure of the brain and its environment. That is, eidetic laws should ultimately be associated with structures in the brain and its environment. Conversely, it could be possible to identify new phenomenological constraints—that is, new eidetic laws—by studying constraints imposed on neural processing by the brain and its environment.²⁹

References

- Aristotle. (1961). *Aristotle, De Anima. Edited, with introduction and commentary.* (W. D. Ross, Trans.). Oxford: Clarendon Press.
- Balduzzi, D., & Tononi, G. (2009). Qualia: The geometry of integrated information. *Plos Computational Biology*, 5(8), e1000462.
- Churchland, P. (2002). Outer Space and Inner Space: The New Epistemology. *Proceedings and Addresses of the American Philosophical Association*, 76(2), 25–48.

²⁸ See the discussion of “Subspaces and applications” in (Yoshimi, 2011).

²⁹ I am grateful to the editors, Tandra Ghose, Scott Hotton, and an audience of graduate students and faculty at UC Merced, for helpful comments and discussion relating to this paper.

- Churchland, P. (2005). Chimerical colors: some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology*, 18(5), 527–560.
- Dreyfus, S. E. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. DTIC Document.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1), 273–305.
- Fekete, T., & Edelman, S. (2011). Towards a computational theory of experience. *Consciousness and Cognition*.
- Frome, F. S., Buck, S. L., & Boynton, R. M. (1981). Visibility of borders: separate and combined effects of color differences, luminance contrast, and luminance level. *JOSA*, 71(2), 145–150.
- Gurwitsch, A. (1964). *The Field of Consciousness*. Pittsburgh, Duquesne UP.
- Husserl, E. (1970). *Logical Investigations*. (J. N. Findlay, Trans.) (Vol. 1). London and Henley: Routledge.
- Husserl, E. (1997). *Thing and Space: Lectures of 1907*. (R. Rojcewicz, Trans.) (Vol. 7). Springer.
- Husserl, E. (2001). *Analyses Concerning Passive and Active Synthesis: Lectures on Transcendental Logic* (Vol. 9). Kluwer Academic Print on Demand.
- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer Verlag.
- Kant, I. (1999). *Critique of Pure Reason*. (P. Guyer, Trans.). Cambridge Univ Pr.

- Kasmier, D. (2003). *Husserl's Theory of A Priori Knowledge: A Response to the Failure of Contemporary Rationalism*. University of Southern California.
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Company Publishers.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–972.
- Palmer, L. (2008). Kant and the brain: A new empirical hypothesis. *Review of General Psychology*, 12(2), 105–117.
- Searle, J. R. (1983). *Intentionality, an Essay in the Philosophy of Mind*. Cambridge Univ Pr.
- Stanley, R. P. (1999). Qualia space. *Journal of Consciousness Studies*, 6(1), 49–60.
- Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4), 327.
- Varzi, A. (2011). Mereology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011.). Retrieved from <http://plato.stanford.edu/archives/spr2011/entries/mereology/>
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3), 441–517.
- Yoshimi, J. (2007). Mathematizing phenomenology. *Phenomenology and the Cognitive Sciences*, 6(3), 271–291.
- Yoshimi, J. (2009). Husserl's Theory of Belief and the Heideggerean Critique. *Husserl studies*, 25(2), 121–140.
- Yoshimi, J. (2010). Husserl on Psycho-Physical Laws. *New Yearbook for Phenomenology and Phenomenological Philosophy*, 10, 25–42.

Yoshimi, J. (2011). Phenomenology and Connectionism. *Frontiers in Theoretical and Philosophical Psychology*, 2, 1–12. doi:10.3389/fpsyg.2011.00288