

# Empathy and Introspection

## Abstract

Introspection extends from an attempt to isolate and report on an individual sensation such as the color or smell of a rose; to the account of a formal meditation or instructions on how to perform one, undertaken and published by Descartes, St. Augustine, St. Ignatius Loyola, and Pascal, often in the form of an inner dialogue; to the verbal thinking of child's soliloquy as studied by Piaget and Vygotsky; to the unstudied talk of an individual lying back on a couch expressing whatever comes to mind to the psychoanalyst over his right shoulder, who uncannily is merged with the figures from the past; to what Heinz Kohut described as vicarious introspection, a definition of empathy. The latter work in particular suggests that the path to a full, robust understanding and deployment of empathy leads through the labyrinth of introspection, a candidate phenomenon that, according to some, either is unreliable and without its own phenomenology or is epiphenomenal. An engagement in detail will be necessary to clear the way to an understanding of how introspection provides the context for at least some forms of empathic data gathering, those highlighting the microstructure of the interrelation of two individuals. How this works out will require some effort to disentangle, and we begin by taking a step back to the early glory days of introspection and its fall from grace as a scientific method, following through with introspection as displaced perception, simulation, meditation (inner dialogue), free association, retrospection, evenly-hovering attention, and vicarious introspection.

## ***The Disappearance and Rehabilitation of Introspection***

The disappearance of introspection was caused by at least three tendencies. First, the research programs at the psychology laboratories at the turn of the 19<sup>th</sup> centuries at Leipzig (Wilhelm Wundt), Würzburg (Oswald Külpe), and Cornell (Edward B. Titchner) defined the introspective phenomena with admirable precision but with so many constraints as to trivialize them. Second, the science of psychology had advanced faster and further than the discipline of interpretation; and results from the various labs that were at least consistent were misinterpreted as being contradictory, especially as regards the possibility of imageless thought explored at Würzburg. Third, behaviorism exploited the inconsistencies and results, attributing to introspection the ultimately misleading view that introspection had to provide insight into cognitive process that extended "all the way down" - that is, to whatever level was supposed to be foundational, whether neurology or an implausible Cartesian "thinking substance."

The research program of the behaviorists such as Watson and Skinner in the reduction of thinking to subvocalized speech is admirably documented by William Lyons and will not be repeated here.<sup>1</sup> Equally relevant is Gilbert Ryle's (1945) ordinary language analysis and translation of mental concepts and events into the idiom of logical behaviorism, which, in principle, dispenses with the mental ghost in the machine.<sup>2</sup> If Ryle provided the final nail in the coffin of inner perception, the stake through the heart of introspection was the discovery that normal as well as brain damaged individuals fictionalize - "confabulate," "tell fables" - about the reasons for their introspected experiences and retrospected decisions. This was demonstrated by Sperry's split brain experiments as well as those conducted in shopping malls. More about this shortly.

The recovery and rehabilitation of introspection, albeit with conditions and qualifications, is due to new understanding of existing facts and research as well as new results in neurophenomenology.<sup>3</sup> Not all of these trends and counter-trends will be treated in equal detail. However, all are worth mentioning up front as inputs to the dynamic interrelation between empathy and introspection.

More was known about the brain by the time of Gilbert Ryle (1949) than the time of David Hume (1739). However, both these penetrating thinkers were working with a preliminary and inaccurate interpretation of introspection that implied it required access "all the way down" to the microstructure of cognition in order to be useful philosophically. The results of neurology and neurophenomenology have demonstrated that the microstructure of cognition is inaccessible to introspection. This is sometimes called "autoepistemic closure." In spite of the limitations of introspection, the quality and phenomenology of first-person experience is an important aspect of what makes human beings human. Rich realms of meaning and significance are available to first-person experience and invite exploration and inquiry even after behaviorist explanations have

---

<sup>1</sup> William Lyons. (1986), *The Disappearance of Introspection*. Cambridge, MA: MIT Press, 1986.

<sup>2</sup> G. Ryle. (1949). *The Concept of Mind*. New York: Barnes and Noble, 1965

<sup>3</sup> Chapter \_\_\_\_ looks at the use of fMRI technology to explore empathy. These results will not be repeated here, however, the their significance from the perspective of introspection will be explored.

run their course. Phenomenal experience in abundance is left over even after one abandons the attempt to find the deep structure of cognition in a stream of experience that is variously coded as internal versus external.

Meanwhile, a more careful look at the results of confabulation and the (mis)interpretation of the results of the early labs indicates that a limited deployment of introspection is appropriate. If introspection makes use of methods that apply controls and checks and balances, then such an approach may provide a way of rehabilitating first person data access and collection. This will result in correcting the very blind spots about one's own motives to which confabulation points, albeit, once again, with limitations and qualifications.

From a completely different perspective, the work of L. S. Vygotsky in interpreting Piaget's results about the language and thought of the child has expanded the linguistic-friendly interpretation of introspection.<sup>4</sup> Piaget famously focused on the child's soliloquy, expressed as a monologue to no one in particular. This monologue is cited as an example of insufficiently socialized, egocentric use of language. Four and five year old children playing together will make comments about their actions and feelings that are not directed at their playmates and indeed are unrelated to any on-going shared activity. Vygotsky's brilliant reinterpretation, tested in a lab of his own in the USSR, suggested, contra Piaget, that the soliloquy is sufficiently socialized speech but insufficiently individualized thought:

If our hypothesis proves to be correct, we shall have to conclude that speech is interiorized psychologically before it is interiorized physically. Egocentric speech is inner speech in its functions: it is speech on its way inward intimately tied up with the ordering of the child's behavior, already partly incomprehensible to others, yet still overt in form and showing no tendency to change into whispering [contra Watson and the behaviorists] or any other sort of half-soundless speech [ . . . ] We should also have at our disposal an excellent method for studying inner speech "live," as it were, while its structural and functional peculiarities are being shaped; it would be an objective method since these

---

<sup>4</sup> Lev S. Vygotsky. (1934). *Thought and Language*, tr. E. Hanfmann & G. Vakar. Cambridge, MA: MIT Press, 1962.

peculiarities appear while speech is still audible, i.e., accessible to observation and measurement (Vygotsky 1935: 45-6).

In other words, the researcher is witnessing the child's symbolic thought process prior to the child's learning not to vocalize and express it. Contra the Watsonian behaviorists, there is no intermediate stage of whispered and subvocalized behavior since the function of the speech is completely transformed when it is recruited to enact verbal thought, which, as thought, is not vocalized. As the child learns to keep his thoughts to himself, the process of thinking, as it were, goes underground. As this process of learning to keep quiet is occurring and prior to its completion, the researcher is able to hear and see the child "thinking out loud," which is the only way it knows how to think. Thinking silently to oneself comes later.

Vygotsky's account of verbal thinking as a soliloquy that is coded as "inner" evokes a rich philosophical history. It recalls the situation invoked in Plato's *Theatetus*, not as a monologue but as a dialogue.<sup>5</sup> There Socrates describes thinking as an internal dialogue that one has with oneself. Judgment supposedly occurs as the participants in the dialogue agree on a conclusion. Presumably the child's modes of thinking advance to the adult level as one learns to answer oneself and comment on one's initial monologue. Capturing the contents of this conversation as it occurs is one form of attentional awareness; and it is a candidate for introspective access. In a related but distinct idiom, this is also the sense of introspection that Ryle covers (and recovers) in the form of unstudied talk, which, however, is not addressed to anyone and is only contingently private if one happens to keep it to oneself and not share it. I am not proposing to identify Ryle's account of unstudied talk with Plato's (Socrates') inner dialogue; but rather I am asserting these are overlapping descriptions of verbal thinking. Introspection is not reducible to such a verbal process, but it is one form among others.

If one candidly communicates to another individual the results of this verbal give-and-take with oneself, the implications are significant for empathy. One is able to overhear, as it were, to eavesdrop on one's own, and, if communicated, another's personal musings,

---

<sup>5</sup> Plato. *Theatetus*, tr. F. M. Cornford. Indianapolis: Bobbs-Merrill, 1957: 189e.

speculations, and ruminations. The access is first-person, but not exclusively so if one can capture the transcript. This provides rich input to an empathic process that enables the listener to grasp the other's experience at the point where it emerges from a source that is contingently as accessible to the one as to the other. The challenge is that many forms of introspection are available, and this provides access to one among many.

Finally, the work with embodied intelligence, anomalous monism, and neuropsychiatry has opened up a logical space for less well-explored problematics even as problems of the mind-body distinction have been quiesced if not totally solved. The subsequent radicalization of issues about the individual's relations with others has swung the pendulum back in the direction of a willingness to consider introspection as a valid form of data gathering among others. In particular, Heinz Kohut's definition of "empathy" as "vicarious introspection," some form of which is useful and indeed essential in accessing the humanness of the other individual, has necessitated a reengagement with introspection in its diverse forms.

In what follows, much of the give-and-take concerns where to draw the line between introspective processes that are misleading – the equivalent of the bent stick in the glass of water – and those that surface and provide useful inputs to the empathic process, albeit with the necessary conditions and qualifications. By mapping the scope and limit of introspection from within, we shall be able to gain a firm foundation by engaging empathy as vicarious introspection and what that means.

### ***The Many Meanings of "Introspection"***

The challenge is that "introspection" has meant so many things to so many thinkers. As Wundt and Titchner use the term, it means attending to and reporting on a stimulus to one or another sensory modality, i.e., applying an identifying concept to a sensory manifold. In ordinary language, "introspective" as an adjective is more commonly used than "introspection," at least until a neurophenomenologically driven revival of interest in the latter. Introspective people tend to be melancholy, quiet, and given to thoughtful

examination of their own ideas and feelings. Folk psychology uses “introspection” as a search light illuminating the stream of consciousness. This is the “inner perception” that, under diverse interpretations, is touted by Locke, Descartes, and Brentano, and alternatively denounced polemically and translated into logical behaviorist terms by Gilbert Ryle. “Introspection” means accessing and monitoring (and possibly transcribing) the verbal thinking that results after the soliloquy identified by Piaget and Vygotsky goes underground; the dyad of the two-in-one identified by Socrates; and what gets thrown away (perhaps because it never really existed) by Ryle, leaving unstudied talk. Finally, “introspection” is a practice in a spiritual discipline and a literary form. “Introspection” means meditation and the closely related inner dialogue articulated in such literary forms as Descartes’s *Meditations*. These diverse meanings are obviously not a complete list and overlap in many ways so that, for example, Descartes is both the philosopher of inner perception as well as one who also shows up as a player, the ghost, in the internal theatre of the mind where the meditative dialogue is enacted.

Based on this overview, this chapter is inclined to cast a wide net in its engagement with introspection. Introspection is the process of paying attention to the occurrence of thoughts, feelings, beliefs, impulses, emotions, and related content in so far as they are mine. This “mineness” is not a term familiar in ordinary usage, but is easily translated into it. The redness of the apple belongs to the apple and is experienced as being a public, shareable, objective property of the fruit “out there” until the individual examines the experience of redness in its own right. Then the experience of red becomes mine. This may happen as a check on the accuracy of the experience or in order to compare current and recollected experience or to grade and evaluate the quality of the apple or as part of training in color perception or in painting a picture of an apple or for a host of arbitrary other reasons. The distinction is made between the red and the experience of red. Only the latter, the experience, is mine. The point is not whether the experience of red is incorrigible (it is not), derivative (it is), comes later (it does), is captured in a concept and is reidentifiable (it is in the case), is an arbitrary abstraction from and partitioning of experience (seems likely), or any other consideration. The point is that the experience of red is mine and gets coded as “inner” in so far as it is “mine.” This implies nothing about

the red being a sense datum, being reidentifiable (or not), being misleading, or being shareable with others. It does imply that a boundary exists between an individual and what is non-individual, i.e., a world of other things, including other individuals, that are not mine. It does imply that the individual is able to bring to his experience or discover in his experience a distinction that gets encoded as the distinction “inner versus outer.” Often the “innerness” of the inner is able to be translated into “my awareness,” “my experience,” or simply “mineness.” However, the language of inwardness is so fundamental a way of carving up experience that I cannot think of eliminating it altogether in favor of mineness. However, mineness remains as a powerful reminder that individuals are open to the world, access the world as a world of experience to which the world is not reducible without remainder, and usefully distinguish a boundary between the individual and the world even if the individual is a fundamental aspect of the world and vice versa.

Introspection as used in this discussion may include fragments of conversations, recollections of experiences, random words, sensations, and affects. It may also include an internal (not vocalized) monologue or even dialogue where I ask myself questions and provide answers. Why did I do that? Am I proud of my behavior? Am I really that annoyed? For the sake of completeness, recollection will be treated as an intentional attitude alongside believing or imagining or empathizing. If an individual has access to the intentional attitudes that constitute beliefs, emotions, desires, recollections, etc., and if such mental acts have phenomenal properties, then such phenomenal properties are candidates for introspective investigation.

There is a strong sense in which introspection is retrospection, even as applied to the current streaming forth of experience in the window of presence of my experience. The human central nervous system has a latency of about three tenths of a second in transmitting a sensation of a hot stove to the brain as an experience of pain. The reflex that operates within one tenth of a second to remove one’s hand even prior to any awareness of pain is a function of the autonomic nervous system, as is taught to high school biology students. So all our experience is of the past. However, the past that is

recollected in memory is different than what happened three tenths of a second earlier and refers to long term memory. Episodic memory is significantly different than the built-in latency of experience. In episodic memory reminiscences spontaneously emerge based on current stimuli whose association is not always transparent as well as the effortful recollection based on intentional recall.

Introspection extends to retrospection, and, for example, A. Goldman, who is a thinker generally in favor of using introspection as a mode of simulation, invites controversy in stipulating that it does not (Goldman 2006: 246): “I prefer to regard memory retrieval as a nonintrospective cognitive act, one that creates a new working-memory state that can then be introspected.” Anyone who wishes to align with Goldman’s analysis at this point, can use “retrospection” as an abbreviation for this stipulation that retrospection is introspection of contents that refer to the past. This extends from trivial to deep memories. For example, an individual tries to remember where he put his car keys by visualizing coming home, emptying his pockets - where? Onto the dresser? Kitchen counter? Hallway table? He asks, “Where did I last have them? See them? Touch them?” He closes his eyes and tries to visualize coming in the door and emptying his pockets. When a person is asked how many rooms does he have in his mansion, then one way that he counts them is by visualizing them as they appeared when touring the house. However, by the time introspection becomes retrospection, it is often a mulling over of a narrative about the past. Such material is contingently known only to one person, and forms the basis of narratives about the past. For example, no one ever discovered that I was the one who hit the gym teacher in the back of the head with a well aimed spit ball fired from the second story window while all alone. If I had not published that fact here, then I could have taken it to my grave. The history of its effectiveness does not die with me, since the gym teach was also present at the incident even though he did not know the perpetrator. There is a surplus of experience and meaning that is not necessarily accessed or captured in a single narrative told by me. The past is not mine, not mine alone, and parts of it escape from my story. Yet if all the witnesses die and all the archives are burned and lost, then the past becomes a tree in the forest that falls with no one present. It does not make a sound. That sound – the sound of the human past – is its effective



history. We are the effect of what happened – the events, decisions, choices – and can wonder about the multiple reasons and causes that resulted in the present state of affairs. The point is not to produce a philosophy of history; but just to balance the fragility of retrospection and the archives that support it with the surplus of experience and meaning that constitutes history as the objective accomplishment of a community.<sup>6</sup>

Into this context, Heinz Kohut, the psychoanalyst most responsible for self psychology and the related explosion of interest in empathy, has famously said that “empathy” is synonymous with “vicarious introspection.” We will engage what this means after further exploring what is introspection. “Vicarious” experience has been the subject of a detailed drill down that will not be repeated here.<sup>7</sup> In this context, empathy will provide contents that are indeed my own experiences and that satisfy the criterion of being “mine,” but with the modification that these experiences are a vicarious experience of an other individual’s experiences, too, and so refer to what is not mine. As noted, the challenge is that there are many meanings of introspection; and we shall have to ask if a vicarious introspection applies to each of them and in what way.

## **The Unreliability of Introspection**

Those who train to become natural scientists get instruction in how to “read” – that is, use the eyes to grasp – a scientific instrument. This is done in order to reduce statistical variability in the reporting of data points. Researchers learn how to take three separate readings, not the variability, and, in many cases, average the result to smooth out variations and obtain the canonical, scientifically refined value of the variable.

Why should introspection be any more or less coherent than our perceptions of the external world? We have received considerable training in mastering the variability of normal perception of the environment. Our perceptions of the external world show up in

---

<sup>6</sup> F. Baez. (2008). *A Universal History of the Destruction of Books: From Ancient Sumer to Modern Day Iraq*, tr. A. McAdam. New York: Atlas, 2008. The term “Effective history” (*Wirkungsgeschichte*) is Gadamer’s key term for Dasein’s relationship with its (positive) tradition (*Ueberlieferung*).

<sup>7</sup> See Chapter \_\_\_\_ on Heidegger’s special hermeneutic of empathy.

introspection as detached contents, no longer regarded as external objects, but as presentations of external objects to me in my experience.

At this point, it is useful to collect here all of the Cartesian reminders that our senses may lead us astray – the bent stick in the glass of water that is not really bent; rail road lines converging at the horizon that do not really converge; the newly risen moon that seems to follow us along as we drive through the night; and a host of other optical illusions. These are normal functions of perception that, in the limit, create an ambiguity, breakdown, or lead astray the perceiver's interpretive mechanism in a disconcerting way. No amount of explanation will cause the illusion to disappear from perception. The ordinarily essential mechanism enabling object constancy causes the object to look larger since it is interpreted as being further away.

When Galileo first invited the learned scholars of the Italian Inquisition to look through his telescope at the satellites around the planet Jupiter - a veritable solar system in miniature- and the rings around Saturn, these scholars literally did not know how to see what they were supposed to be viewing. They said that what appeared through the device must be the work of the devil. In effect, these viewers said, "This pattern is not mine." And, of course, no one knows for sure what that pattern really was – whether a blur or dancing points of light or exactly what Galileo claimed to see. But they disowned it. It was the work of the devil, operating within the confines of the telescope tube. "Not mine." Not theirs. A new way of seeing was needed in order to make sense out of what was supposedly visible to mere visual inspection. As is well known, the result was that Galileo ended up under house arrest, so where the boundary is drawn, between what is mine and what is not mine, should not be taken lightly.

Why should it be any different with an individual's attention to her or his own experiences as introspected experiences? When combined with my lack of training in managing experiences given in my awareness as my own - I will say "my own" rather than "internal" - the seemingly inevitable "optical illusions" of my own introspective reports distort my beliefs in line with my wishes, skews recollections by my own selfish

interests and desires, and results in unreliable, untrustworthy, biased introspective reports that are weak candidates as input for even the simplest report much less further cognitive processing.

There is an entire school of suspicion – Nietzsche, Freud, Marx, and their followers – that take as the premise of its thinking and philosophizing that the vast majority of what an individual reports about her- or his own experiencing and thinking is not accurate, not complete, misleadingly superficial in relation to the person’s own authentic wishes, desires, and hopes. Though one doubts Nietzsche, Freud, or Marx would care about the split brain experiments and research on confabulation in shopping malls that will be discussed shortly, such results tend to confirm their suspicions and conclusions about the limited accuracy of the individual’s initial assertions of self-knowledge. However, “limited accuracy” is significantly different than “always wrong.” It is ironic that Nietzsche’s ruminations were almost entirely introspective; and Freud was committed to transforming the contents of the unconscious into consciousness by a process of interpreting introspectively surfaced free associations and evenly-hovering attention. In particular, Freud will be the subject of further drill down in the interest of “fixing the flux” of introspection in such a way as to make it accessible to other individuals in empathy. However, the way lies through the ascent from introspecting fine-grained sensations and affects to verbal thinking and individual thinking as the dialogue of the two-in-one. In principle, one might proceed either top down or bottom up; but we shall start at (or near) the bottom.

### **Wundt’s Attempt to Improve the Quality of the Data Provided by Introspection as Reaction to Physical Stimuli**

Just as scientist train in how to read a pressure gauge or measure a quantity of liquid in a Pyrex cup by reading from the bottom of the meniscus, so too work and discipline were required and researchers would benefit from training in introspective data capture. Such was the conclusion of Wilhelm Wundt who resolved to provide for introspective experiences a rigor and discipline analogous to that targeting the physical sciences. This

would in effect clean up introspective data and make it suitable for scientific employment. However, while the sciences of perception and psychology were advancing rapidly, that of interpretation was not.

In short, Wundt's research program and observational framework did not extend to the phenomenal correlates of mental acts such as calculations, manipulations of symbols, and logical inferences. This was the very stuff that engaged the Würzburg school of imageless thought for whom the phenomenal quale accompanying intellectual operations were most engaging. In contrast, for Wundt, his laboratory apparatus and results were designed to capture the results of sensory stimuli. Amidst the political battles between the two various laboratories, the main casualty was the truth or at least the would-be consistency of the data. The outcome of Wundt's efforts at his and separate labs in America was the empirical results from different labs were misunderstood as being contradictory when more attention to the explanatory framework, analysis of variables, and operational procedures would have shown the results to be consistent and even mutually complementary. After reviewing the work of the Würzburg introspectionists and the results generated at E. B. Titchener's Cornell lab, two present-day researchers, Monson and Hurlburt, found that the two laboratories actually reported similar observations, but these were given different interpretations by the scientists (Goldman 2006: 229).<sup>8</sup> The work of Monson and Hurlburt has been treated like a correction to a front-page inaccuracy in a leading newspaper - placed bottom right on a page in the middle of the paper. Thus, the engaging work of Monson and Hurlburt has received less notice than the initial "scandal" about the "inconsistent" results on imageless thought.

Even without the fiasco resulting from basic misinterpretation, Wundt's excruciating attention to detail in the name of science contributed, paradoxically enough, to the devaluation of introspection. The constraints applied to perceptions and reports of

---

<sup>8</sup> C. K. Monson and R. T. Hurlburt. (1993). "A comment to suspend the introspection controversy: Introspecting subjects did agree about 'imageless thought'" in R.T. Hurlburt, ed., *Sampling Inner Experience in Disturbed Affect*. New York: Plenum. "Imageless thought" was championed by Oswald Külpe's Würzburg school in contrast to that of Wundt, who regarded physical sensations and affects as the foundation for cognition. Of course, these groups were talking past one another. Külpe's approach to mental acts would at least indirectly support Brentano and Husserl.

sensations squeezed the context out of experience and, in effect, the life out of it. The greater part of what falls under the science of psychology – the interesting mental phenomena such as thoughts, emotions, volitions – could not be studied by the experimental method of examining inner sense:

. . . [T]he first psychological apparatus, that came to birth with the first psychological laboratories . . . were, for example, the tachistoscope, a machine for presenting visual stimuli for very brief and accurately timed periods, and the metronome or its more sophisticated relative, the chronography, an apparatus for accurately recording reaction times, such as between, say, the presentation of the stimulus and the introspective report on the ensuing sensation.

Wundt believed that he had so hedged in and controlled introspection that it could be admitted into psychology as scientific inner perception. It might be more accurate to say that he had reduced introspection to inner perception, for he believed that the more robust and complex mental phenomena, such as thoughts, volitions, and feelings, were not sufficiently amenable to experimental control to be the object of scientific inner perception [ . . . ] In making psychological investigation scientific and experimental, Wundt was accused of trivializing it and making it impossibly boring [e.g., by William James]. Further it could be argued that Wundt had so hedged in and reduced introspection and introspective reports that it was no longer clear how they differed from simply reporting on ordinary perception (Lyons 1986: 5-6).<sup>9</sup>

Introspection was constrained by rules about what to report, how to report it, what to avoid in reporting on the stimulus, what not to do (such as inferring what should have been introspected), and a list of rules too long to be cited here. The more conditions and qualifications that were specified, the more bizarre introspection became, and the more distant from anything that any ordinary person might do or label as “introspective.” Introspection barely survived its transformation into a tool for the scientific investigation of psychology and the physiological processes supporting perception.

A review of the approach to introspection of the classic psychology laboratories of Wundt and Titchner shows they were investigating the introspection of physical elements and contents of space and time as they impact awareness. For Wundt - and his colleagues and competitors - introspection was far from being a single, unitary phenomena. But if one had to say what most characterized the approach of Wundt, then it would be that

---

<sup>9</sup> William Lyons. (1986). *The Disappearance of Introspection*. Cambridge, MA: MIT Press, 1986.

introspection was capturing a reaction to a physical stimuli. When all is said and done, the operational protocols developed did well in unambiguously defining reporting on sensations of the kind that the logical positivists proposed as sense data. This does not mean the experiment identified the elusive (and illusive) sense datum. It means that, given a consistent method of delivering stimuli to different sense organs, consistent and repeatable reports applied to sensations will be forthcoming. Tachistoscopes, chronographs, and, just as importantly, clipboard, pen, and paper, were deployed to “fix the flux,” i.e., render the individual’s experience of and reaction to sensory stimuli stable and repeatable. The mountains labored, a mouse was born. At no point were persons asked to provide an assessment of their reasons, motives, intentions or complex psychological states in decision making. No decision was being contemplated. However, had the researchers done so, the results would not necessarily have been any different, even without the conflict of interpretations lurking in the background and exploding, than the results produced years later, in experiments in confabulation.

### **Introspection as Confabulation**

Confabulation, the tendency of individual persons to invent a reason retrospectively for a choice made, is an important constraint on the accuracy of introspectively collected data and reports. That some first-person reports are wildly unreliable and inaccurate is highly probable. Basically, “introspection as confabulation” taken literally means that introspection, even if a real enough phenomenon as inner input, is a fiction as output and produces fictional output. Confabulation serves as a stern warning to those who would believe everything they are told based on introspective reporting. Still, as a standalone research result confabulation no more invalidates introspection than an optical illusion disproves vision. Confabulation does, however, contain a lesson learned and a caution about introspection’s limitations. The controversy is about where to draw the boundary about the (un)reliability of first person, introspective reports.

In Sperry’s split brain experiments, the corpus callosum of the brain of the person suffering epileptic seizures was surgically severed to stop the debilitating epileptic

attacks.<sup>10</sup> The corpus callosum functions as a switching center, transferring data between the two main hemispheres (right and left) of the brain, coordinating the verbal (left) hemisphere with the spatially-oriented, “intuitive” right one. It is also important to know that information captured by the sense organs on the right side of the body is mapped onto the left hemisphere and vica versa. (You didn’t think this was going to be simple, did you?)

Now the scene changes to Sperry’s psychology-physiology laboratory. When the right brain was informed (via the left eye) of a word on a flash card – e.g., pencil – then the split-brain subject was able to pick it out from a collection of things ready to hand but otherwise hidden from his view by feeling for it behind a partition that blocked vision. The subject was able to pick out the item by touch alone based on the word he was shown on the flash card; but he was unable to say what word he was shown on the flash card since the data had been flashed to the left eye, which, in turn, maps to the nonverbal right brain hemisphere. The inference is that the input does not reach the left hemisphere that controls verbal production from the right hemisphere where received the initial stimulus word, because the connection between the two hemispheres had been surgically severed. The subject cannot say the word – it is not introspectively available to his awareness. But he can pick out a pencil from among various objects when asked to do so. So the information was available to the system – presumably in the right, nonverbal part of the brain.

Now the experiment advanced to the story telling (confabulation) part. When further questioned about what they knew and why and how they were able to get the right answer, such subjects “confabulated” – they invented a story, a complete fiction. They had no idea how or why they picked out (e.g.) the pencil (assuming that indeed was the word match object). If such subjects claimed the story was based on introspective awareness, then introspection was as much a story as the tall tale told about why the pencil was chosen. Note, however, that doubt about introspection played little or no role

---

<sup>10</sup> M. Gazzaniga. (1967). “The split brain in man” in *Readings from Scientific American: Perception: Mechanisms and Models*, intr. R. Held and Whitman Richards. San Francisco: W. H. Freeman, 1972: 29f.

in the results reported by Sperry and, following Sperry, Gazzaniga. The latter focused on the ability of the brain to support multiple centers of consciousness only one of which was able to give a report using language. However, the work did imply a simple scientific definition of introspection similar to Wundt's – expose the subject to a visual stimulus and then ask him for a report. In Sperry and Gazzaniga's case, provide the subject with a stimulus – the word “pencil” – and then ask them to identify the object in two different ways – verbally and tactilely (by touch). He could only do so tactilely, not verbally. As things stand, the phenomenon cast in doubt the accuracy of introspective (verbal) reports no matter how sincere. It is easy to generalize to all forms of introspection and the behaviorists have not hesitated to do so. What is warranted, however, is skepticism about the accuracy of verbal reports based on introspection of data not available to the individual's attentional awareness. As a cautionary tale about the accuracy of verbal reports based on introspective awareness the results stand, though limited to individuals whose cerebral commissure has been severed. Such a restriction does not apply to the next case.

A similar result casting doubt on the value of introspective reports was provided in a widely cited article by Richard Nisbett and Timothy Wilson.<sup>11</sup> The researchers performed interviews in a shopping mall, displaying panty hose, and asking respondents which pair were preferred and why. The pantyhose were identical so whatever reasons given for preferring Brand A to Brand X were necessarily invented – confabulated. In fact, according to the researchers, preference was caused by the position of the display of the items with those in the right position being preferred. Such location-oriented preferences are common, and I am inclined to admit their validity. People do not have privileged information about the *causes* of their preferences and closely related behaviors.

As noted by the researchers in subsequent debates about the meaning of this experiment, what people *do* have available is a wealth of content about their autobiographical memory and history – the tree on which they inscribed their initials in high school when no one else was looking and other contingently secret details. No one else has access to

---

<sup>11</sup> R. Nisbett and T. Wilson. (1977). Telling more than we can know. *Psychological Review* 84: 321-259.



such information. But this does not necessarily help provide a useful analysis when trying to understand an individual's own motives, character, preferences, and patterns from the past that influence a decision or given segment of behavior.

This is surely a cautionary tale for anyone who wants to make use of introspection, but it does not result in the wholesale invalidation of introspection as subsequently acknowledged by the researchers. The respondents were commenting on their reasons for making a choice because that is what the authorities running the experiment requested. But the choice was influenced by physical position as well as personal preference. Had the experimenters asked what caused them to prefer the pantyhose located at [position] D on the right rather than A on the left side, then the respondent might have introspected – that is, reported - that she was sensitive to location. Of course, the experimenter's question would have been different, and, thus, so would the experiment have been different.

Under any interpretation, it is a brilliant experiment from which no amount of 20-20 hindsight can undo the fundamental value – be skeptical about what people tell you – believe half of what you see and none of what you (merely) hear. Still, it does suggest that additional conditions and controls on introspective data gathering will be useful. The distinction between reasons and causes may usefully be invoked. The same phenomenon can be described and redescribed to highlight the cause for a decision and a reason for a decision. In this experiment, the cause was not available to attentional awareness and so was not identified as input to the reason-giving process. Instead the subject noted a quality of the panty hose itself, such as its fineness or strength. The conclusion? People are good at giving the right answer for the wrong reason, where “reason” can include causal triggers of which the person is transiently unaware. Second, the authority of the questioner in his scientific, white lab coat (if he was indeed wearing such) can become a consideration. The initial reaction might well be simple ignorance – I have no idea why I picked Brand X. It just seemed nicer. That's not a reason. That's an admission of ignorance. Such a reply should not be coded as a reason when none was forthcoming. Third, empathy can provide one such aspect of control. Going forward, the researcher's

empathy for the experimental subjects can be expected to disclose such phenomena as pressure to give the right answer, fear of giving the wrong answer, wanting to “look good” and wanting not to “look bad” On the part of the experimental subject. This is likely to be so regardless of what the researchers tell the candidate respondents up front, which is consistent with such communications in any case.

In addition, the researchers, whether they realized it or not, are extending work previously conducted with children to adult respondents. Jean Piaget emphasized the ability of the child (ages 8 to 12) to provide a justification for many phenomena for which the child has limited understanding.<sup>12</sup> Piaget (1930: 156) attributes the “need for justification at any price” to the child’s pseudo-logical and pre-causal grasp of reasons and causes. For example, when asked how two sayings of folk wisdom relate to one another - “White dust will never come out of a sack of coal and “We must work to live” - the child will suggest a fantastical answer - that “A person needs money to buy coal.” The ability to merge two sayings that are otherwise unrelated according to an adult way of thinking is attributed by Piaget to syncretism, a primitive tendency of pre-logical thought to merge related contents by means of superficial associations. According to Piaget, an adult would be more candid in acknowledging the two sayings are unrelated. It is not part of Piaget’s analysis that the adult respondent frequently behaves similarly to the child in providing a clever but anomalous response, so to speak as the adult’s “inner child” takes over and give free reign to its imaginative association. The evidence is that, when put in a position where an answer is required, the syncretic tendencies that Piaget claims are left behind by logical and adult modes of thinking reemerge, at least transiently, in answers given in shopping malls.

My suggestion is that such confabulations and confabulation-like phenomena are the equivalent of optical illusions in the realm of introspection. Absent specific steps, tactics, and procedures to control, eliminate, or reduce such distracters, the use of introspection will be what it already is today – of limited value and requiring rehabilitation to deliver the potential value it has. In short, the existence of instances where introspection goes

---

<sup>12</sup> Jean Piaget. (1930). *The Language and Thought of the Child*, tr. M. Gabain. New York: Meridan, 1974.

astray – misfires – and does so systematically and predictably is no more justification for throwing out introspection than the skeptic’s argument that once our senses deceive us they cannot be trusted at all, ever again.

A long list of control and constraints must be marshaled in order to improve the accuracy and reliability of our reports of perceptions of the physical environment. Why should it be otherwise with introspection? Confabulation belongs with such phenomenon as rationalization (as opposed to reason giving) where distorted logic (including informal fallacies) is used to make us feel justified or look good. When asked by someone in authority, such as a scientist in a white lab coat with a clip board for marking down your score, a person feels compelled to provide the “right” answer whether due to habit, friendliness, or intimidation. Years spent in school giving the “right answer” are a pattern that is hard to break. A person tends to give a reason for what he wants or intends to do anyway such that he is cast in a favorable (or at least neutral) light. In general, people are not liars, but they tend to be a tad casual about telling the whole truth and nothing but the truth, especially in a situation where deep personal disclosure is requested. The list of things that lead our introspection astray is long – confabulation, the tendency to wishful thinking (“wish fulfillment”), wanting to look good in front of one’s peers or in one’s community, wanting to avoid looking bad, rationalizations such as wanting to be right and justified, and actual hallucination of desired outcomes and conclusions based on the sketchiest of evidence (“she really does love me – not!”). This list is not complete. All of these things can be controlled only imperfectly and by means of disciplined attention to details of experience and the context.

Without wishing to be flippant, if one’s corpus callosum has been severed, then one must be cautious about reports based on data delivered to only one brain hemisphere. This is usually not a problem outside neurological laboratories where such isolated inputs can be manipulated, but one can easily imagine situations of using peripheral vision in piloting an airplane where communications with the copilot might be impaired. Obviously a physical check up with a doctor is appropriate up front. At a less extreme level, people tend to placate and sincerely agree with those they respect or fear. So those who wear

white lab coats such as doctors, researchers, and scientists, have a significant authority over others, which in itself cause the other to try to guess what is wanted and do it. The power of suggestion is useful in educating and informing; but it is much less useful in surfacing factual details (“the truth”) about painful situations that may be unclear in the present or from the past.

Thus, it is no more appropriate to throw away the evidence of introspection because we lack training in it and because our introspective reports sometimes deceive us than to throw away the evidence of everyday perception because our senses sometimes deceive us.

### **Introspection as Displaced Perception**

The modern author who engages introspection at the level of elementary sensations familiar to Wundt and Titchner, without, however, being committed to their research program is Fred Dretske.

Dretske (1981, 1995) is not interested in sense data, he is interested in information.<sup>13</sup> He conducts a raid on the inarticulate, pressing back the boundary of experience to get access to a patch of blue, not as an isolated, reidentifiable sense data, but as input to an information supply chain that results in multiple forms of cognition, especially belief. From the perspective of this work on empathy, the issue is what happens when this information processing apparatus is directed, not at a patch of blue but, at the emotional life of an animate organism expressing itself. To appreciate the value of such a shift (which Dretske does not necessarily conceive or endorse), we shall have to engage the details of his key distinctions around displaced perception.

Dretske collects reminders that identifying the content of introspection requires looking outward rather than inward.<sup>14</sup> Accepting that knowledge, including belief, is always

---

<sup>13</sup> Fred Dretske. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: CSLI Publications, 1999.

<sup>14</sup> Fred Dretske. (1995). *Naturalizing the Mind*. Cambridge, MA: The MIT Press, 1999.

conceptual, Dretske redescribes introspecting a patch of blue as a metarepresentation of blue – a conceptual representation of blue – that in effect displaces the perception of blue in the direction of the information the perceiving system (“the self”) has about the experience of blue. This kind of *epistemic seeing* is not as obscure as it might first sound. When a person interrogates his or her experience of the blue of the sky, there is not some additional experience but attention to the blue of the object (the sky) experience. Mental events are constituted by the relation these events bear to external things and situations. One of Dretske’s paradigm examples of a representational system is a scale that shows a person how much he weighs by looking at the pointer to the weight on the dial, not at his bulging midsection. The representation is oriented towards the external state of affairs, the weight, not towards the individual’s bulging midsection (e.g., 1995: 41). The scale has all the information it needs to represent both the world of weight as well as introspection about it. Self-knowledge in this elementary sense of knowing what an individual is sensing becomes an example of displaced perception.

The knowledge a system allegedly gets from introspection is based on a further (cognitive) processing of experiences. Of course, what the scale that shows the weight and the gauge that shows the pressure lack is any mechanism or method of articulating the information by means of a concept or metarepresentation. The scale gauge does not know the weight of the person; but it could know it, if it had the right concept(s) and connecting beliefs.

Under my interpretation of mineness, Dretske’s insight about displaced perception turns on which properties get coded as “mine” and which get coded as not mine and so as external. So called secondary properties such as smells, aspects of touch, sweet and sour and saltiness as they impact the palate, generate representations in the interaction between individual and milieu that are experienced as nonmental and as external states. When an individual directs his attention to such “external” states this attention and its direction is described as external attention, phenomenologically as well on the level of folk psychology (e.g., see Metzinger 2003: 34). The point is that the externally directed attending to the smell of the rose as well as its redness are both a form of introspection as

displaced perception. The smell of the rose gets encoded as an inner experience of an external property of the rose. The difference is that the primary properties (color, shape, substantiality) are introspected as displaced from the outside inwards and misinterpreted as sense data (by the sense datum theorists) while the secondary properties (smell, taste, warmth) are displaced from inside outwards and misinterpreted as public properties of objects whereas they are just a function of interaction with the sense organs of the perceiving person.<sup>15</sup>

An account (redescriptions) of introspection corresponds to each of these two forms of externalism. The displaced perception of the experience of a secondary property is a redescription of the experience of sweetness from being mine to being not-mine, and as such becomes rather a quality of the sugar. In effect, the individual attributes these qualities into the world as if they were objective correlates of phenomenal properties, the sweetness being in the sugar, the saltiness in the salt, and the fragrance in the rose.<sup>16</sup>

While Heidegger and Dretske make strange bedfellows, Dretske provides an account that of what neither Heidegger nor anyone has been able to explain completely, why a mood (*Stimmung*) that is part of a milieu at a party gets experienced as mine (Heidegger 1927: H136).<sup>17</sup> Of course, Heidegger defines a mood as the way in which the world is disclosed to human existence (*Dasein*) even though existence turns away from what is disclosed by coding it as internal. But the mystery remains – the individual is in a good mood or a bad mood, but that is encoded as inner, as mine, and has nothing to do with the environment. The answer lies in a way of being that is in parallel with Dretske’s displaced perception, what can be described as “displaced affectivity.”

---

<sup>15</sup> By an argument that Kant makes in his *Prolegomena*, the redness of the rose is as much an appearance as its smell. Both are displaced perceptions, but the one misleads from the inside out; the other from the outside in.

<sup>16</sup> This is the mechanism that T. Lipps focused on in his theory of projective empathy. It is how we animate nature with qualities that are a function of our interaction with nature. It is also the basis of emotional contagion.

<sup>17</sup> M. Heidegger. (1927). *Being and Time*, trs. J. Macquarrie & E. Robinson. New York: Harper and Row, 1962. The pagination with “H” refers to the standard German edition used in the margin of the translation. A detailed drill down on Heidegger’s approach is engaged in Chapter \_\_\_\_ on Heidegger’s Special Hermeneutic of Empathy.

Paraphrasing Dretske's analysis of epistemic seeing in a Heideggerian context, if you want to find out about your mood, do not look inward, look outward at the situation you are in, look at the "there," look at the context. The mood is a displaced sensory-affective complex about how and why the situation matters to you. The gaiety or fear that sweeps through the individuals at a party and in a crowd is a sensory-affective complex that would be individually available for direction and control of behavior and may even be available for transient attentional awareness (introspection); but the panic or gaiety is not a source of cognition as it functions in causing the mass behavior of the crowd. That is why the behavior is "mass behavior," "crowd behavior," rather than the coordinated team of a group of individuals trying to follow a plan. The individual in the crowd who is gripped by it does not have a functioning concept of a process of emotional contagion.

In a mood the individual experiences happiness, gaiety, panic or rage as being internal; but they are functionally caused by being part of a crowd or group in which the communicability of affect is transmitted by means of a process shared at the level of the organism, for example, hypothesized to be mirror neurons (the underlying mechanism is not critical path for the argument at this point). Of course, emotional contagion is *not* empathy, but rather relies on a mechanism that also provides an input to the empathic process – in which affects are subsymbolically transmitted through a crowd or from one individual to another without awareness that it is not an individual emotion but a communicability of affect.

The process of displaced perception, epistemic seeing, recruits a concept to enable the primary property to be reidentified. In parallel, a process of misplaced affectivity the world and things in it come to matter to the individual. This corresponds to a vicarious experience, which likewise is not a complete empathic experience, but at least has reference to the other as the candidate cause of the vicariousness of the experience. A vicarious feeling is an affect caused by another that gets encoded as internal, as mine.

A system gets information about itself by perceiving, not itself but, what it is already perceiving externally plus a conceptualization about it (such as belief) (Dretske 1995:

53). This becomes the source of the limited, deflated first-person authority that we do have, especially about the content of our external experiences as they are reflected by displaced perception. This is also why introspection, according to Dretske, has no phenomenology of its own. The phenomenology is derivative on perception.

### **Ryle on Introspection**

Thus, displaced perception comes close to being what Gilbert Ryle described as “heeding”:

But when we speak of a person minding what he is saying, or what he is whistling, we are not saying that he is doing two things at once. He could not stop his reading while continuing to pay attention to it, or hand over the controls of his car while continuing to exercise care; though he could, of course, continue to read but cease to attend, or continue to drive but cease to take care. . . . We commonly speak of reading attentively, driving carefully and conning studiously, and this usage has the merit of suggesting that what is being described is one operation with a special character and not two operations executed in different “places,” with a peculiar cable between them.<sup>18</sup>

That is, there is no second mental place in which the mind’s eye of attention observes while I read or drive. The result of heeding what I am doing is akin to introspective attending to it, but has the advantage of not positing a special place outside of space and time for this activity.

Thus emerges the classical theories of inner perception (Descartes, Brentano) that were used by Gilbert Ryle in his reduction to absurdity of the ghost in the machine, the inner theatre populated by an executive function that operates as the homunculus, the latter caught in an infinite regress, trying to catch itself, according to Ryle, in the act of thinking about its thinking, monitoring its experience without monitoring the monitoring.

Besides being currently supplied with these alleged immediate data of consciousness, a person is also generally supposed to be able to exercise from

---

<sup>18</sup> Gilbert Ryle, (1949), *The Concept of Mind*. New York: Barnes and Noble, 1965: 138. Although Ryle is a student of ordinary speech, he does not hesitate to coin an idiom when it suits his purpose. Thus, “concepts” are like populations which display “behavior” that is naturally of a logical kind. His examples are meant to map the “logical behavior of heed concepts.”



time to time a special kind of perception, namely, inner perception, or introspection. He can take a (non-optical) “look” at what is passing in his mind (Ryle 1949: 14).

According to Ryle, the notion of inner perception is based on the category mistake of transferring observation from the sphere of body to mind without first asking whether the mind-body distinction is intelligible in any case. The idea of inner perception or inner inspection is just a metaphor, but a misleading one when made the basis of a non-metaphorical theory of mental functioning. Ryle’s account is filled with many negative results, all consequences of his *reductio ad absurdum* of the attempt to undo the body-mind dichotomy through inner perception.

However, rather than perform a critique of the scope and limits of introspective data gathering from a commitment to fix the flux, Ryle throws out the introspective baby with the ghost of consciousness bathwater. He commits to the translation of introspective experiences into behaviorist terminology by doing performing the translation.

“Attending” becomes “heeding,” which adverbially modify all manner of behavioral dispositions, especially mental ones that manipulate symbols such as calculating, reading, (symbolic) thinking. “Free association” becomes “unstudied talk,” in which people can catch themselves in the act of transforming a dispositional emotion into an occurrent one. “Unconscious emotions” are handled as “dispositions.” The inner stream of consciousness becomes the adverbial modification of dispositional behavioral activities and intentional attitudes. The disappearance of introspection is near.

Unfortunately, from the perspective of logical behaviorism, Ryle is forced back in the direction of the flux by his commitment to the phenomena themselves. Ryle is a good phenomenologist. He cannot overlook the evidence of ordinary language that points to the irreducibility of mental events such as moods or frames of mind to nonoccurrent dispositions.

. . . To say that someone has done something, paying some heed to what he was doing, is not only to say, e.g., that he was ready for any of a variety of associated tasks and tests which might have cropped up but perhaps did not; it is also to say

that he was ready for the task with which he actually coped. He was in the mood or frame of mind to do, if required, lots of things which may not have been actually required; and he was, *ipso facto*, in the mood or frame of mind to do at least this one thing which was actually required (Ryle 1949: 141).

Amazingly enough, I think this represents a passage in which Ryle is struggling for words. Heed concepts are “mongrel categoricals” (1949: 141). They are hypothetical in that they serve as inference tickets for a number of possible predictive outcomes (“inference ticket” being Ryle’s key phrase for a “warranted inference”). Yet, they also describe current matters of empirical fact such as an individual *concentrating* on the task at hand that are not visible in overt behavior. Ryle paraphrases this rather infelicitous locution, “mongrel categorical” by “mood or frame of mind.” He seems to be trying to say that “paying some heed . . .” is not only a conditional prediction of further occurrences, but is also a report of an actual occurrence. It is:

. . . To accept the dispositional account would apparently involve us in saying that though a person may properly be described as whistling now, he cannot be properly described as concentrating now; and we know quite well that such descriptions are legitimate (1949: 139).

In short, the concept of heed cannot be completely unpacked through the distinction between dispositions and occurrent behavior. Being “ready for the task” is being in a frame of mind known as “readiness.” Ryle is struggling mightily not to introduce intentional attitudes in order to save the phenomena. “Concentration” and a whole host of mental attitudes and contents reemerge here.

A “mood or frame of mind” is completely parallel with Dretske’s displaced perception, though the displacement is in the reverse direction. In displaced perception an aspect or property of an external object – the redness of the apple – is represented as an internal awareness of the (externally directed) experience of the apple. The internal and external are ways of coding the experience; and, in the case of introspecting, the property of red, externalism rules the process. In contrast, in the case of a mood, the experience of properties of an external situation – people laughing and having fun at a party or the anxious and urgent actions of the staff in a hospital emergency room – are represented as

internal awareness of the externally directed experience of the situation. It works out nicely in English, the mood is not in the individual; the individual is in the mood. But the individual experiences the mood as being in him or her – that is precisely what it means to say the mood is “coded” as being internal. The individual finds out what is the external situation by exploring the mineness of his or her own mood. The analysis is completely parallel in the instance of a particular emotion such as hope or fear that is communicated from another individual in a process of affective resonance with another individual, only the language is not quite so accommodating. In no case is there a ghostly theatre of the mind. Yet such a theatre is a compelling illusion that returns no matter how often one disabuses oneself of it.

### **Dretske’s Return to First-person Access and Authority**

Dretske’s account of displaced perception is designed to “disappear” introspection as not fundamental but even if introspection has no phenomenology of its own, it does have a phenomenology that belongs to someone else - perception. The burden of the position that introspection has no phenomenology is the requirement to “explain away” all the introspective dynamics that are reported. Dretske is clear that he does not want to maintain what is useful and sensible about first-person authority.

Unlike Ryle, whose logical behaviorism is committed to dispensing with consciousness not only as an inner theatre but as a target for cognition, Dretske is still committed to an “externalism of the right sort,” in which first-person authority is not jeopardized (Dretske 1995: 54). Externalism of the mind allows for privileged access, albeit of a deflationary sort, because introspection is a process in which information about internal matters is gained in the act of representing external objects (Dretske 1995: 54). The entire rich external world of sensations and content is captured and reflected in the mental representations of the mind.

Although Dretske does not invoke “mineness,” this would have been useful to him. The phenomenology of introspection is the qualitative mineness that accompanies the experience of blue of the sky. Introspectively, the blue is mine – my experience of blue -

not the sky's. But one learns about the experience by looking outward at the sky that is not mine, not inward at the experience that is mine. Likewise, this is the case with moods. The phenomenology of my good mood is the qualitative mineness that accompanies the experience of the party. Introspectively, the gaiety is mine – my experience of gaiety – not the party's. But one learns about the gaiety by considering events at the party, not by looking inward at the experience of gaiety itself.

However, the issue is not the content, which we do see, but the attitude that we do not see. Dretske:

The problem is in knowing that you believe and experience it. The problem centers not on the content, but on the attitude (the relation) one has to that content [ . . . ] We are all very good – in fact, absolute authorities – about what we think and experience, but not very good (in fact, I think, very bad) about the attitudinal aspect of these mental states (Dretske 1995: 55).

In other words, people are good at identifying the content of their beliefs, but if people are asked to distinguish how they know the difference between belief, conjecture, imagining, expecting, recollecting, hoping, fearing, empathizing, or other intentional attitudes, then they are hard-pressed to say what is the difference. If they have mastered the meaning of the words – “belief,” “conjecture,” etc. – then they just know. The relation of the intentional attitude to the object, the content “it is raining” may be one of belief, in which case I reach for my umbrella, or hope or conjecture or doubt or imaging, in which cases I cross my fingers, raise my eye brows, or other relevant gesture. The point is not that some behavior is necessarily connected with the intentional attitude, but rather that we can demonstrate that we live in an understanding of such intentions and know how to go on and explain ourselves in all kinds of contingent circumstances.

The language of the vehicle-content distinction has its limitations, but is pervasive. In introspecting, an individual is readily aware of the content but less so when it comes to the vehicle. It is by no means used in exactly the same way (but there is enough overlap to ground an alignment) by Metzinger, who explicitly makes the distinction; Dretske, who in the above-cited passage distinguishes attitude (the relation) and content; Husserl,

who distinguishes intention and quality; or Searle, who distinguishes attitude and mode as aspects of the intention.<sup>19</sup> However, in every case, the attitude/quality/vehicle is transparent in standard, normal, non pathological circumstances, to introspection. Introspection does not see it. It is invisible, though, arguably, functioning in perception and related intentional acts. It is as if pouring the content into the container makes the container visible, like pouring orange juice into a glass pitcher – the pitcher was transparent prior to containing the colored juice. That is, the content is opaque – the person can identify in it introspection (displaced perception under Dretske’s interpretation); but the attitude is transparent, i.e., invisible, one just sees right through it.

The reason one just sees right through it is that there is no additional perception *of* the intentional attitude as such to displace. Under this description, the intentional attitude constitutes the subject’s contribution to experience, independent of the content of sensation. And Dretske has no way of accounting for any other kind of introspection than epistemic seeing. We can think about the relationship between belief and the object of belief, but the relationship is not introspectable, at least not as content like a color or sensation or other property. Of course not, it is a relationship, not a property. You cannot get from the content (matter) to the attitude, when the content is abstracted as a standalone given, because the same content (“it is raining”) can be the target of many diverse attitudes as indicated – I believe it is raining, I hope it is raining, I fear it is raining, I remember that it is [was] raining, etc. Relying on what he says in an earlier work (1981: 177), Dretske has a valuable clue: the intentional attitude is the manner in which the content is *encoded* by the system, not the content as such. But this remains undeveloped without introducing the notion of mineness as suggested here.

---

<sup>19</sup> E. Husserl. (1900/21). “Investigation V: On intentional experiences and their contents” in *Logical Investigations*, tr. J.N. Findlay. Vol. II. London, Routledge, 1970: 119 (§20); Thomas Metzinger. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press, 2003: 4; John Searle. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge, UK: Cambridge University Press, 1983: 7.

## **Introspection as Simulation**

In trying to describe something as simple as searching for just the right word, a linguistic tidal wave of simulation is brought to the effort to translate what is basically a momentary, microscopic event (or small set of related events). Metaphors relating to information technology, computing, and simulation by computing of would-be introspective events loom large in the description:

Introspectively, one draws pictures. Thus one describes oneself as “looking for” a word. Whenever it is baffled or momentarily vacuous, the search, the act of scanning, suggest circuitry. The relevant sensation or, more cogently, the vulgarized images we make up of what are subliminal processes, leave one with a compelling notion of nervous probes “trying this or that connection,” recoiling where the wire is blocked or broken and seeking alternative channels until the right contact is made. The sensation of a “near-miss” can be tactile. The sought word or phrase is a “micromillimeter away from” the scanner; it is poised obstinately at the edge of retrieval. One’s focus becomes excited and insistent. It seems to press against a material impediment. The “muscles” of attention ache. Then comes the breach in the dam, the looked-for word or phrase flashing into consciousness. We know nothing of the relevant kinetics, but the implication of a correct location, of a “slotting into place” is forceful, if only because of the muted but unmistakable impression of release, of a calming click which accompanies the instant of recall. When the right word is found, compression gives, and a deep-breathing currency – in the dual sense of “flow” and “integrated routine” – resumes. In contrast, under the spurs of stimulants or histrionic occasion, or in the strange weightless tension of tiredness of mind, resistance seems to diminish the verbal circuits and synapses multiply. Every bell chimes. Homonyms, paronomasia, acoustic and semantic cognates, synecdochic sets, analogies, associative strings proliferate, undulating at extreme speed, sometimes with incongruous but pointed logic, across the surfaces of consciousness. The acrostic or cross-word yields faster than our pencil can follow. We seem to know even more than we had forgotten, as if central sediments of memory or reserves normally unrecorded, because lightly imprinted or laid down without deliberate marking, had been galvanized. At yet another level of banal experience there are short-circuits and wires fuse. The identical morpheme, tonal combination, or atrophied phrase forces itself on the inner ear, insistently, like a bulb going on and off pointlessly. Some part of the memory circuit is trapped (Steiner 1975: 289-90).

Introspection is accessible through language and is drenched with language; but it is not limited to language, an inner dialogue, verbal thinking, or models based on computing simulations. Images of computing simulations at the hardware level are common in

accessing introspection; and Goldman's simulation theory (ST) of understanding other individuals welcomes the rehabilitation of introspection.

At this point it is clear that additional perspectives on the issue of introspection will be useful. Goldman (*Simulating Minds* 2006) proposes to treat introspection as a perception like process in which attention is the organ (method) of introspective orientation like the eye for sight. The choice of Goldman is guided by the rich distinctions as well as engaging limitations of his approach. Some of the simulations (models and analogies) will require "as if" processing to grasp how the other feels being me with or without my character traits, preference functions, and historical patterns of experience. Unfortunately, Goldman gets it almost exactly backwards in his account of introspection as a process with "input – process (transduction) – output" where the output is the attitude or intentional quality (e.g., believing, desiring, imagining, being angry, recollecting, seeing) being identified and classified. He asserts that function, phenomenal, and representational properties are *not* suited as inputs to the classification process, but that neural properties are. Notwithstanding his claim, the following shall trace how all these are suited as inputs, but with the exception of phenomenal properties, none are introspectable as such in any sense of introspection, displaced or otherwise, discussed in his chapter or this one.

Goldman proposes to use introspection as a perception like process to classify (recognize) intentional (propositional) attitudes (beliefs, desires, remembering, etc.) based on the input of various candidate properties. He makes a nice point that attention is the organ of introspection that is analogous to seeing in the case of visual perception. The relevant sense of introspection that he deploys is "the process of analyzing or classifying selected [mental] states" as being one or another propositional attitude (Goldman 2006: 246). The candidates properties include functional, phenomenal, representational, or neural properties (Goldman 2006: 247).

### **Three Problems with Goldman's Approach.**

First, with the exception of phenomenal properties, none of these groups of properties – functional, representational, neuronal – are available for attentional recognition,

classification, or monitoring (in any sense). In general, these are not introspectable properties; and, on the contrary, are subject to autoepistemic closure – that is, as introspecting subjects we do not have access to the genesis, operation, or history of the functional, representational, or neuronal process. They are levels of analysis to which an individual does not have introspective access. Their generation and history is not available. If they are inputs to the process of introspection, which might well be the case, then we are unaware of it as a matter of empirical fact as well as by definition.

It is true there are some border line redescriptions that suggest the intentionality of representation does have its own phenomenology. The content of an intention is opaque to attentional inspection, which provides an indirect access to the vehicle, the intentional attitude, provided one can tease out the distinction from the phenomena. Likewise with neurology. Goldman argues quite cogently that there is an internal milieu – not exactly neurology – but perhaps close enough and involving the organism's chemical-endocrinological equilibrium that percolates up into consciousness through the ventromedial nucleus of the dedicated thalamocortical relay (Goldman 2006: 252).

“Interoception is the perception of the physiological status of bodily tissue [ . . . ] Introspection, as I propose to understand it, is a further response to these sensations . . .”

Goldman 2006: 252). Well and good. By the time, these sensations percolate up out of the pulsating cauldron of biological infrastructure, they are phenomenal properties.

Neurons and their related process have properties such as sodium ions gradients that pump charged electrons across the neuronal cell membrane interface, recovery times, and other chemical and biological transformations. A similar logic applies to any of the other properties. If an individual is introspectively aware of their operation, the functional and representational properties have become phenomena available to attentional monitoring.

It is a further question whether this ever actually occurs, but if it does, then these properties are phenomenal by definition. Indeed each of these has its own phenomenology in the elementary sense that they sometimes percolate up and pass over the threshold of conscious awareness and are then available for introspective processing. This may be on an exception basis as in the functionality of attention deficit disorder in



which the attentional function becomes conspicuous in being rapidly attracted by one thing and then another in sequence. What most individuals experience as insignificant background events become attentionally salient for the individual with attention deficit disorder.

One thing that tends to make introspection uninteresting and causes it to fall into disrepute is that many of the introspected feelings and sensations of the so called internal chemical milieu are relatively insignificant. Ordinarily there is a vague, diffuse sense of well-being and being okay. If something goes wrong and a person becomes sick, then these contents become important in providing clues as to how well one is doing – symptoms start showing up - dizziness, feelings of fever, chills, etc. But even then, as Dretske point out, I look at the reading on the thermometer to confirm the way an individual, including myself, actually feels, yes, I have a fever.

Second, strictly speaking if the analogy is between the eye as the organ for seeing and attention as the organ for examining the stream of presentational conscious as a perception-like inner sense, then the functional, representational, phenomenal, and neurological properties should be input to attention, not introspection. This may be a slight imprecision on the part of Goldman, but it is of the essence. Remember in this case the individual is undertaking to classify the kind of intention (“propositional attitude”) that is being deployed in “believing that” “imaging that,” “desiring that,” “remembering that,” and so on through a relatively long, but finite, list of mental attitudes. Introspection is being deployed to provide an answer to the question as to what is the so called quality, modality, form, of the intentional attitude. Is it belief or desire, etc? Engaging the phenomenal, representation, functional, and neurological inputs to the stream of introspection is in effect to ask about the infrastructure and presuppositions of the totality of consciousness. This is too ambitious an undertaking. Therefore, Goldman is likely using “introspection” in the narrow sense of the term as a substitute for “attentional availability to answer the issue at hand introspectively.”

The third issue is Goldman's confusion about whether unconscious states can be introspected – that is, made the target of attentional activity. “So it would be unwise to push the thesis that the scope of introspection is restricted to conscious states” (Goldman 2006: 246). This is not just a standalone howler. It works further damage as Goldman draws out the inference. This causes him to disqualify phenomenal properties as the canonical input to an account of introspection (attention). As he tries to identify the introspectively discriminated properties associated with all mental-state types, he arguably jumps to further unwarranted conclusions: “If it is granted that nonconscious mental states can be introspected (or monitored), that is a problem for the phenomenal-properties answer, because nonconscious states presumably lack phenomenal properties” (Goldman 2006: 249). Granted that nonconsciousness states lack phenomenal properties (which are, by definition, conscious), still it is *not* a problem. Nonconscious states are simply not introspectable. Unless a person is sleep-walking, the person is conscious in an elementary and entry level way without attending to anything in particular as the individual goes about his business on “automatic pilot,” getting things done. As we shall see, such states are not the focus of attention, but they are *not* therefore nonconscious; they are immediately available for attentional awareness if the individual chooses to attend to them or is forced to do so (say) as when a beeper goes off, demanding attention, and indicating the need to call the office.

### **The Disentangling and Clarification of Attention**

The disentangling and clarification what are the inputs to attention and the role of attention in introspection requires answering the question, “What then is ‘attention?’” Goldman gathers some interesting reminders, but never quite gets around to a definition. He notes (“calls attention to”) the role of attention in descriptive experience sampling (DES). A person is given a beeper and agrees to write down in a notebook the answer to one question (phrased in various ways) when the beeper goes off at any arbitrary time: “What was occurring in your inner experience at the moment of the beep?” For example, one individual discovered that he had frequent angry thoughts about his children. The

individual had been unaware of those thoughts prior to this experiment in descriptive experience sampling (DES).<sup>20</sup>

The point is that people who are awake – conscious and aware, not sleeping – go about their business on automatic pilot in a state of what might be called “entry level” consciousness. Along the same lines, Dretske quotes David Armstrong as referring to “minimal consciousness” (Dretske 1995: 104-5). Freud refers to the “preconscious” to distinguish such contents that are contingently out of awareness at a given moment (such as your phone number) from dynamically unconscious contents that cannot be introspected even if the individual were to try. However, as regards minimal consciousness, individuals are ready at a moment’s notice to attend to whatever salient features of the environment are useful for their well being – such as a turn in the road confronting a truck driver behind the wheel. When the beeper goes off, that’s attention. When the traffic sign indicates “Bend in the road,” what gets activated is attention (or the truck ends up in a ditch). However, other than the analogy with the eye of inner sense, no definition of “attention” is forthcoming from Goldman. Here is one that applies the distinctions invoked by Goldman in turn.

Neurologically, attention is the allocation of additional neurological and biological resources (e.g., oxygen, adrenaline) to the details of a situation, object, or process that is on-going in the window of presence. Informationally, attention is the allocation of additional computational resources such as caching the result of afferent (inbound) stimuli in short term memory, drilling down on the details of the content to bring out salient features by comparing or contrasting them with related features, finding distinctions where there were previously not thought to be any. Functionally, attention is focusing on the granularity (whether fine or coarse) of the relationship between the input and output and the operation of transforming the one into the other. Representationally, attention is the recognition of the relationship between the intention and that of which the

---

<sup>20</sup> R.T. Hurlburt and C. L. Heavey. (2001). “Telling what we know: describing inner experience,” *Trends in Cognitive Sciences* 5: 400-403 cited in A. I. Goldman, *Simulating Minds*. Oxford: Oxford University Press, 2006: 242. [Editorial note: This experiment is also discussed from a different perspective in Chapter \_\_\_\_ on Empathy and Unexpressed Emotions.]

intention is a representation of – whether the identification requires ascent from a content-laden experience that is retrospectively analyzed or some other recognition of a form of intentionality. Phenomenally, attention is paying heed to what one is doing, what one is experiencing, what one is relating to.

Therefore, we are now in a position to distinguish properties associated with all mental-state types (“intentional (propositional) attitudes”) and their role in introspection. Remember we (and Goldman) are looking for a transduction function, treating attention like the function of the eye in a perception-like process of introspection, in which the input has certain properties =  $x$  and the output is the classification, “Yes, this is a belief it is raining, not a desire it is raining, nor a fantasy it is raining, etc. Contrary to Goldman, an entire hierarchy of inputs to attentional introspection will have to be distinguished. Only the phenomenal will be available to consciousness and attentionally available for introspection.

**Functional:** Functional properties are relational and rule-governed, mapping inputs to outputs. The focus of attention is often demanded by compelling events in the environment – watch out for that bus! However, that is not exclusively the case. The spontaneous focusing of attention is a choice available to individuals as they decide to apply themselves to a vast array of arbitrary activities, intellectual or physical. The functional analysis of such intentional attitudes as belief, desire, fantasy, empathy, etc. looks at operations such as the spontaneous synthesis, identification, reproduction, recognition, projection, and whatever other causal mechanisms play a role in providing a coherent manifold of experience to guide and manipulate concentration. Introspection does not synthesize objects or other individuals, for these are coherent targets of experience prior to our attending to them. In other words, even minimal, entry level consciousness allows of an entry level synthesis of the manifold of experience. However, what attention contributes to the flux of introspection is transiently and temporarily fixing the flux of introspection by reifying and hypostatizing the intentional operation – making an object of the ongoing process. The phenomenology is a misleading and distorted one that reifies the intentional object to which it attends, rendering what should be a process

as a thing, what should be a subject into an object. There is a pseudo-phenomenology but it is like what Kant would have described as a dialectical illusion – it hypostatizes the intentional attitude or propositional attitude into an object and misleads the would-be introspecting individual to results that are dubious philosophically such as inferring a thinking thing, an immortal soul, the illusion of a thing self (which, however, can never be reidentified, a homunculus like ghost in the machine).

In attempting to avoid the reification trap, another risk opens up. As Goldman notes: “. . . It [Introspection] can refer either to a process of inquiry, directed at mental states, or to a process of answering such an inquiry. In the former sense, introspection is inwardly directed attention, which chooses selected states for analysis. In the latter sense, it is the process of analyzing or classifying selected states. In the remainder of this discussion, we focus on the latter process: analysis or classification” (2006: 246). In asking (analyzing) what is the input to the transduction process of attention, the properties identified will therefore be identified as input to the very attentional process that identifies them. While this is not a regress, it does send the inquiry into a processing loop. This is perhaps why Goldman decided to reject three out of four forms of input – functional, representational, phenomenal. The neurological escapes the loop by being backstopped as the sense of inner equilibrium of a chemico-endocrinological bodily milieu.

However, a completely different approach is available from pathology as when the functionality is missing or disturbed. This avoids the processing loop at least from a functional perspective. Functionally, we know attention best when it is missing or hyper active. “Attention deficit” – whether the result of a medical diagnosis or not – is experiencing seemingly trivial events in the environment as like many beepers going off in sequence, seemingly demanding one’s focused concentration. This may be due to a low threshold for distractibility, an absence of the usual filters and blocking systems that keep most people on automatic pilot doing whatever they are doing, or the spontaneous operation (“discharge”) of a separate attention mechanism. The latter case would of course contradict Gilbert Ryle’s famous assertion that attending (“heeding”) is a modification of whatever activity is being undertaken – reading attentively, driving

attentively – and not some separable process. An alleged standalone attention would be like some sentry standing guard against the arrival of the enemy who never shows up. But upon more careful inspection this is not a counter example to Ryle, but rather a privative mode of attending with the underlying activity a form of expectation (“watch out for the enemy”) that just seems empty. In contrast, attention deficit does seem to be a counter-example to the allegation that attention is always dependent on another activity. Attention deficit posits an attentional activity that is not given enough material on which to attentively operate, and so it finds its own objects of attention seemingly at random. This produces the characteristic behavior of students who turn on the TV, activate instant messaging, and enable their cell phones prior to sitting down to study. In other instances, if the individual is in a quiet milieu – say trying to write a chapter in a book such as this one - without interruptions from the environment, having turned off the phone and TV (etc.), then the individual will invent his own distractions such as craving for a snack or “need” for a drink of water. These are content, but what makes them salient is a free floating attention that highlights them when they would otherwise be ignored or fade into the background as indistinguishable noise. It is to be noted that the different levels interact and the functional and representational properties of attention will elaborate the internal biological milieu for their own purposes, attending to them or not. The internal milieu at another level also includes random thoughts, day dreams, and one’s mind wandering, that defeat concentration on the task at hand, assuming there is a task.

**Representational:** The inputs to the attentional process are representations of propositional attitudes. At the risk of a bold statement of the obvious, believing relates to its content as a belief pure-and-simple, that is, it relates doxically; desire relates longingly; imagining relates fictionally or pretentiously (to coin an existing word - in the manner of a pretence). In classifying the intentional attitude that is operational in a given situation, identifying the relationship between the content and the attitude is the key to identifying the type of mental state. Attending to the relationship between the content and the attitude of an on-going occurrence of belief, desire, emotion, enables the classification of the attitude to occur. This is introspecting in order to classify an

intentional attitude, *not* in order to satisfy an inquiry such as answer the question “Why the heck did I do that?”

Along the same lines, Searle suggests direction of fit as the defining characteristic of attitude (“mode”) with varieties of belief having a mental state to world direction of fit and desires having a world to mental state direction. If I am confused about whether I want it to rain so badly that I imagine it is raining and now believe it is raining – a form of wish fulfillment – I look at whether I am trying to make the world look like my belief or whether I am trying to make my belief correspond to the world. The latter would be an objective approach in comparison with wish fulfillment. In this instance, we are on the right track since the direction of fit provides the clue to the attitude. As a farmer facing a drought, I come to my senses and realize my intense wish for rain brought into being a belief in the likelihood of rain, which, however, is just that, a wish, not a realistic belief. That part of the emotions not reducible to desire or belief – the irreducible affective part of the emotion – has a reciprocal direction of fit – back-and-forth. It discloses the world like a direction of fit mental state to world, and, at the same time, provides a motivation to make the world different, because the crops matter to the farmer, in a world to mental state direction of fit. The reciprocity of the two directions is characteristic of the emotional dimension, though Searle does not make it explicit. It is only partially a joke to note that is why the emotional dimension often leaves an individual spinning in a circle, at least emotionally. The individual is trying to align two simultaneously resonating directions of fit rather like a multi-stable picture (the “duck-rabbit” or Necker cube) and the tension characteristic of an emotional reaction is qualitatively a function of the oscillation.

**Phenomenal:** How does the feeling of a belief differ from that of a desire or a hope? And how do I know that it is the attitude that I am experiencing, not the content? We know that it is not the content because the content such as “that it is raining” is indifferent as regards the attitude. It is only when the content is implemented by an attitude that it gets suffused, as it were, with the attitude. The latter is reflected in the content which takes on the quality of the attitude rather like tofu takes on the taste of whatever food it is

cooked with. Or apple juice take on the shape of the container (belief, desire, hope) into which it is poured. There is no separate phenomenology of the intentional (propositional) attitude in and of itself that might be introspectable in the stream of inner sense. It remains dispositional and therefore invisible (transparent) until it implements some particular “aboutness” whether that is the weather, the comfort level of an individual’s biological equilibrium, or the content of the page I am reading in the book in my hands. In the case of desires and emotions there is a strong feeling (affective) component to the entire constellation of desire and emotion, including both the attitude and content (quality and matter). In the case of beliefs, doubts, conjectures, imaginings, and other intellectual type operations, the accompanying feeling – an intellectual quale - is more subtle. In the case of belief and doubt, the feeling evokes a feeling of trust or mistrust in the positing of what is being asserted. In the case of imaginings, the feeling of pretending predominates, but exactly what that feels like is hard to specify and is a function of what is being imagined, accompanied by the negation of reality.

**Neurological:** As noted above, there is an internal milieu involving the organism’s chemical-endocrinological equilibrium that percolates up into consciousness through the ventromedial nucleus of the dedicated thalamocortical relay (Goldman 2006: 252).

“Interoception is the perception of the physiological status of bodily tissue [ . . . ] Introspection, as I propose to understand it, is a further response to these sensations . . .” (Goldman 2006: 252). Well and good. Neurons and their related process have properties such as sodium ions gradients that pump charged chemicals across the neuronal membrane interface, recovery times, and other chemical and biological transformations. None of these properties are available for introspective access. However, by the time, these events percolate up out of the pulsating cauldron of biological infrastructure, they are phenomenal properties, corresponding to the interoceptive perception of the physiological status of the body.

Ultimately, none of the ways of redescribing the subsymbolic structure of introspection is adequate for Goldman. He highlights the neurological as the input to the transduction procedure (“function”) whose output is the classification of the intentional attitude.



However, this aspect of the neurological morphs into the inner milieu of chemical-endocrinological equilibrium from which the kind of distinctions are missing that would be useful to propositional content. The phenomena do not support the conclusion, and Goldman realizes this even if he does not explicitly acknowledge it.

However, a telling observation occurs at this point. Do we ever phenomenally experience an intention that makes explicit the distinction between content (matter) and attitude? What would a pure intentional attitude be like and would we ever find one swimming in the stream of consciousness? The issue has occurred to Goldman too.

### **Moving to an Alternative Approach**

Thus, Goldman moves to an alternative approach. He proposes and posits a proprietary introspective code similar to lexical items for encoding linguistic meanings (Goldman 2006: 262). The inner dialogue is not explicitly named, but it *lives* as a grasping of meanings, the introspecting subject, in the act of knowing. If an individual wonders whether he really *believes* that it is raining (or that religion is the opiate of the masses) or merely *conjectures* it or perhaps even *imagines* it, then this individual straightaway “reads off” the intentional type (i.e., the attitude/quality/vehicle) from the operation of the act of belief or conjecture. A distinction is introduced between meaning and the introspective experience of meaningfulness. This is not introspection in the usual sense of the term. However, there is precedent for it. The precedent is to be found in the Husserlian *wesensschau*, intuition of meaning (“essences”). Under this interpretation, one reads off the meaning of belief, desire, empathy, from the act itself.

Make no mistake, that the essence that is intuited in categorical intuition is not reducible to a Fregean sense (Sinn), though in cases of linguistic expressions, there is significant overlap and convergence between Husserl and Frege.<sup>21</sup> The overlap does not extend beneath the level of explicit aphoristic statements. It is a contingent matter of fact that

---

<sup>21</sup> Marina Paolo Banchetti. (1993). “Follesdal on the notion of the noema: a critique,” *Husserl Studies* 10: 81-95, 1993. This was discussed in detail in Chapter \_\_\_\_ on Empathy and Intentionality.

the *epoche* makes available the noema of constitutive acts of believing, desiring, seeing, empathy, etc., not the noesis. It is the noema that is directly accessible, not acts of constitutive noetic empathy. Husserl was himself driven in the direction of this conclusion by the incomplete access to the noetic acts constitutive of experience

When all else fails and an individual is still not sure whether he believes or desires or hopes or fears that *p*, then the individual employ an ascent routine. Robert Gordon illustrates the ascent to classification of the intentional attitude by means of belief:

“...The way in which adults ordinarily determine whether or not they believe that *p* is simply to ask themselves the question whether or not *p*” (1996: 16).<sup>22</sup> While contingent obstacles may block a robust empirical procedure for determining whether or not *p*, contra Goldman (2006: 24), the method is completely generalizable to other intentional attitudes based on the mode of fulfillment of the proposition *p*. It is not restricted to just belief. If an individual is trying to determine whether the intentional attitude is desire (wanting), imagination, hoping or fearing or other emotion, recollecting or expecting, then a procedure to distinguish whether the proposition is fulfilled by a truth value, an evaluation of volitional worth, an assessment of affective well-being, an estimation of aesthetic pleasure, or other candidate condition of satisfaction. It is important to note that it is no more necessary that properties such as being desireable, pleasureable, attractive, useful, exist as that any candidate belief be instantiated. For the sake of an intentional attitude, fiction will do as well as fact. Indeed sometimes even better.

Here simulation as an ascent routine is very powerful – the individual imagines the outcome – he posits the result and draws out the consequence(s) – this feeds back to the propositional attitude and refines it – yes, he realizes, that it must be that he is primarily hopeful and he hopes all the more fervently. This makes it particularly surprising that Goldman is so short in dismissing it (2006: 240) perhaps because Goldman has overlooked a robust definition of attention that cuts across the trap of the inner/outer duality. With an ascent routine, a kind of epistemic seeing (inspired by Dretske) is

---

<sup>22</sup> R. Gordon. (1996). “Radical simulationism” in P. Carruthers and P. Smith, eds. *Theories of Theories of Mind* (11-21). Cambridge: Cambridge University Press.

deployed. The individual looks “outward” at the satisfaction conditions of the belief or desire in order to access what is the intentional attitude. This is an approach that would work for the third-person as well as the first-person.

The bottom line for empathy as an attitude/quality/vehicle that intends the other is this. The externalism of the intentional attitude (in this case, empathy) is preserved in that, speaking in the first-person, I attend to the other in order to gain further fine-grained distinctions about my experience of the other; and the introspective availability of the empathic attitude allows first-person access (albeit one that is subject to error and revision) to the experience as mine. The “mineness” gets coded as “internal” and indeed is inward in that the particular content and the way of encoding it right now is available only to me. Working from the bottom up, in introspection the vicarious experience of the other is encoded as mine, though as caused by the experience of the other. The sensation of the other is encoded as mine, but as caused by the sensation experience of the other. The pain of the other is encoded as mine, but as caused by the pain of the other.

Working from the top down, as the vicarious experience occasioned by the encounter with the other is subjected to further cognitive processing in empathy by distinguishing self and other. The vicarious experience of the other is encoded as if it were mine but as caused the experience of the other; the sensation of the other is encoded as if it were mine but as caused by the sensation experience of the other; the pain of the other is encoded as if it were mine but as caused by the pain of the other. This perhaps does not completely exhaust the microstructure of introspecting particular sensations or affects; but it goes a long way towards explicating the details. Now let us continue ascending the chain of phenomenological redescriptions of introspection towards a top down perspective. A different way of encoding the experience of introspection is to be found in a cultural artifact, introspection as meditation.

### **Introspection as Meditation: The ‘Voice Over’ of the Internal Dialogue**

While introspection may seem to be the ultimate act of a private individual in ruminating on the issue of who the individual is and what life is about, such meditation is not an

inherently isolated activity. It is a culturally acquired practice, a form of life that requires training and discipline. But, one might object, what about Descartes, closing himself up in a warm room to meditate on what he can know without doubt and with certainty? We all too easily forget that a meditation is a literary form that takes a rhetorical position. It is that the reader (or listener) is getting a glimpse into the world of the author's on-going stream of consciousness. It is amazing that, to the best of my knowledge, no one else has called attention to the rhetorical stance of the introspective mediation.<sup>23</sup> It is not a fiction that the author (Descartes) is sharing many things that, up until time of publication (1642), were contingently private, perhaps known only to the author alone. What *is* a fiction is that these things were in any way intrinsically private or inaccessible to others. The meditation is the demonstration of the accessibility in so far as it is a narrative device shareable with its audience.

Thinking is defined as the dialogue of the two in the one, and described by Socrates in Plato's *Theatetus*.<sup>24</sup> If that definition is accepted tentatively for the sake of discussion, then one interpretation of introspection is a conversation in the form of a dialogue on which the meditating thinker silently eavesdrops.<sup>25</sup>

As one reads the first of Descartes' *Meditations*, the rhetorical stance is that the reader is being given a glimpse of and following along with the dialogue Descartes is having with himself.<sup>26</sup> Descartes asserts that his senses may deceive him and a wise man never entirely trusts those who have once cheated him. The interlocutor argues back and responds that only a madman would doubt that he is sitting by the fire, holding this paper in his hands. Descartes then answers himself back as Descartes: "A fine argument!" and that he may well be dreaming, which is a transient form of madness to which even

---

<sup>23</sup> Lyons asserts that introspection is a cultural artifact useful for understanding other individuals, and he then proceeds approvingly to quote Hobbes's argument from analogy for the existence of other minds. William Lyons. (1986). *The Disappearance of Introspection*. Cambridge, MA: MIT Press, 1986.

<sup>24</sup> Plato, *Theatetus*, tr. F. M. Cornford, Indianapolis: Bobbs-Merrill, 1957: 189e.

<sup>25</sup> See also "The two-in-one" in H. Arendt. (1971). *The Life of the Mind*. New York: Harcourt, 1978: 179. As I read it, this is not Arendt's position or is an over-simplification of it, as she does not mention "introspection" explicitly. She takes off from Socrates positive statement that it would be unfortunate that he (Socrates) "being one should be out of harmony with himself" (1971: 181).

<sup>26</sup> R. Descartes. (1642). "Meditations on first philosophy," 2<sup>nd</sup> edition, in *Descartes: Philosophical Writings*, tr. E. Anscombe and P.T. Geach. New York: Nelson, 1966: 62.

ordinary individuals are subjected daily (nightly). This casts into doubt all empirical sciences. But what about arithmetic, geometry, and the mathematical ones, wonders the interlocutor, reversing roles? Don't these stand fast, hopes Descartes. To doubt these, an evil genius is required, as suggested by the interlocutor, the eventual answer to which requires a proof of God by Descartes. Thus, without splitting too many hairs about who is in what role, the point is that the conversation has the form of a dialogue.

The obvious question is who is the interlocutor? One of the contents of introspection is the running commentary on our experiences and interactions in the world provided by the "voice over". The voice over is providing a cynical and, properly speaking, skeptical commentary on Descartes' project of attaining certain knowledge. In the individual's everyday experience, the voice over is usually commenting on how our neighbor looks ("that is a strange haircut"), what we really think about what he said ("where did he get *that* idea?" or what is happening to us ("I hope I'm not late")). The commentary is so diverse and heterogeneous that it defies easy description or listing, but it can be omnipresent and persistent. If you do not believe in such a thing as the voice over, who was it that just told you to doubt such a thing? The voice over even reminds you, "There's no such thing." The voice over gets coded as "mine"; and, arguably, it does support the survival of the organism in which it is implemented and animated. But as a function of patterns of the past that previously worked well enough to ensure survival, it is not particularly creative, innovative, or even adaptive going forward in new and challenging situations not previously envisioned. In short, it is common.

The interlocutor is "the voice over." This term is used as an update to Descartes' experience and analogously to the voice over the events in television commercials that provide a commentary on the action in the commercial in which the product is center of attention. This is just another tactic used by the marketing influencers to "get inside the head" of the audience.

The "voice over" expresses – but is not limited to – the accumulated experiences of parents, teachers, friends, bosses and subordinates, opponents – real and imagined –

filtered through one's hope and fears, doubts and aspirations, agreements and disagreements, affections and antipathies. For example, after a stressful day, i.e., one filled with rapid fire activities, chemical enhancements such as double espressos, and multiple simultaneous conversations trying to convince and persuade the boss, wife, or teacher in the face of opposition and even resistance, it is useful to take a deep breath and relax. That is when an individual is able to hear just how loud the voice over has become. It is loud enough to drown out attempts at certain forms of thinking that require great attention to fine details. The force with which these contents are presented by the voice over borders on the compulsive – one cannot voluntarily make it be still by means of a single, simple statement to oneself to stop and be quiet. If one goes for a jog, say, to distract oneself, the voice over will continue to stream off as if being played back; and, by the end of the jog, the accumulated stress will have been discharged, leading to a lower level of compulsiveness for both the body and the consciousness embodied in it, freeing up cycles for self-guided mental activities to be undertaken. This experience of “streaming off” is the experiential (“phenomenological”) basis of Lyons description of introspection as the perceptual replay of experience, about which more will be said shortly (Lyons 1986: 113f.). There is another method that readily comes to mind for quiescing the voice over.

Recall the running commentary on an individual's experiences and interactions in the world provided by the “voice over” - commenting on how our neighbor looks (“that is a strange haircut”), what we really think about what he said (“where did he get *that* idea?” or what is happening to us (“I hope I'm not late”), etc. Often people are having a conversation with themselves in the presence of the other - a lively give-and-take with the voice over - which is one reason so little communication with the actual, other person is occurring.

One of the consequences of being listened to empathically and heard is that the voice over is quiesced. One of the consequences of empathy as a gracious, generous listening is that the compulsive quality of the “voice over” is stilled. The voice over's commenting on everything the individual says to the (real) person with whom he is talking as well as

commenting on what the individual thinks to himself is quieted. The volume on the voice over is tuned down. It does not need to be so loud, because the person has been listened to and heard. The person feels understood and does not need to argue either with himself or the other. As a result of the stilling of the constant commentary, the individual is then able to participate in a real conversation – this time with another person present live and in person - and be fully present to the other and as another to the other.

As a result of quiescing the voice over, the individual is able to listen with an attention and accuracy that was previously not possible or very rare. Although empathy is deployed as a way of influencing the introspection of a given individual, it is not the only way. It should be noted that similar results – quiescing the voice over - are reported as a consequence of certain forms of solitary meditation or spending time in a sensory deprivation tank, the latter regardless of whether one meditates or not (a separate spiritual technology).

From another perspective, this reenacts and resolves the controversy about imageless thought that erupted between Wundt and Külpe in the glory days of the laboratories in Leipzig, Würzburg, and Cornell. In fact, both are experientially grounded, and intermediate phenomena such as the “voice over” can support either approach, depending on the experiential intensity with which the replay is implemented. Starting from epistemic seeing, one gets various forms of auditory, visual, and tactile imagery. Starting from verbal thinking and the internal dialogue, one gets forms of symbolic manipulation (“thinking”) such as calculating, following rules, taking turns in formulating opinions on a given topic (“dialogue”), means-end reasoning, “as if” speculating about conditions contrary to fact, are highly opaque – i.e., available to introspection.

It has become common place to acknowledge that much more is going on in consciousness than to rule-governed symbol manipulation, including by the proponents of a language of thought.<sup>27</sup> In fact, symbol manipulation and verbal thinking captures

---

<sup>27</sup> J. A. Fodor, (1983), *The Modularity of Mind*, Cambridge, MA: MIT Press, 1983: 120 where Fodor acknowledge “epistemic boundedness” as the in principle unavailability of certain forms of self-knowledge.

only a small subset – albeit a significant and important one, central to sharing one’s thoughts with others – of the processes that form human intelligence. As a rule of thumb, the closer the processes are to the high level cognitive functions of the brain and the farther away the processes are from the peripheral perceiving functions, the more visible (i.e., opaque) they are and susceptible to attentional grasping. The results of neurology have demonstrated that the awareness of the individual of the processes subserving thinking does not reach very far down or inward in the direction of brain processes; cognition is floating on top of a complex biocomputer, “wetware” that can be separately investigated using third person methods such as MRI or PET technology. The point is to highlight the structure of one form of introspection – manipulation of symbols, including a dialogue with oneself. In some instances of thinking, the individual has to try to imagine what is the response to the inquiry being engaged - for example, “Was that a duck or a rabbit that I saw?” – in an effortful way. In other instances, the answer seems to spontaneously be suggested in a form of words that sound as an echo out of one’s thoughts including “the voice over.”

### **Introspection as ‘Free Association’**

Of more immediate interest is the form of empathy that corresponds to introspection as the conversation with oneself. Here Freud is an important witness. If one delves deeply into oneself and with the necessary combination of acceptance, suspicion, and interpretation of what is available to introspective attention, then one can attain a measure of self-knowledge. It is an essential part of the cunning of psychoanalytically oriented therapy that an accepting milieu is established in which what was hidden (whether pathology or everyday unconsciousness) gets expressed. While the results are conditional and subject to revision, a measure of self-understanding is attained. At least enough is attained to transform the pain of neurotic symptoms and maladaptive behavior into a combination of productivity in one’s work, affection in one’s relationships, integrity and cohesiveness of one’s sense of self, and, optimally, humor and an enhanced appreciation of art.



No conversation is quite like the standard psychoanalytic dialogue. Freud asked his patients to face away from him and to try to have a conversation with him without his being visibly present face-to-face in the ordinary way that people talk. This single act, facing away, significantly transformed the conversation from one between two individuals – Freud and his patient – to one between the patient and patient – with Freud able to eavesdrop on the process as if it were a contingently personal one known only to the patient himself. Thus, if the patient allows it, the doctor aligns himself with the patient's own attentional processes, indeed sometimes attending to connections between contents and attitudes of which the patient himself is unaware or only dimly aware at the time. Temporarily the doctor transiently forms a part of the patient's psyche by being associated with characters and roles from the patient's past and current life situation.

What is occurring here is a method of structuring introspection so that rich material is surfaced yet a measure of control is maintained over the process, allowing data capture to occur in such a way that the material can be communicated to another and fixed in the context of human interrelations. This includes the possibility that there are thoughts, attitudes, emotions that occur to an individual that the individual does not wish to acknowledge. Such contents are objectionable whether for reasons of custom, ethics, habits, or other implicit or explicit standards. The attitude towards them is denial, rejection, projection, or other mental manipulation.

In talking with a patient about the distress that has caused a patient to seek consultation with him, Freud (and Breuer) developed the technique of free association as a way of rendering heretofore unacknowledged ideas and affects available for attention. The method of free association is a description of a form of introspection that aligns with the inner dialogue of the meditation.

However, the rhetorical stance of this approach reverses the standard definition of mediation. A meditation is supposed to be personal and (contingently) private; but it is published to the world in book form. Free association is supposed to be a conversation with another, and so public, but it is actually a private soliloquy, which, however, another

is enabled to overhear and transiently participate (in strict confidentiality). The significance for empathy will become obvious. Freud recommends to the would-be patient:

Your talk with me must differ in one respect from an ordinary conversation. Whereas usually you rightly try to keep the threads of your story together and to exclude all intruding associations and side-issues so as not to wander too far from the point, here you must proceed differently. You will notice that as you relate things various ideas will occur to you which you feel inclined to put aside with certain criticism and objections. You will be tempted to say to yourself: 'This or that has no connection here, or it is quite unimportant, or it is nonsensical, so it cannot be necessary to mention it.' Never give in to these objections, but mention it even if you feel a disinclination against it, or indeed just because of this. Later on you will perceive and learn to understand the reason for this injunction, which is really the only one that you have to follow. So say whatever goes through your mind. Act as if you were sitting at the window of a railway train and describing to some one behind you the changing views you see outside. Finally, never forget that you have promised absolute honesty, and never leave anything unsaid because of any reason it is unpleasant to say it (Freud 1913: 147).<sup>28</sup>

This passage has been quoted at some length because it is so rich in distinctions relevant to introspection. First, the patient lying on the couch in Freud's consulting room is positioned so that the listener is behind him or her. Thus, Freud's suggestion to the prospective analysand to describe to "some one behind you the changing views you see outside" the window of the imaginary railway train. The person behind is, of course, the analyst. As the analysand lies back on the couch, based on the photography by E. Engelman of the consulting room at Bergasse 19, Vienna, 1938, he or she would have a view of cabinet with various primitive totemic statues and a plaster relief of a female figure (Jensen's *Gradiva*). Such statues and images helped to stabilize an evanescent idea and keep it from escaping altogether, according to Freud (1929/39: vi). Furthermore, though not explicitly discussed in this passage, the invitation is to allow the statues to evoke memories and figures from the past of the person lying there giving an account of his or her experiences relating to the illness.<sup>29</sup>

---

<sup>28</sup> S. Freud (1913) "Further recommendations in the technique of psychoanalysis: On Beginning the treatment, tr. J. Riviere in *Therapy and Technique*, intr. P. Rieff. New York: Crowell-Collier, 1964: 147.

<sup>29</sup> S. Freud. (1929/39). *The Diary of Sigmund Freud: 1929-1939: A Record of the Final Decade*, tr. Michael Molnar. New York: Charles Scribner: ii-iii.

In addition to positioning himself out of the line of sight of the patient, since, according to Freud, the doctor does not want to be looked at for eight hours a day, the advantage of the unconventional physical positioning is to allow the free play of the imagination of the analysand (speaker) in narrating events. In addition, the voice coming at the analysand from over his or her shoulder literally takes on the quality of a “voice over” – over the shoulder – and an imaginary quality that allows it to be amalgamated with significant figures such as parents, partners, spouses (at least transiently and occasionally).<sup>30</sup>

What is not pointed out by Freud here is that the material that wells up out of the depths of consciousness has *already* been edited. The material – whether memory, fantasy, or nonvisual sensory-affective constellation - that is accessed for reporting – within the metaphor, appearing in the window - has already been selected, condensed, elaborated, tidied-up, simply in order to be able to be communicated at all. What Freud is cautioning against by his explicit remarks is *additional* editing that will prevent the recollection, etc. from being communicated at all.

For example, say that the recollection is of one’s earliest memory – a view of one’s nursery from the perspective of inside the crib as one tries to climb out of it but instead falls into the narrow space between the bed and the wall and recalls feeling helplessly trapped, suffocated, unable to breathe, and really afraid – really afraid until Mom comes to the rescue and plucks one out of the abyss. There are no words spoken here within the story, though the experience has to be captured and communicated as a narrative – being squeezed through a narrows channel and united with Mom, this time from the outside instead of inside the womb. The narrator really did climb out of his crib and fall into the narrow space between the wall and the crib; and this experience provided a narrative framework for expressing an earlier, even more primitive experience that left an impression prior to the availability of any language to structure or fix it in verbal memory. The primary sensation here is that of being squeezed in a narrow crevasse, what

---

<sup>30</sup> H. Smith. (2001) “Hearing voices: The fate of the analyst’s identifications.” *Journal of the American Psychoanalytic Association* 49:781-812.

the narrator described as the space between the crib and the bed, accompanied by a feeling of helpless, being unable to breathe, and anxiety. The empathic listener, attending to the sensation of suffocation and the accompanying affect of fear, gets a sense of what it was like for an individual to experience the trauma of birth.

Considering introspection as a form of literature, a meditation, in which an individual has a conversation with himself led, in turn, to a structured interaction between patient and doctor. The patient free associations based on whatever occurs to him about the details and situation of the illness causing him to consult the doctor, who, in turn, listens with an empathic and even hovering attention (about which more shortly) and responds (interprets) on that basis. But what is it that the individual is supposed to be sharing without censorship or selection, at least to the best of her or his ability? At this point, introspection has become retrospection.

Thus, introspection is not restricted to the so called inner monologue, personal soliloquy, or voice over. There is a rich store of images, sensations, and affects which well up out of the stream of consciousness, usually but not exclusively having their source in the past or imaginative variations on the past. However, once again, the individual's access to this subsymbolic content is through language and the conceptual reidentification made possible through language.

This is why this chapter began the engagement with introspection from the bottom up – salient sensations and affects are often the access to deep experiences, even though these are initially inaccessible without being couched in a verbal form. A raid on the inarticulate has occurred, and, at the risk of a spatial metaphor, the booty with which one escapes does not always shine as brightly and mean the same thing in the robber's tent of science as it does in the mythical mansion of originary experience ("mythical" because the "myth of the given" remains). The gamble is that what does survive the transition from introspective process to publicly shareable narrative is the meaning of experience, which can be captured, recovered, and fixed. There is something there, but by the time introspection is done trying to process it, it is no longer a process; it is an opaque

component of the self that has been hypostatized into a *subject*-object. This results in the sense of introspection as the fixing of the flux, the attempt, only partially successful, to identify one's subsymbolic stream of consciousness.

### **Introspection as Retrospection**

The metaphor of the window of the train compartment in which one is sitting fits nicely with William Lyons positive account of introspection as a replay of perception:

. . . [W]hen we "introspect" in order to discover things about our cognitive, appetitive, and affective lives, we engage in a process of perceptual "replay." We "replay" or recreate – at least with "edited highlights" or in "dramatized form" – what we think we said, or would have liked to have said or did or felt. For example, if we "introspect" in order to discover whether we love someone, we "replay" how we look or reacted and what we said or did, or might have or were going to or had it in mind to) when he or she, the beloved, was present on the topic of conversation or when the beloved was noticeably absent. "Introspection" is not a special and privileged executive monitoring process over and above the more plebeian processes of perception, memory, and imagination; it is those processes put to a certain use. . . .

. . . [M]y central contention is that "introspection" is our employment of perceptual memory and imagination to find out about our motives, thoughts, hopes, desires, and the like by finding out about our published, revealed-in-speech-gesture-expression-and-behavior motives, thoughts, hopes, desires, and so on, for these are all we have conscious access to. . . [W]hat we can experience [and use is] . . . the perception-based processes of memory, imagination (and dreaming, daydreaming, hallucination), the rest lies hidden in the brain" (1986: 113-114).

The engagement with Freud will continue, and include the perspective of retrospection. Lyons does not make much of the distinction between processes that are dynamically unconscious such as objectionable desires that are repressed due to conformity with social standards, which Freud is interested in detecting and unmasking, and those cognitive processes to which the individual is autoepistemically closed and will never have access because, e.g., they are neurologically below the threshold of consciousness by definition. The former are precisely the contingently unconscious desires that Freud is allowing to betray themselves through their indirect influence on the free associations of the analysand. Notice, however, that such desires, intentions, ideas, beliefs, and affects will show up as disruptive content, not standalone vehicles.

Granted that attending to a memory introspectively invokes the use of additional cognitive resources, anyone who thinks that introspection involves concentrating his attention in order to discover mental contents (affects, volitions, cognition) of consciousness will be disappointed and soon become disillusioned with the prospect of using an introspective technique at all.

The harder one concentrates one's attention, the more elusive and one-sided are the results. It is not merely that the person only finds answer to questions he had already posed, but that he suggests the answer to himself. In a text cited later in this Chapter, Freud explicitly warns against such an introspective bias as a part of his account of a use of evenly-hovering attention, Freud's positive account of introspection on the part of the analyst. This concentration of attention leads to the disrepute that has plagued introspection and, by implication, empathy.

Furthermore, it is important to note that memories, which are often the content of introspection, behave in a parallel manner. That is, recollections are rarely disclosed by an intense concentration of attention. A forgotten name or foreign word may be on the "tip of the tongue." Typically, exerting extra effort, trying harder, will get one nowhere. But if one is distracted by something for a period of time, then all-at-once the name or word will just pop into one's mind. The memory retains a certain autonomy of its own. Although taking the right attitude towards the memory can make the difference between a reliable and an unreliable memory, it is not completely subject to voluntary control.

The metaphor of looking out the window of the train and reporting on the passing landscape points in the direction of introspection as retrospection of memories, dreams, fantasies, and aspects of a sensory and affective content. This is not exclusively so, since one may comment on the sound of the tea pot whistling and how it makes one feel. But how it makes one feel often invokes a pattern of events into which one was thrown in the past - I was making a cup of tea least New Year's and the door bell rang . . . Of course, our access to these contents will be by way of a verbal, narrative description. But in at

least some instances it might be more appropriate to paint a picture, a series of pictures, or even make a short film. These memories and dreams and fantasies and ideas and affects are coded as “inner.” They do not hover in space before the individual as if they were a power point slide projected on the wall. The individual has a vague sense that something = x is occurring, running off, in the dark just behind the eyes or over the right shoulder or just at the edge of one’s peripheral vision or in one’s chest region. However, as soon as one attends to it, it evaporates, disappears. The vague sensation runs back into its box, as if it was a beetle, and pulls the lid shut behind it. By the time one has attempted to capture it in verbal form, it is no longer an ongoing process, it has been reified into an object that is derivative on the verbal narrative with which, therefore, it is in agreement in essential details.

Thus, to return to Freud’s injunction about free association after this roundabout analysis, the metaphor of describing what is happening at the window of the railroad compartment in which one is riding has a particular target. What appears there is the input to the process of attending to a dream, screen memory, or other experience. Freud does not explicitly extend the analysis to mood, but they belong on the list. It is not the target of the individual’s communicative activity in all its evanescence and allusiveness. It is what is available for attentional introspection without being conceptualized or captured as a metarepresentational concept or cognition. It is the unfixed flux that one calls out as it goes by without the chance of reidentifying what one saw. Was it a cow or a horse? A duck or a rabbit? A senior citizen (“old woman”) or a vase? Were the flowers yellow dandelions or daffodils? What the bread wheat or rye? What is given recedes towards a point on the horizon labeled “myth of the given.” Yet something = x is captured and analyzed in such a way that the individual has more experience and knowledge - albeit corrigible, subject to revision, and finite - than that with which he began.

This is parallel with what happens in writing down a dream or a screen memory. The material that emerges in the dream or recollection is rich in sensory images, powerful affects, and so-called secondary properties such as tastes and smells. In fixing the material into verbal form, the material undergoes a secondary elaboration that transforms

the sensory and experiential into the verbal and linguistic. “Secondary elaboration” is what is required to put the dream, screen memory, or fantasy into a form that is grammatical enough to be communicated at all. Even if the narrative jumps around, includes side issues, mentions “unimportant” details, and is nonsensical in its transitions and contents, it is still a narrative. The narrator must take care to describe “the bright yellow color of the dandelions that girl handed to the narrator” or the “delicious, fragrant taste of the piece of black bread that the peasant lady handed to the children.” Yes, that is what was contained in the experience being described. But the inevitable result is that the immediacy of the experience recedes and the fixed format endures. The experience cannot always be recovered from the narrative, though the experience cannot be accessed at all without the narrative.

### **Introspection as Evenly-Hovering Attention**

In many ways, the invitation of free association is for the analysand to give her- or himself over to the dangers of fictionalizing introspection, provided analysand’s integrity is intact and what is expressed is what one really believes in the moment. It is precisely such an approach that will enable the pathological unconscious content, intentions, influences, etc. to be expressed, even if only indirectly in slips and symptoms. Since the slide into certain amount of story telling – confabulation – is just human nature in trying to give an account of one’s neurotic symptoms, the tactic is to make a virtue of necessity. Do not try to prevent what will happen in any case; but structure and influence it in a way to minimize self-deception and prepare the way for unmasking.

In contrast, such is not the case with the introspective process in the analyst, at least not as regards his listening. The exercise of “deliberate attention” is “dangerous,” according to Freud. This does not mean dispensing with the use of attention altogether; but points towards a special use of attention. Freud has a specific recommendation as to how the analyst can avoid skewing his listening, making biased selections of material by preexisting expectations and inclinations.



This recommendation will enable an empathic listening to be formed in an introspective context, and, vice versa, a rich, introspective milieu to be formed in an empathic context. In particular, symmetry between empathic and introspective attitudes is enabled. The physician is listening with empathic receptivity to the other, and he is also introspectively receptive to any patterns that may spontaneously emerge in the other's free associations as well as in his own reaction to what is being communicated:

The technique, however, is a very simple one . . . It rejects the use of any special expedient (even that of taking notes). It consists simply in not directing one's notice to anything in particular and in maintaining the same "evenly-suspended attention" (as I have called it) in the face of all that one hears. In this way we spare ourselves a strain on our attention which could not in any case be kept up for several hours daily, and we avoid a danger which is inseparable from the exercise of deliberate attention. For as soon as anyone deliberately concentrates his attention to a certain degree, he begins to select from the material before him; one point will be fixed in his mind with particular clearness and some other will be correspondingly disregarded, and in making this selection he will be following his expectations and inclinations. This, however, is precisely what must not be done. In making the selection, if he follows his expectations he is in danger of never finding anything but what he already knows; and if he follows his inclinations he will certainly falsify what he may perceive. It must not be forgotten that the things one hears are for the most part things whose meaning is only recognized later on.<sup>31</sup>

This recognition of meaning is largely retrospective. It occurs "later on." The point is that the physician becomes aware of this recognition in himself without the use of any "deliberate concentration of attention." The use of the latter is counter-productive – even "dangerous" from the perspective of getting at the truth, to quote Freud - in that it reduces the chances of highlighting anomalies or surprises. The concentration of attention leads either to one's "never finding anything but what" one "already knows" or simply falsifying what one receives to confirm one's like and dislikes.

Note that "evenly-suspended" is *not* a lack of attention. It may be correctly described as a state of calm, quiet, alertness. One may well ask whether it would not be better to admit that a certain amount of selection is inevitable; and, therefore, to approach the

---

<sup>31</sup> S. Freud. (1912). "Recommendations for physicians on the psychoanalytic method of treatment," tr. J. Riviere, in *Therapy and Technique*, intr. P. Rieff. New York: Crowell-Collier, 1963: 118; *S.E.* 12: 111.

analyst's free associations with expectations instead of hypocritically pretending to be free of prejudice. However, this involves a distortion of what Freud was maintaining. He does not wish to deny that a selection of relevant material will occur. In fact, only a small fraction of the patient's remarks are important. What Freud wishes to point out is that no one can know in advance what will turn out to be meaningful in terms of the treatment. It is not the physician's conscious expectations that provide the criterion of selection in this context:

It will be seen that the rule of giving equal notice to everything is the necessary counterpart to the demand made on the patient that he should communicate everything that occurs to him without criticism or selection. If the doctor behaves otherwise, he is throwing away most of the advantage that results from the patient's obeying "the fundamental rule of psychoanalysis." The rule for the doctor may be expressed: "He should withhold all conscious influences from his capacity to attend, and give himself over completely to his 'unconscious memory.'" Or, to put it purely in terms of technique: "He should simply listen, and not bother about whether he is keeping anything in mind" (1912: 118).

Here Freud uses "unconscious memory" in the descriptive, not the dynamic and explanatory, function. That is, a person's memory is more extensive, has a wider range, than what he can actively recall at any one moment. The point is that the physician's mind is not a tabula rasa even if he is not bothering to keep anything in mind. The free associations to which he is currently being exposed are capable of arousing analogous fantasies, memories, dreams, reflections in him. Although he could not be sure in advance what would be aroused, it is the unconscious which leads the way in explicating the meaning of the situation in which interrelation is unfolding. The criterion of selection is what resonates with the unconscious.

Neither the patient nor the doctor makes a prior criticism or selection, though the one is talking and the other listening. So there is nothing that, in principle, excludes free associating and an evenly-hovering attention from being united in one and the same person. This state of affairs must have occurred in the case of Freud's own self-analysis, though mediated by frequent dream interpretation using journaling and written notes. This state of affairs also occurs in the case of the analyst who is simultaneously listening

to the analysand and attending, in an evenly-hovering manner to whatever associations, feelings, and ideas occur to him while listening, though, it should be noted, the analyst does not explicitly share the associations aroused by the analysand's discourse with the analysand.

Furthermore, the analyst is on guard against the inherent risks of an introspective method by means of the analysis of analyst's own counter-transference. The analyst is a human being, and, in spite of his own analysis and training analysis, is still subject to blind-spots, biases, limitations, and all the hazards of being human. For the psychoanalyst, the provisioning of such a method of guarding against the analyst's own blind spots and distortions of the use of introspection is the analyst's own self-analysis and the on-going analysis of the counter transference.<sup>32</sup> Thus Freud:

Other innovations in technique relate to the physician himself. We have begun to consider the "counter-transference," which arises in the physician as a result of the patient's influence on his unconscious feelings, and have nearly come to the point of requiring the physician to recognize and overcome this counter-transference in himself. Now that a larger number of people have come to practice psychoanalysis and mutually exchange their experiences, we have noticed that every analyst's achievement is limited by what his own complexes and resistances permit, and consequently we require that he should begin his practice with a self-analysis and should extend and deepen this constantly while making his observation on his patient. Anyone who cannot succeed in this self-analysis may without more ado regard himself as unable to treat neurotics by analysis (Freud 1910: 81).

One cannot guide another's introspection or help another to understand the influence of the unconscious on his or her behavior, emotions, and thoughts without plumbing the depths of one's own unconscious in direct confrontation with one's own limitations. It is interesting to read in this same article Freud's warning against the enormous power of authority and the power of suggestion that accompanies it. His recommendation is a modest one – to continue to work with patients through the analysis of the unconscious and not through suggestion and hypnosis.

---

<sup>32</sup> S. Freud. (1910). "Future prospects of psychoanalytic therapy," tr., J. Riviere, in *Therapy and Technique*, ed. P. Rieff. New York: Crowell-Collier, 1963: 80-1. *Standard Edition*, XI: 141–51.

When evenly-hovering attention works as designed, it discloses to the analyst the subtext of the analysand's conversation with himself. Unconscious issues are allowed to influence the conversation indirectly and to surface explicitly. The analyst's response ("interpretation") can have a soothing and calming effect if it leaves the analysand with the awareness that the latter "has been heard" for the first time ever in his pain and unhappiness. When accompanied by working through of the content and its integration with the healthy and functionally coherent aspects of the analysand's self, this provides the elements of a cure. It can also require repeated passes when the material is multidimensional, highly conflict-generating and fragmentary. Intermediate results can be incomplete and upsetting, which results in turn become the target for further gracious listening that leaves the analysand with the sense of "being heard," "being gotten" in his struggle for completeness.

Some authors have observed some mental pathology can be dissolved simply through making explicit the complex of context and attitude through a process of introspection, usually undertaken in the context of psychotherapy (Metzinger 2003). Just paying attention to things makes them better (in a psychological reenactment of the famous Hawthorn effect).<sup>33</sup> There is more than a grain of truth to the method by which addition processing resources are allocated to integrating the functionally incoherent aspects of the individual's experience into a coherent self representation. What Freud would add to this observation is that the process of making explicit what was unavailable will usually be accompanied by a process of emotional (affective) discharge. Absent such an "abreaction," the change in structure will remain theoretic and the improvement will not occur.

So far, the emphasis on therapy and technique makes it seem as though empathy is derivative on introspection. As a tactical tool to be deployed in order to bring an other's

---

<sup>33</sup> The famous study at Bell and Howell factory in Hawthorn Illinois showed that when the illumination was turned up, productivity improved. For control, the lights were then turned down – productivity improved then too. People could see their work just fine either way – paying attention to the process improved it (at least within the parameters of this particular situation). Hence, Hawthorn effect.

experience closer to the listener that is indeed the case. However, if you look again at the basic rule of psychoanalysis – to free associate and share the results with the listener – then it should become clear that, as it were, empathy is the tail that wags the introspective dog. Only in so far as the analyst has created an empathic milieu will the analysand succeed in overcoming his own censorship, hesitations, and resistances to communicating the inevitably embarrassing material. Only in so far as the analyst has created an empathic milieu, will the analysand trust the analyst, providing grist for the analyst's evenly-hovering mill. Empathy is indeed an important tool and tactic; but it is also strategic. In particular, the availability of the gracious and empathic listening in the context of analysis is what inspires the analysand to share himself without reservation, or at least without the usual reservations that one encounters in a conversation between acquaintances.

Thus, the discussion of an introspection and empathy leads through Freud's account of free association and evenly-hovering attention. The difficulty is to unite the spontaneity of free associating with the receptivity of evenly-suspended attention in one and the same person. It is the tension between spontaneity and receptivity that makes the use of an introspective method so difficult without the help of a second person's participation. Here, too, is the source of that resistance to introspection to which Dr. Kohut has called attention (Kohut 1959: 466). Most people would rather avoid the tension, and either make a premature criticism or select only that which they already know. However difficult this tension is to reconcile in practice, still it can be done. What is perhaps needed is further examination of the relation between introspection and empathy.

Now instead of introspecting the redness of the apple using Dretske's epistemic seeing – displaced perception – looking externally in order to get information about the state of the observer – consider the case where the seeing is seeing of the expressions of life of another individual – laughing, crying, expressing feelings, telling a story about how her father abandoned her mother and her and her mother was less difficult but still a tough case. What is aroused is not a patch of blue. What is aroused is a trace of anger, a sense of sadness, a reflection of the affects constellated by going through the experiences

presented in the story. This is not yet empathy, but it is a vicarious experience of what the other endured, suffered, or enjoyed. By making the trace of affect available for further attentional processing, including cognitive processing, the seeing (or listening) is a vicarious introspection of the other. This is Heinz Kohut's definition of empathy as vicarious introspection.<sup>34</sup>

### **Introspection as Vicarious Introspection - Empathy**

Once empathy and introspection have been linked together, it is hard to find an instance of one without implying the other. Thus, introspection implies that one is empathizing without oneself; and empathy implies introspecting the feelings and experiences of another vicariously. The former does not present any particular problem, but the latter may be easily misunderstood or distorted and deserves further clarification.

It is relatively easy to say what "vicarious introspection" is not. It is no kind of Russellian knowledge by acquaintance with another's sensation. Nor is it any kind of Schelerian intellectual intuition of the alter ego. No privileged access to another's fantasy life is entailed, and the situation remains an interhuman one in which expression of the emotional life of the other and receptivity to it mesh with one another.

What the locution "vicarious introspection" wants to call to our attention is how the emergence of an interpretation from empathic receptivity is mediated and facilitated by introspection. We have encountered the issue of the formulation of an interpretation based on one's empathy before. Now we can make further progress with it.

In vicarious introspection one is *not* introspecting the feelings, sensations, or experiences of the other at all. Such an operation would be to misplace a displaced perception, using Dretske's expression. Rather one is introspecting a vicarious feeling, sensation or experience aroused in oneself by the other's expression of feeling, etc. In introspection one comes to realize that this feeling is not an endogenous feeling, arising from purely

---

<sup>34</sup> Heinz Kohut. (1959). "Introspection, Empathy, and Psychoanalysis," *The Journal of the American Psychoanalytic Association* 7 (July 1959): 459-83.

inner processes in oneself, but is a vicarious feeling that is part of being receptive towards another's self-expression. Using the first person "I" for clarity, it is myself, not the other, who is the target of introspection. The other is the target of my empathy; I am the target of my introspection.

If, in defining "empathy" as "vicarious introspection" the emphasis is misplaced – misplaced, not displaced - from the vicarious to the "introspection," then a misunderstanding results. This makes it seem that empathy consists in introspecting another's inner life. This goes too far. It is not telepathy. Empathy consists in introspecting one's vicarious receptivity towards the other, and it is through the vicarious dimension of experience that one makes the connection between self and other.

It is worth citing an example to see how well this analysis accords with the current technical use of "vicarious introspection," Dr. Wolf discusses Freud's method of dealing with the many free associations with which he as analyst was inundated in listening hour after hour to his patients:

Did Freud arrive at his insights by an examination of the associations? I believe that he did not proceed in such a manner. Rather, Freud's guesses came from within himself, from an introspected awareness of his own mental states which Freud then connected with the mental states of his analysands.<sup>35</sup>

In the next sentence Wolf clarifies what he means by "examination."

Freud did not deliberately set out to draw conclusions from his experiences. Instead of that it seems that as he listened and immersed himself in the patient's presence, thought occurred to Freud which led him to insights about the patient (1976: 8).

That is, Freud's "guesses" were not based on deliberate conclusions drawn as to the possible cause of his state of mind, but rather on "introspective awareness." According to

---

<sup>35</sup> E. Wolf. (1976), "The role of empathy in the work of psychoanalysis" unpublished manuscripts: 8. Even here empathy does not arise from introspection without the addition of an aspect of interpretation. This is what provides the connection "with the . . . analysands." Without interpretation the "guesses" remain subjective. In being subjected to the interpretive rule, they are candidates for objectivity.

Wolf, the use of “introspection” is directed towards the self not the other. The perception is displaced, but from the other individual to the self, the opposite direction from Dretske’s epistemic seeing which is oriented towards the things in the environment. The vicarious experience is encoded as “mine” – that is, as an inner experience of the other’s affect, sensation, mood, etc.

When Wolf writes of “guesses” I think it is clear from the context that what is meant is an unvalidated interpretation, a hypothesis about the meaning of some symptom, etc. These spontaneously occurring “guesses” or “thoughts” “led” Freud “to insights about the patient.” I think it is correct to say that these thoughts and guesses occur as a kind of unreported, spontaneous process of free association in the analyst. Put in simpler words, as Freud was listening to his patient, he was also receptive and open to his own feelings.

### ***Vicarious Introspection and the Constitution of a Psychoanalytic Fact***

The importance of the question of the relation of empathy to introspection is considerable. Nothing less than the constitution of what is a “fact” is psychoanalytic psychology turns on the interrelation of these two terms.<sup>36</sup> A representative text in which all four—“empathy,” “introspection,” “vicarious introspection,” and “fact”—are connected occurs in Kohut’s “Introspection, Empathy, and Psychoanalysis” to which we have alluded more than once:

Let us consider a simple example. We see a person who is unusually tall. It is not to be disputed that this person’s unusual size is an important fact for our psychological assessment—without introspection and empathy, however, his size remains simply a physical attribute. Only when we think ourselves into his place, only when we, by vicarious introspection begin to feel his unusual size as if it were our own and thus revive inner experience in which we had been unusual or conspicuous, only then begins there for us an appreciation of the meaning that the unusual size may have for this person and only then have we observed a psychological fact (1959: 461).

---

<sup>36</sup> What is said in this section is primarily representative of the Chicago School of Psychoanalysis including Drs. H. Kohut, A. Goldberg, M. Basch, E. Wolf, and many others.



The transformation of a “physical attribute” into a “psychological fact” (of psychoanalytic psychology) occurs in three steps. First, with regard to the other person, “we think ourselves into his place.” Next the introspection of vicarious experience occurs based on an emerging, transitory, partial identification with the other individual. The experiences that are thus revived belong to the content of the introspection. Finally, an “appreciation of . . . meaning” of the person’s size accompanies the institution of a “psychological fact.”

Admittedly, this text is only an approximation to Kohut’s fully worked-out view, for it is prior to the introduction of the central notion of a transitional self-object. This latter is discussed in a vocabulary that is unashamedly peppered with the use of “cathexes” and is strongly energetic in *The Analysis of the Self* (1971). Furthermore, the earlier paper is incorporated into the texture of the later text on the self.<sup>37</sup>

I think it is fair to say that Kohut consistently maintains that introspection and empathy are necessary constituents of psychoanalytic facts (Kohut 1971: 37).<sup>38</sup> What he says is primarily written from the perspective of clinical theory of transference, though the discussion sometimes goes over into metapsychology as well. One of Kohut’s major contributions to psychoanalysis is the discovery of two hitherto undifferentiated forms of transference. The opening of the field of narcissistic transferences reveals two polar styles of analysis involving the activation of the grandiose self and the idealized, “omnipotent” object in a mirror and idealizing transference. Though it is an oversimplification, these correspond to a maternal, nurturing role and a fatherly, goal-oriented role. Note that we are no longer arguing about the constitution of empathy and introspection, but are instead using them to constitute something else (“facts”).

What it means to “think ourselves into the other’s place” is hopelessly ambiguous. Am I imaging myself in the other’s situation as I am or as he is? And if I already know how he is, then why is it necessary to imagine at all? I already know him. Suppose that I am of

---

<sup>37</sup> Compare Kohut 1959: 467 and Kohut (1971) *The Analysis of the Self*.

<sup>38</sup> H. Kohut. (1971). *The Analysis of the Self*. New York: International Universities Press, 1971.

short stature and hate it. Given Kohut's example, I would just love it being tall. I would be delighted no longer to talk to people while either staring at their middle chest or being constrained to look up at them if I want to make eye contact. However, that is not the question. The question is how does this person – this other person - feel about being tall. If I knew that, then I wouldn't have to imagine myself in his position. So how does the process get started?

Starting with the experience of meeting a tall individual that I have never encountered before and who walks into my class or office, I may fall back on something like an argument from analogy to bring what is experience-distant closer to what is familiar. But this is an argument that already supposes that stranger and I belong to the same community of fellow travelers capable of arriving at an understanding. Feeling the other's "unusual size as if it were one's own" entails an analogous kind of recollection of experiences in which I may have been unusual or conspicuous. However, the introspection that takes place is not of the other but of one's own experiences.

In this passage, the vicarious experience is driven by the imaginative thinking oneself into the other's place. This revives experiences that are encoded as inner in which one has been "unusual or conspicuous." In a coincidence of opposites, the quality of being conspicuous due to short stature would work almost as well as being too tall in capturing the quality of the experience of unusually sticking out in an uncomfortable way. The result is an "appreciation of meaning" of what is implied by being personally conspicuous and unusual. In simplest terms, the point is that a physical fact - 7 feet tall – is given meanings in terms of human interrelations – pretty girls who are 5 feet tall won't date him; can't find a decent suit off the rack; height is not all that is required to be good at basketball, still a klutz; can't fit in coach on an airplane and can't afford business class.

The grass is greener on the other side of the hill, and that is also the case here. When you cross over the hill, the grass is not greener after all. So even though I do not know full well, experience extends to a range of experiences that enable a bootstrap operation that ascends to an experience of the other. The tall individual knows all the disadvantages. He

constantly bumps his head going down the stairway into the basement. He has to travel on airplanes on business, and does not remotely fit into the seat in coach class. Long trips get really painful. It is hard to find clothes that fit without expensive alterations. The person is not particularly good at basket ball, and what good is being tall if you can't play some hoops? In short, it is misleading to put oneself in the place of the other without considering the particular character traits, personal preferences, and historical patterns of experience that go along with "being oneself." It is clear that Kohut appreciates this in that he starts collecting reminders about how being usually tall and conspicuous would be, having carefully built "unusual" into the description.

Still, one cannot help but get the impression that Kohut's interest in observation is as much a gesture in the direction of the authority of a certain paradigm of what is science (circa 1959) as opposed to a genuine conviction that what is constituted by empathy and introspection are "observables." The intention here is not to deny the analyst the use of his eyes and ears or even to suggest that he merely uses his senses in a different way than the experimenter (though the latter is true); but rather the idea is to disclose the function of interpretation in the constitution of psychoanalytic facts. We return to the hybrid character of empathy, and the way that aspects of two polar dimensions, receptivity and interpretation, are intertwined in this concept.

### **Without Empathy, the Inner Life of Man is Unthinkable**

The following consideration indicates that for Kohut a psychoanalytic fact is not principally an observable, but involves an aspect of the understanding of meaning. Only after vicarious introspection revives experiences in which we have been unusual, "only then begins there for us an appreciation of the meaning that the unusual size may have for this person." Kohut recognizes that an "appreciation of meaning" is involved in the constitution of a psychoanalytic fact. Empathy is not only tactical; it is strategic. It is not only empirical; it is constitutive of the inner life of the human being. Thus Kohut:

Empathy is not just a useful way by which we have access to the inner life of man – the idea itself of an inner life of man, and thus of a psychology of complex

mental states, is unthinkable without our ability to know via vicarious introspection – my explanation of empathy . . . what the inner life of man is, what we ourselves and what others think and feel (Kohut 1977: 306).

Coming toward the end of *The Restoration of the Self* (1977), this statement might mistakenly be taken as simply rhetorical or inspirational (though it may be these also). This statement, however, should be taken at face value. When it is taken in this way, it is astonishing. It is an assertion that the very idea of the mental life of the human being (man) - what we ourselves and others think and feel – is “unthinkable” without the ability to access (and know) others by means of empathy. Empathy is constitutive of the mental life of human beings, what we ourselves and what other think and feel.

An obvious first question is whether Kohut says both too much and too little. Does he say too much? What basis is there for asserting that empathy is the foundation for the complex psychology of mental states – thoughts (and beliefs?) as well as emotions, affects, and sensation? Without making it too easy to skate through on a technicality, it is useful to recall that most beliefs do not arouse our empathy because they relate to what are *prima facie* physical facts, not psychological ones. Empathy is not constitutive of the physical world, but of the psychological one, understood in the sense of what makes us human. Kohut’s interest would be in empathizing with beliefs and related mental states, presumably including human actions, in so far as they are vehicles for meaningful human relations. Even the simplest imaginable belief takes on a wealth of human meaning. The belief that it is raining *means* that the whole world seems to reflect my sadness; means that I can’t go out and play baseball with the guys from work – no fun allowed; means that I get to stay home and read – which I love to do; means that the wife and I get to stay home by the fire – fun allowed; means that on the day of Mozart’s funeral in the movie *Amadeus* it was raining too – the heavens were crying. The list is without limit. What it does not mean is a mere report about the weather. The affects and emotions are what make beliefs and actions matter to human beliefs and actions, so empathy brings them along too.

Kohut is explicit that the discussion extends to both oneself and the other – “what we ourselves and what others think and feel.” Empathy is the further “up stream” cognitive processing of the vicarious experiences disclosed in introspecting one’s openness to the other person. In relation to being open to oneself, empathizing with oneself reenacts the entire analysis of introspection undertaken earlier in this chapter in order to position and make sense out of introspection. Of course, the content is going to be the recollections, memories, patterns of experience and behavior in the past and present that provide access to character traits, autobiographical material, and one’s sense of identity. We can readily see why it is very useful to have another person - a committed and gracious listener – to mediate the accessing of one’s most intimate meditations. Thus, while it is in principle possible to empathize with oneself, all of the risks and hazards of confabulation and limitations of introspection loom large. The techniques of journaling, analysis of the counter-transference, and interpretations inspired by suspicion that things are not what they at first seem to be, are essential. Very few individuals have been able to get far with a self-analysis – Freud and Jung come to mind – and even then the results have inspired volumes on how incomplete the job was. Still, as an ideal point on the horizon towards which to work, empathizing with oneself remains a meaningful task.

Does he perhaps say too little? If the inner life of man is a “we” that is not a mere, rhetorical editorial “we,” but an actual first person plural form, then it is an easy next step to conclude that empathy is constitutive of the community of human beings, starting with the self and other. In this particular passage, Kohut is primarily arguing for the constitution of a scientific community – including psychoanalysts who define the realm of their data gathering activities through empathy. For example, the Kleinians are okay even if somewhat confused, while Alexander goes beyond the pale in rejecting empathic methods in favor of education and suggestion (Kohut 1977: 307). Of course, this is a caricature of both what Kohut says as well as his would-be Kleinian and Alexandrian colleagues; but the point is that empathy is used as the foundation for community, here understood as the “in group” of analysts who “get it” about the scientific use of empathy.

While this latter statement goes beyond what Kohut explicitly asserts – that empathy is the constitutive basis of community – it is consistent with his approach. It avoids the misinterpretations of empathy as compassion, on the one hand, and a sixth-sense perception (“telepathy”), on the other hand (Kohut 1977: 304). The essential tools of psychoanalytic transference and the interpretation of resistance themselves become meaningful within the field of observations constituted by the analyst’s empathic and disciplined (“scientific”) immersion in the experience of the other.

If there were any doubt about Kohut’s commitment to the establishment of a community of fellow travelers on the path of empathy – the unity of observer and observed - then it is Kohut’s account of the first psychoanalytic cure through the application of empathy.

The mutation that opened the door to the new field of introspective-empathic depth-psychology (psychoanalysis) took place in 1881, in a country house near Vienna, in the encounter between Josef Breuer and Anna O. (Breuer and Freud 1893). The step that opened the path to a whole new aspect of reality – a step that established simultaneously both the novel mode of observation and the novel content of a revolutionary science - was made by the patient who insisted that she wanted to go on “chimney-sweeping” (1893: 30). Yet it was Breuer’s joining her in this venture, her permission for her to go on with it, his ability to take her move seriously (i.e., to observe its results and to commit the observations to paper) that established that unity of observer and observed which forms the basis for an advance of the first magnitude in man’s exploration of the world (Kohut 1977: 301-2).

This documents the paradigm of the establishment of the first psychoanalytic community of self and other in Breuer’s deployment of empathy to cure Anna O. The famous “talking cure” was Breuer’s gift of empathy, which should not be underestimated given the stereotyped authoritarian approach to medicine characteristic of that time and place.<sup>39</sup> Obviously this is not the first use of empathy as every parent, teacher, and human being knows. Rather is the first disciplined, scientific use of empathy.

---

<sup>39</sup> In a video on the web site of the Chicago Institute for Psychoanalysis, Kohut refers to Breuer’s (and Freud’s) invention of the “talking cure” and explicitly calls out Breuer’s “gift” of empathy. See <http://www.chicagoanalysis.org/video.htm> (site checked on 2008/11/11).

Consider an analogy. In order to know some phenomenon I must be capable of being affected by it. Musical sounds are a constituent of my experience and understanding because of my capacity for hearing. Even if on some occasion my hearing breaks down or is destroyed, the concept of “musical sound” still makes sense. It is still thinkable, conceivable, so long as we admit the intelligibility of the capacity for hearing. We may debate the origin, quality, or meaning of some configuration of sounds. We may marshal other empirical tests, authorities, or perspectives to scrutinize some sounds. But we agree about the possibility of agreement, even if we never arrive at agreement in this particular case. However, in a universe without hearing as a general capacity, in a universe without organisms sensitive to - receptive to, capable of apprehending, able to be affected by – sounds, neither music nor even noise is thinkable. It would be a logical contradiction to assert “No sounds are audible” [because there is no such thing as hearing] and “Some sound is audible.” We are no longer capable of conceiving of sound, much less understanding or knowing them.

A parallel consideration applies to empathy. As Kohut’s remark suggests, without empathy, the very idea of the mental life of the human being unthinkable. Note how this is immediately qualified in the direction of a community of individuals by invoking “what we ourselves and what others think and feel.” This is because empathy, as a general competence in human interrelations, makes this inner life intelligible and meaningful by constituting it as a field of study in the first place. Thus, Kohut writes: “Empathy does indeed in essence define the field of our observations” (1977: 306). Here the phenomena (feelings, emotions, thought, and, presumably, meaningful behavior) are dependent on that function which makes possible our access to them. Empathy is that function on the basis of which the experiences studied by depth psychology are opened up and constituted as accessible and knowable. Because empathy is that without which the constitution of our psychological life does not make sense, it is the condition of possibility of that life.

Now let’s explicitly shift this consideration in the direction of the community of human relations. Empathy is the function through which human relations makes sense, insofar as

without empathy we would not even be able to conceive of human beings as capable of expressing and being receptive to the expressions of feelings. A communal field of experience in which feelings are expressed and receptively apprehended, but which is completely lacking in empathy would be unthinkable. Empathy is the organizing principle on the basis of which these experiences are made accessible.

Granted that we do have these experiences of expressing and being affected by others' feelings, of becoming aware that the feelings of another have an impact on our own, we may ask: How is this possible? We find that it is necessary to posit some capacity or competence – let us call it “empathy” - upon pain of contradiction if we refuse it. A world with expressed and receptively experience emotions, but without empathy, would be an absurdity in the strict sense. It would be a world of musicians without hearing – the frantic movement of bows across violin strings and fingers on ivory piano keys would be in vain for neither the musicians nor the listeners would in principle be capable of hearing the music. Similarly, without the capacity to empathize with the feelings of another, we would be just bodies located physically in space alongside one another – no interhuman connection would exist at all.

This line of reasoning, which Kant called a “transcendental argument” (1781/87: 592, 624;A737/B765, A788/B816), provides a principle that answers a question of the form: Granted that we have certain experiences, what must the constitution of our mental functions be like in order to account for the very possibility of such experiences? In this case, human experience in encountering other individuals indicates that we are affected by their feelings and that our feelings (in turn) affect them. What is being proposed is that there must be a functioning capacity for being receptive to the feelings of others in order for the recognition, identification, and further understanding of feelings in another to be “thinkable,” conceivable in any sense. Empathy, then, is just such a condition of possibility for describing others (and by implication ourselves) as capable of being affected by feelings.



Here the intimidating term “transcendental” requires some unpacking. It can be paraphrased as “not capable of being contradicted by experience, but nevertheless relating to experiencing and providing the framework or structure within which that experience becomes meaningful.” Empathy is what makes possible the experience of affecting and being affected by the feelings of another person. This experience, in turn, is the basis on which we are subsequently justified in positing the existence of the capacity for empathy. But in this experience empathy is itself presupposed, for without empathy the experience itself could not occur. Thus, the argument has the force of logic. It is concerned with what is “thinkable,” “conceivable,” without contradiction. Yet it is more than mere logic, for it concerns the realm of experience. Empathy is that on the ground of which being affected by the feelings of others is constituted as a realm of accessible experiences in the first place.

What Kohut has in fact given in the cited passage is an example of such a “transcendental argument,” although in an implicit and abbreviated form. In tracing a path from the experience of self and other in community to empathy, in which empathy is the presupposition of community, empathy becomes a component in a kind of metapsychological equation. Invoking empathy as an interhuman competence helps to make understandable the experiences we do in fact have. It makes intelligible how we are able to be receptive to the feelings of others. But, at the risk of paradox, how do we explain this principle of understanding?

Indeed that is the proper question since a show stopper objection occurs at this point. The skeptic might argue back that it is quite conceivable that we would all be emotional zombies – where is the logical contradiction in that? There is a wide spectrum of diseases of empathy – extending from autism to sociopathy – which, in greater or lesser degree, exemplify the loss of empathy without incurring anything illogical. It is true that the lives of such individuals are challenging and characterized by intermittent or frequent breakdowns in behavior, understanding, and communication. Yet the capacity for logical thinking is unimpaired, and, in some cases, is even enhanced, resulting in elaborations of consistent, systematic thinking.

The answer is direct. The statement “empathy is constitutive of the mental states that make us human and human members of a human community” does indeed lead to a contradiction if the subject term is withdrawn. “If there is no empathy, no mental states of the other individual are able to be experienced.” This statement directly contradicts the assertion that “Some mental states of the other individual are able to be experienced.” This leads back to the further contradiction in terms of humanness which is the realm of experience under-girding the connection between empathy and mental states. Empathy gives us our humanness of which mental states are a paradigm example. Withdraw the one, the other cannot stand. Of course, the skeptic is still able to question the premise, claiming, in effect, that the first transcendental argument requires another one to establish the conditional upon pain of not really attaining a principle that is both logically rigorous and applicable to experience.<sup>40</sup>

One partial answer to the skeptic is to backtrack on the universality and necessity of empathy. Empathy is indeed a form of receptivity to the expressions of the affective, volitional, and intellectual life of the human being; but it is entirely accidental and contingent that we are constituted that way. It is a primitive capacity not further analyzable in itself. Of course, one can analyze the neurological substrate, mirror neurons, though these are no less contingent than the empathic resonance which they ground. One can analyze the intentional structure of empathy, with its necessary correlation with the intending of the other in community, though, once again, the intentionality is no less contingent than the empathic understanding which it grounds. One can analyze the informational infrastructure of empathy, the communicability of affect and related propositional content. One can analyze the functional operation of empathy, with the mechanism of partial identification. None of these analyzes, while

---

<sup>40</sup> Michael Forster argues this is a risk and failing of transcendental arguments in general. I do not use the technical vocabulary of mathematical and metaphysical synthetic a priori principles in this paper for obvious reasons. I see no easy answer to Forster’s challenge to Kant’s transcendental arguments. If Kant has a backstop that halts the infinite regress, then it is Kant’s “Highest Principle of All Synthetic Judgment” (A158/B197), which delves into the transcendental understanding of the synthetic unity of the manifold of intuition as the “to which” all further questioning is referred. I acknowledge this is not a complete answer for Kant or anyone (such as Kohut) who wishes to follow him. See Michael N. Forster. (2008). *Kant and Skepticism*. Princeton, NJ: Princeton University Press, 2008: 64f.

enriching of our knowledge of empathy, are a backstop to the regress of the transcendental argument. All are ultimately contingent and a feature of the way we humans are thrown into the world. The contradictions that results from subtracting empathy from the equation are a function of human interrelatedness and limited to the sphere of human interrelations, not generalizable to all possible (logical) worlds.