

Who am I? What is it?

The subject-object relation

Sunny Y. Auyang

Mind is not some mysterious mind stuff; no such stuff exists and the universe comprises only physical matter. It is an emergent property of certain complex material entities, not brains alone but whole human beings living and coping in the physical and social world. This thesis involves three ideas: materialism, emergent properties, and intentionality. The first two belong to the mind-body problem and the status of mental properties in the material universe. The third refers to the mind-world relation, the symbiotic relation between subject and object in cognition and experience.

How is it possible that physical things, which are external to human subjects, become the intelligible objects of their experiences? The subject-object relation is peculiar because it includes both the possibility of errors and the possibility for the subject himself to doubt and separate illusion from reality.

Most theories regard the subject and object as entities that exist independently of each other, and intentionality as an extrinsic if not optional relation added on later. This paper offers an alternate model in which intentionality is an intrinsic relation in which the subject is aware of himself only when he can encounter objects in the world. Thus, intentionality not only relates but also defines the subject and object.

Borrowing ideas from group theory and its employment in physics, the model represents the mind-world relation in terms of two variables: object and mental frame. By accounting for the systematic transformations among mental frames, it explains how the variations in subjective perspectives lead not to relativism but to objectivity and a notion of reality based on human conditions. The full sense of *self* includes both a first-person and a third-person sense in which one recognizes oneself as a member of a community. Intersubjective understanding includes empathy and forming theories about others' mind, and both are built on the presupposition that we share an objective world.

Intentionality as the Mind-World Relation

You see clouds gathering. You believe it's going to rain and decide to take an umbrella. Seeing, believing, and acting are some of the most prevalent activities in our everyday life. They all involve some kind of relation between us and the world. This relation is complex. Your vision is qualitatively different from the camera's detection of light. Your taking the umbrella is

qualitatively different from the wind's wrenching it. Your relation to the world has an inalienable and essential mental aspect, by which you find the world meaningful and expect future events that may not actualize. The mental element in seeing, believing, choosing actions and other commerce with the world is *intentionality*.

Intentionality is one of the most basic characteristics of our mind without which we cannot have the kind of experience that we do. It explains what it is to see things that are physically outside us; what it is to believe that certain events happen. Everyone knows in his guts what it is to see or to believe, because they exercise these mental abilities every minute of their waking hours. Hence many people take them for granted and see no problem with them. When philosophers and scientists try to frame concepts and theories to spell out the characteristics of these mental abilities, however, they encounter great difficulties, for these abilities are highly sophisticated. Simplistic models tend to reduce intentionality to interactive behaviors. They are unsatisfactory because they make us mindless. To see implies to recognize, and we realize that recognition is fallible. Similarly, we not only believe that a certain event happens but are aware that our belief can be wrong, thus we can spontaneously doubt. To doubt and to be critical of one's beliefs are the corner stones of scientific research but are not limited to it. Cross cultural studies have found the notion of false belief to be shared by people all around the globe (Lillard, 1997). Thus intentionality is a relation between us and the world that includes not only the possibility of errors and illusions but also the awareness of the possibility. How to articulate this relation theoretically is a big problem in the scientific study of mind.

The characteristics of a relation depend on the characteristics of the entities that it connects. Let us call the entities related by intentionality the *subject* and *object of experiences*. Thus a model of intentionality should also give an account for the peculiarities of the subject and object. What makes things objects that are intelligible and meaningful to us? What makes a person a subject who both has a sense of self and makes sense of objects? What facilitates various subjects to understand each other? These questions about objectivity, subjectivity, and intersubjectivity underlie *the relation between mind and the world* that intentionality signifies.

Depending on whether I am a fully bodied person or merely a brain in a vat or a nonphysical soul, I would relate to the world in qualitatively different ways. How my mind relates to the world depends on what kind of entity I am and hence how my mind relates to my body. Thus an answer to the mind-world question presupposes an answer to the mind-body question. This constraint poses another difficulty for models of intentionality. For it demands an explanation of how, if mind belongs to a physical entity, can it extend beyond the entity's boundary to know the outside world.

I will briefly survey the prevailing model, which I call the closed mind controlled by mind designers (Fig. 1a). Then I will offer an alternate model, the open mind emerging from intricate infrastructures (Fig. 1b). The bulk of this paper is devoted to analyzing one part of my model, the structures of the open mind or mind-open-to-the-world.

Briefly, models of the closed mind posit an inner and an outer realm separated by a veil of Mental Representations. The subject or the closed mind, which is not necessarily embodied, is a preexisting entity that inhabits the inner realm independently of objects on the outside, for it has

access only to Mental Representations. My model of the open mind abolishes the inner/outer dichotomy, so that mind is in direct contact with the world. The subject is not a preexisting entity but arises simultaneously with the intelligibility of objects and intersubjective understanding.

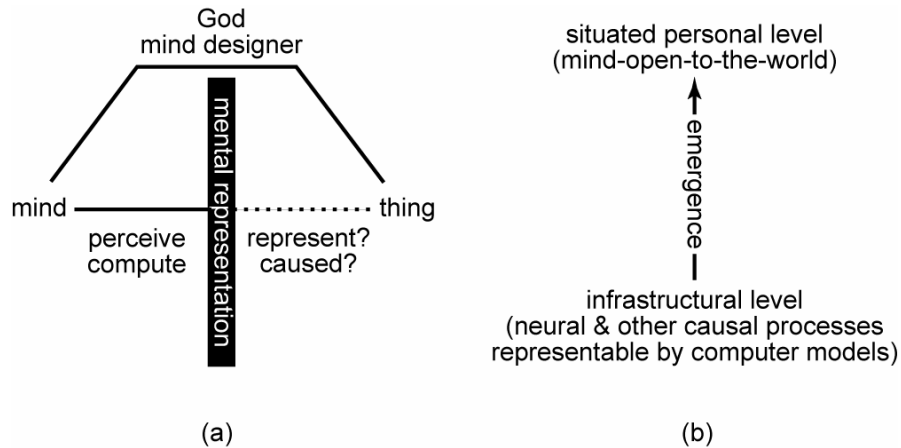


FIG. 1. (a) Models of *the closed mind controlled by a mind designer* posit an inner mental realm and an outer physical world, separated by a screen of Mental Representations. Mind, closed from the world, perceives or computes with Mental Representations, the meanings of which are known only to the mind designer. (b) Models of *the open mind emerging from intricate infrastructures* posit the open mind as a high-level property of a person engaged in the natural and social world. As a complex physical entity, a person has at least two organization levels connected by the relation of emergence. The engaged-personal level, on which we lead our conscious mental life, is analyzed into various mental faculties. The infrastructural level consists of many unconscious processes, which cognitive scientists represent by computer models.

The Closed Mind Controlled by Mind Designers

The closed mind controlled by mind designers is an old idea with many guises, and recently it is revitalized under the aegis of computationalism. Its source lies in the worry about errors. What if all my experiences are illusions created by an evil demon, René Descartes asked. In his radical doubt, he retreated to a pure thinking subject that is self evident in introspection. This subject exists by itself, independently of the existence and structures of the world, which need not be its concern at all.

Under Cartesian influence, John Locke split perception, the primary relation between mind and the world, into two parts by introducing a "veil of perception" that separates mind inside and things outside. The first part concerns the relation between mind and the intermediate entities that constitute the veil of perception, as Locke (1690, p. 563) wrote: "'T's evident, the Mind

knows not Things immediately, but only by the intervention of the *Ideas* it has of them." The second part tries to relate Ideas to things by claiming that certain simple Ideas are caused by things. But if mind sees only Ideas, how can it ever get any inkling that they are caused by things? To answer, Locke (1690, p. 564) appealed to the Deity: The Ideas are "the product of Things operating on the Mind in a natural way, and producing therein those Perceptions which by the Wisdom and Will of our Maker they are ordained and adapted to." This explanation is hardly satisfactory. Many who struggle with the problem find the two parts in the Lockian theory of perception to be incompatible. Once the veil of perception descends, it closes the mind absolutely. Mind cannot possibly breach it to know things and their causality. Pushing the empiricist rationale, George Berkeley arrived at the conclusion that Descartes dreaded: To be is to be perceived.

Here are all the essential ingredients of the computational representational theory of mind (CRTM), which Jerry Fodor (1981, p. 26) acknowledged to be "a Good Old Theory" that would please Descartes and Locke. Besides changing the names of the ingredients, CRTM generalizes the old theory from perception to all mental activities. What Locke called Ideas are now variously called sense stimuli, symbols, activation vectors, and generally, Mental Representations. While the old Ideas are mostly images, the modern Mental Representations can take any form, discrete or distributed. Whatever they are called and whatever form they take, they are all internal entities private to a subject. They constitute a veil – now called the veil of formality – that screens mind from the world. Like Ideas, Mental Representations create the inner/outer dichotomy and a two-part theory of mind. From the inside, the computational theory of mind employs a methodological solipsism to account for how mind operates on Mental Representations in its private chamber. From the outside, the Representational theory of mind tries to explain how the internal Mental Representations represent things in the physical world. It tries, but its effort seems to be no more successful than Locke's, so that how mind relates to the world remains an outstanding problem.

The paradigm example of CRTM is the computer. The computer operates on meaningless symbols, a form of Mental Representations. How do the symbols acquire meanings? Who interprets them? As Zenon Pylyshyn (1984, p. 63), an ardent supporter of CRTM, admitted: "the interpretation typically is effected by the user, not the machine." Thus computer users and programmers play the role of God in Locke's causal theory of perception.

Engineers do have a godlike status to the machines that they design and build. But even here, CRTM has a difficult time explaining why the machines that depend on external designers for the meanings of their Mental Representations have intentionality. Meteorologists who write and use a weather program knows how the program's symbols relate to rain or shine, but in what sense do the machines running the program know it? John Searle (1983) and Steve Horst (1999) are not alone in arguing that the machines at best have "as-if intentionality" or "derived intentionality." They criticize CRTM for confusing as-if intentionality of artifacts with the intentionality that people naturally exercise, hence to cover up important characteristics of mind.

The inner/outer dichotomy created by the new veil of formality proves to be as unbridgeable as the dichotomy created by the old veil of perception. Like a cut at the throat, the dichotomy kills the autonomous person who copes in the world and spontaneously finds it meaningful. In his

place are two parties: a closed mind playing with meaningless Mental Representations and a mind designer who controls the meanings of the Representations. Mind designers are as mysterious as God; how they manage to interpret Mental Representations is not explained by CRTM and other models of the closed mind controlled by mind designers.

Mental Representations and the inner/outer dichotomy they create are not limited to CRTM. They affect wide ranges of discourse on mental phenomena, underlying for example both internalism and externalism in the debate on meanings. Since I will reject them, let me pause to distinguish them from other notions.

"Representation" is one of the most confusing words in the literature of mind, because it has several meanings. I capitalize *Mental Representation* to emphasize that it is a technical term for entities that are radically different from ordinary representations such as words, pictures, signs, mathematical equations, flight simulators, computer generated virtual reality. We use words and symbols because we have the mental ability to understand their meanings. This ability is precisely what Mental Representations allegedly make possible. Explanations of Mental Representations that presuppose it are viciously circular, which is what many CRTM arguments turn out to be.

Words and other ordinary representations are physical objects in the publicly accessible world. A user knows explicitly that representations represent other objects that are also accessible to him, thus he can separate virtual reality from the real world. In contrast, Mental Representations are entities in the head that are not necessarily physical. They are private to a subject and exhaust what the subject is exposed to. The subject neither knows that Mental Representations represent anything nor has access to anything else, including what outsiders may say the Representations represent. The representative nature of Mental Representations is known only to external mind designers whose ability remains unexplained. In short, ordinary representations enable people to cope in the world, whereas Mental Representations close the mind from the world.

The distinction between ordinary representations and Mental Representations clarifies a potential confusion. Scientists in all fields use theoretical concepts and computer models to represent various phenomena, but these representations usually do not imply that the phenomena contain peculiar entities identifiable as Representations. We use the word "hexagon" to represent the structure of benzene, but benzene itself does not contain any Representation, it is hexagonal, that is all. Similarly, cognitive scientists use computer models to represent many processes. To claim that the represented processes contain Mental Representations, however, is a totally different matter, the justification of which is dubious.

The Open Mind Emerging from Intricate Infrastructures

I see trees planted firmly in the ground, not images floating somewhere in my head. I believe that waters fall from the sky, not that the proposition "it's raining" appears in my head. In seeing and believing, I engage in relations with things and events in the public world, not with my private Mental Representations. I need no mind designers or other Big Brothers to interpret my

mind and tell me what I mean when I see a tree or believe that it rains. I reject models of the closed mind controlled by mind designers.

Models of the closed mind are popular in the interpretations of cognitive science not only because of the concern of artificial intelligence (AI) with designing artifacts. Most processes studied in cognitive science, from priming in cognitive psychology to brain activation in neuroscience, are unconscious and hence are not directly meaningful to the subject to whom they belong. You are aware of seeing a tree but are not aware of the processes going on in your brain that makes your vision possible. Neuroscientists, however, pour tremendous effort into experimentally probing these underlying processes and theoretically representing them with computer models. Computer models of neural and cognitive processes are similar to computer models of the weather. They are theoretical tools that scientists use to represent some phenomena or another. However, they are often erroneously identified with the phenomena represented, giving rise to the idea of a computational mind operating on Mental Representations.

To account for the results of cognitive science as well as our ordinary experiences, I propose a model of an open mind emerging from intricate infrastructures. The model explicitly posits two organizational levels of a person, *the mental or engaged-personal level* and the *infrastructural level*, which are connected by the relation of *emergence* (Fig. 1b). The engaged-personal level is the arena of our conscious experiences and mental life. It is analyzable into various mental faculties such as vision, audition, memory, and language using. The unconscious processes investigated in cognitive science belong not to the level of mind but to the level of its infrastructure. They constitute not a peculiar closed mind but what I call the infrastructure of the open mind, which is analyzable into various infrastructures underlying various mental faculties. Noam Chomsky's "language organ," for example, is the syntactic infrastructure underlying the syntax of our speech, and Universal Grammar is the linguist's theoretical representation of the syntactic infrastructure. To study the underlying processes of prominent phenomena is a staple of natural science, and here cognitive science falls in league with other sciences.

Multiple organizational and descriptive levels are familiar in cognitive science, but their significance and consequences still need to be stressed. Because different levels usually have radically different properties, concepts applied to different levels should be clearly distinguished. Many infrastructural processes are causally connected to external lights and acoustics, and scientists often say that they "represent" the lights and sounds. For instance, they say that edges are "represented" in the brain by the differential activation of certain neurons. This "representation" is a technical concept for the infrastructural level and should not be confused with the Mental Representations that occur on the mental level. The infrastructural "representations" imply no conscious cognizance, Mental Representations do.

To explain how our mental experiences and the unity of consciousness emerge from the self-organization of myriad infrastructural processes is called the *binding problem*. Substantive solutions to the problem await future science. Elsewhere I explain it by adopting a general notion of *emergence of high-level properties*, which I delineated from existing scientific theories about other complex phenomena (Auyang, 1998, ch. 6; 2000, ch. 3).

There is much advantage in examining concrete examples instead of arguing in the vacuum about the meanings of emergence. By looking at how theoretical science treats phenomena such as the emergence of fluid turbulence from the self organization of molecules, we can solve a puzzle that haunts the explanations of intentionality. All neural and infrastructural processes occur underneath the skin of a person. Does that imply that our mental structures too can only involve factors underneath the skin? An affirmative answer is tacitly or explicitly assumed in many models of intentionality. Searle (1983), for example, argues that it must be possible for the brain in a vat to have intentionality. However, if intentionality is confined to factors underneath the skin, its significance of relating to the world evaporates, and we relapse to a model of the closed mind. Fortunately, this need not be the case, as existing theories of high-level emergent phenomena assure us.

Contrary to the stipulation of reductionism, scientific theories usually employ different scopes of generalization for different organizational levels. For example, mechanics does not generalize over initial conditions; statistical mechanics does, and hence is able to explain the irreversibility of processes that is forbidden by the laws of mechanics. It is not unusual for a high-level property to include factors that are spatially outside the confine of lower-level properties. Furthermore, there are properties that have absorbed factors of an environment into the characterization of an individual entity situated in that environment. These *situated properties* of individuals are custom made for the *situation*, which in turn is characterized self-consistently together with the situated properties. I refer the readers to the many examples that I gave elsewhere (Auyang, 1998, § 14; 2000, § 13). Here I only want to observe that although the notion of situated properties-consistent-with-the-situation is not familiar in the philosophy of mind, there is nothing mysterious about it. On this ground I will argue that intentionality or the open mind is not and cannot be the property of a brain in a vat or a person in absolute isolation. It is a situated property consistent with the intelligible world, belonging only to a person radically engaged in the natural and social world, manipulating things and communication with other people. Therefore I call the level where mental experiences occur not the merely personal but the *engaged-personal* level. The remaining of this paper is devoted to analyzing the structures of the engaged-personal level, which I call *mind-open-to-the-world* or briefly *the open mind*.

Mind-open-to-the-world

Before the rise of computer worship, the major motivation for introducing Mental Representations is to explain illusions. Theoretical models of mind that hope to account for the possibility of errors in our judgments need at least two variables, an objective variable germane to what is the case, a subjective variable germane to what we think to be the case. Mental Representation was originally introduced as the subjective variable; errors occur when a Mental Representation does not match with the thing it purportedly represents. Unfortunately, this move crowds out the object variable, because the subject has no way to compare Mental Representation to things, to which he has no access. Mental Representations close mind to what is the case and rob it of the ability to know or even to doubt.

What is the difference between a house and your visual experience of it? In describing your

experience, you do not refer to any mental imagery. You describe features of the house itself, with the restriction that they belong only to one side of the house, because this is the side accessible to your viewpoint. Your particular perspective, not some mysterious Mental Representation, is your subjective contribution to your experience. Perspective is a better variable in models of intentionality than Mental Representation, provided it is suitably generalized.

We see, touch, think, wish, remember, often about the same thing or event, as I hold, smell, and taste a piece of cheese. We have many possible *psychophysical modes*, ranging from perception, proprioception, motion, and emotion to language and inference. Each psychophysical mode consists a spectrum of *perspectives*, which may be spatial, temporal, intellectual, cultural, or linguistic. Vision and audition involve spatial perspectives; recollection and anticipation involve temporal perspectives; description and explanation involve intellectual perspectives; speech production and comprehension involve linguistic perspectives; convention and custom involve cultural perspectives. Intellectual perspectives change when we learn more about a phenomenon, just as visual perspectives change when we draw closer to an object. I combine psychophysical mode and perceptive in the *frame of mind*. A person is in a specific mental frame when he adopts a specific perspective in a psychophysical mode.

I posit frame of mind as a *variable* to replace Mental Representation. The difference between the two marks the difference between the closed and the open mind. Mental Representations are introvert and oblivious of embodiment; one looks into oneself, which even a nonphysical soul can achieve. Mental frames are extrovert and necessarily embodied; their *psychophysical* modes tie them to our bodies and their perspectives put us squarely in the world. Thus mental frames belong not to abstract souls but to engaged persons.

How does the mental frame variable help to account for the possibility of doubt, error, and knowledge? Consider William Shakespeare's *Macbeth* (II, i, 34-40):

Is this a dagger which I see before me,
The handle toward my hand? Come, let me clutch thee!
I have thee not, and yet I see thee still.
Art thou not, fatal vision, sensible
To feeling as to sight? or art thou
But a dagger of the mind, a false creation
Proceeding from the heat-oppressèd brain?

One can hardly hope for a better candidate for Mental Representations than Macbeth's dagger, yet Macbeth did not regard it as such. His immediate experience was seeing an object in the world that incited his suspicion. He doubted spontaneously because he knew from past experiences in normal frames of mind that daggers do not hang in thin air. To resolve his suspicion he resorts to another psychophysical mode and hence another frame of mind, not unlike someone moving to a new position to gain a better look of a thing. When the visual and tactile frames of mind yielded conflicting experiences, he thought of various hypotheses. Finally he concluded "There's no such thing." Thus Macbeth decided, without the help from any mind designer, that he was hallucinating.

Relativism and its sibling philosophies have turned variable perspectives into perspectivism, where mental frames degenerate into subjective caprice and viewpoints into the antithesis of objective knowledge. However, Macbeth's commonsense reasoning reveals the error of perspectivism. Perspectives are not *merely* subjective; by adopting and relating various mental frames, Macbeth was able to make an *objective* judgment about his vision. Relativism errs because it has neglected the other variable necessary in the description of mind: what is the case. In discarding the objective variable, it also kills the relations among mental frames that Macbeth grasped. Hence it ends up with a set of incommensurate frames.

My model of mind-open-to-the-world employs two variables, *object* and *mental frame*. The two are interrelated, so that we have a particular experience of an object from a particular mental frame. As a variable, mental frame has many values that account for many possible perspectives. Contra perspectivism, however, I insist that the mental frames are not incommensurate but can be related by *transformations*, especially those transformations that leave some values of the object variable *invariant*. Thus by comparing experiences in various mental frames, we can subtract away the varying subjective contributions and extract the objective features as *that which are invariant under the transformations of mental frames*. When we are unable to identify any invariant features, we doubt ourselves and admit errors. This is the logic that Macbeth employed; it is also common sense.

When a boy tries to figure out a toy, he does not merely stare at it but turns it in his hands, thus probing it from all possible perspectives that he can manage. From different intellectual perspectives, we describe an event in different ways. "The sun rises everyday" and "the earth rotates about its axis every 24 hours" are two representations of the same process. The transformation among perspectives is as important in scientific research as in everyday life. Thus Richard Feynman (1965, p. 168) observed that "every theoretical physicist who is any good knows six or seven different theoretical representations for exactly the same physics." The stress on the *same* physics differentiates the physicists' objectivity from relativism. Various representations accentuate various aspects that are illuminating under various contexts. Knowing how they transform into one another is crucial for understanding complex phenomena and investigating the unknown.

If we compare objective knowledge to distance measurement, then the transformation among perspectives is like triangulation. In triangulation, we determine the distance of an object by looking at it from two viewpoints. It works only if we also know the distance between the viewpoints, and its accuracy increases as we move the viewpoints further apart. Similarly, in the quest for objective knowledge, we look at a phenomenon from two or more perspectives. It works only if we can relate the perspectives, and its accuracy improves as we secure invariance from a bigger and more diverse group of perspectives.

A transformation Schema as a Model for Intentionality

If openness to the world is an essential characteristic of the human mind that makes objective knowledge possible, then it should also reveal itself in scientific theories, because the theories

embody our most refined knowledge about nature. By comparing the structures of common sense and quantum and relativistic theories, I have delineated a shared general conceptual structure (Auyang 1995, ch. 5; 2000, ch. 8). Here I borrow a schema lifted from the group theoretic structure of physical theories to explain how the two variables of object and mental frame account for the objectivity, subjectivity, and intersubjectivity of people with open mind.

The schema, which I call the *transformation schema*, is depicted in Fig. 2. It contains two variables, object x and mental frame f . Both variables are necessary for objective experiences; an experience $p_i = f_i(x)$ is analyzable into two aspects: an object x conceived in a mental frame f_i . Various experiences are related by the frame transformation $f_2 \cdot f_1^{-1}$.

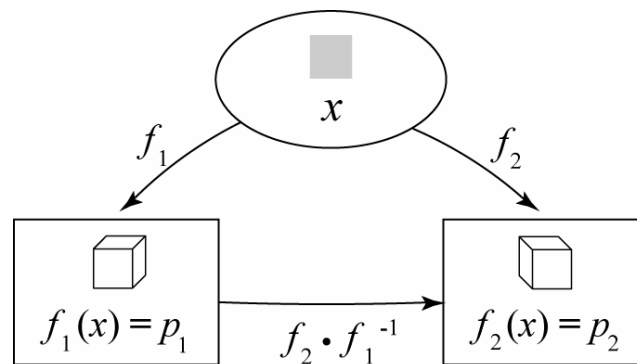


FIG. 2. The *transformation schema* represents the structure of *mind-open-to-the-world* with two variables: object x and mental frame f . In it, an experience $p_1 = f_1(x)$ is analyzed into a subjective mental frame f_1 and the object x the experience is about. The objectivity of the experience p_1 is secured by the fact that its object x can also be experienced in other mental frames such as $p_2 = f_2(x)$ and is invariant under transformations $f_2 \cdot f_1^{-1}$ across all relevant mental frames. The invariance under systematic transformations implies that the object is independent of subjective mental experiences it.

If x is a house and f_1 is your mental frame of vision from the front, then $p_1 = f_1(x)$ is your experience of the house as seen from the front, which appears like a facade. With the experience p_1 alone you are unable to decide whether x is a house or merely a stage pop. You can find out by going around the corner, from where you have another experience $p_2 = f_2(x)$ of the same x from another mental frame f_2 . By reckoning with your own spatial locations relative to the house, you are able to transform your mental frame, as represented by $f_2 \cdot f_1^{-1}$. It is important to note that the transformation is not a simple mapping between two internal Mental Representations, which even a closed mind can manage. Rather, it is a composite mapping that refers to the object x : $f_2 \cdot f_1^{-1}(p_1) = f_2(x) = p_2$. This composite transformation indicates the intrinsic openness of mind to the world.

By systematically transforming your mental frames, f_1, f_2, f_3 , and so on and comparing your experiences p_1, p_2, p_3 , and so on, you abstract the features of the object x . You ascertain that it is

a house with front and back, inside and outside, roof and cellar. Conversely, because you can extract coherent and invariant features from the transformations, you judge your experiences objective. Otherwise, you would join Macbeth's company and decide that you are hallucinating.

The transformation schema occurs in physical theories also. In special relativity, for instance, the f s represent coordinate systems. Connecting two coordinate systems is a Lorentz transformation $f_2:f_1^{-1}$. The objective spatial-temporal structure is the x that is invariant under all Lorentz transformations, and $f_1(x)$ is the spatial-temporal structure measured in a particular coordinate system. As descriptive schemes associated with various motions of an observer, coordinate systems are mere theoretical entities. Let us call them "subjective elements" in the theory of relativity. The spatial-temporal structure x , represented in the coordinate-free form, is independent of any and all subjective elements connectable by transformations in the Lorentz group. It is intelligible to us only through the elaborate theoretical apparatus of descriptions and transformations.

As a model for our mental characteristics, the transformation schema differs in two essential ways from its employment in physical theories. First, physics takes inanimate systems as objects of investigation. Thus it deems only the invariant x to be objective and interprets f as the theoretical tool for physicists. The science of mind takes physicists themselves as objects of investigation. Thus it deems the whole schema, including *mental frames-transformations-invariance in its entirety*, to be objective. The whole schema describes objectively the mental states of people who are capable of experiencing and knowing the objective world.

In physics, the transformation schema is a part of mathematical theories that are substantiated to details. Adapted to the science of mind, whose complex objects defy representations of comparable refinement, the transformation schema functions not as a mathematical model but merely as a crude schema. The schema posits only the *general concepts* of object and mental frame but does not demand that all the *specific values* of the general concepts are available. It is like a tax form with many systematically connected blanks. There is no requirement that a taxpayer fills in all the blanks, but it is important for the taxpayer to know that the blanks for certain deductions are available should he needs them. Similarly, the transformation schema provides many blanks for mental frames, objects, transformations. These blanks are partially filled in most mental states, but there is no requirement that they be all filled in all cases. The important point is that by grasping the general concepts of object and mental frame, a person is aware of the *possibility* of blanks to be filled. Thus when the substances for some blanks are missing, he can spontaneously doubt or admit ignorance. And when some blanks are filled, he is informed and gains knowledge.

Suppose $f_1(x)$ represents a person's seeing a bright star in the morning and $f_2(x)$ his seeing a bright star in the evening. However, the viewer does not know that the morning star and the evening star are the same planet Venus because he does not know that x is invariant in the two mental frames. He will gain knowledge about nature when he learns the identity represented by the transformation $f_2:f_1^{-1}$. Or consider Macbeth's case. He had an experience p_1 whose object x failed to be substantiated as the object in other mental frames. Consequently he judged that the first content occurred in an illusory rather than the visual frame of mind.

Concepts are susceptible to generalization. Once we grasp the general concepts of object, mental frame, transformation, and invariance, we can generalize them to include more frames and more transformations. With language and other forms of communication, we take account of other people's perspectives; with tools and equipments, we extend our observational range to things invisible to the naked eye; with concepts and theories, we multiply our intellectual perspectives. As we secure a broader and broader basis of judgment in scientific inquiry, the probability of errors in our objective knowledge drops precipitously. In this sense we can say that our knowledge converges asymptotically on truth. Importantly, this convergence does *not* posit any transcendent view of Naked Reality. It has presupposed the human condition that we can only experience objects from particular mental frames. Thus it does not fancy that we can have the view from nowhere. Instead, it demands us to take account of the views from everywhere that we can manage.

The Existential Self and the Subject of Intentionality

Intentionality means that *my* mental state is directed at an object or that *I* am thinking about an object. The subject is an individual, but not merely an individual to whom we can refer as *this* or *that*. It is an individual who can say *I* and adopt a first-person perspective. What structures must a subject have to have a sense of self as we do?

Models of the closed mind controlled by mind designers have very little to say about the characteristics of the subject, expect that it is an entity transparent to introspection and exists independently of the world or even of any embodiment. Notions of the self have attracted much interest recently; the special issues on it in *Journal of Consciousness Studies* have run over 400 pages. Their key note paper is the inner self of the closed mind, which has incited much criticism and many alternative suggestions. I will offer a notion of *existential self* based on the transformation schema. Consistent with the open mind, one finds himself not in introspection but in worldly engagements. The self *exists* in the Greek sense of "standing out" by differentiating himself from and relating himself to things and other people.

The closed mind knows itself first, and then runs into the intractable problem about other minds. I take the opposite approach in which a person with open mind knows himself as he knows others. How do we represent the cognizance of other people in our model of intentionality? Our model has two variables, object x and mental frame f . Consequently, it accommodates two views of other people, which cognitive scientists call "the theory-theory" and the "simulation theory" of mind (Carruthers and Smith, 1996).

The "simulation theory" resembles what people ordinarily call *empathy*, in which a person projects his own thinking, feelings, and situations to other people. This is represented in our model by extending the range of the frame variable f to cover the psychophysical modes and perspectives of other people. Suppose you and I look at a picture x . Then in the theoretical representation of my mental state, $f_1(x) = p_1$ is my visual experience, $f_2(x) = p_2$ is what I attribute to you as your experience, and $f_2 \circ f_1^{-1}$ is my notion of intersubjective understanding. Notice that this intersubjective communication is not a mere two-way relation in which I broadcast my mind to induce an experience in you. It is a three-way relation among you, me, and an object, for $f_2 \circ f_1^{-1}$

$f_1(p_1) = f_2(x) = p_2$ refers to the picture x , so that our minds meet at the picture that we together see. We can understand each other because we share a publicly accessible world. Joint attention to objects develops very early. Infants three months old can follow the gaze of an adult to look at the object the adult looks at (Hood *et al*, 1998).

The "theory-theory" asserts that people have a theory of mind, by which they attribute various mental properties to other people. Psychological concepts such as believe, desire, know, and feel are elements of the theory of mind, which is often called folk psychology or commonsense psychology. This is represented in our model by treating other people as values of the object variable x . Objects are what we postulate theories about in our intellectual frames of mind, and we treat people as objects with mental properties. Theory appears to be more difficult than empathy. Not until a child is four or five years old does he realize that other people may be ignorant of something that he knows (Wellman, 1990).

In short, an object x in the world and a sharer of some mental frames f_i together constitute the general concept of *someone*, as distinct from *something* that is merely an object. An instance of *someone* is an object to whom a person can extend the values of the frame variable; an instance of *something* is an object to which he cannot.

With perception, proprioception, and other psychophysical modes, a person realizes that he is a participant in the world. With the concept of *someone*, he realizes there are special participants with mental frames similar to his. Now he comes to realize that he is one of those special participants, one whose mental frames are precisely his own. Thus he arrives at the notion of himself as *One*, who is simultaneously oneself, a particular instance of *someone* to whom mental properties can be attributed, and a flesh-and-blood object that is a part of the world.

With the notion of *One*, a person sees himself from a partially detached position, and hence gains a notion of *third-person self*. The third-person self is expressed in many occasions, including many narratives that is usually called "first person." For instance, memory researchers have distinguished *field* and *observer memories*, the former recollects from the first person and the latter from the third person viewpoint. A field memory of childhood sees the world with a child's mentality; an observer memory sees oneself as a child (Robinson, 1996). In observer memory, a person still refers to the child as "I," and this "I" is a third-person self.

As a part of the world, *One* is an object about which a person thinks. The person is also the thinking subject who says *I*. I experience this. I think about that. *I think* or *I am aware* is the most primary hallmark of mentality, without which one cannot know anything, let alone forming the reflexive idea of *One*. Some people take this primitive *I* to be an inner entity, the pure thinker or the Cartesian Ego that is accessible to introspection. This inner self has been criticized to be illusory by a long list of philosophers and scientists. David Hume went in to look for it and came out empty handed. Immanuel Kant followed Hume and argued that "no fixed and abiding self can present itself in the flux of inner appearances." Those who think there is such a self simply confuse the unity of consciousness with the consciousness of a unity (Kant, 1781). These insights about the *nonexistence* of an inner self are captured in our model by the fact that *I* or subjectivity is *not* represented by any definite term in the schema Fig. 2. A definite value of x represents *One*, the third-person self, but not the *I* who think about the third-person self.

Instead of a definite term, the *I* or *first-person self* is represented by the *whole transformation schema*, the whole Fig. 2. For the whole schema represents the minimal mental abilities without which an entity cannot have experiences, hence cannot possibly have any sense of *I*. The schema presents a unitary framework with many interrelated elements. The emphasis on the schema as a whole highlights the unity of consciousness wherein I am aware of various mental frames and experiences as *mine*. Thus whenever I perceive an object from a particular mental frame, whenever I conceive a state of affairs from another frame, whenever I transform from one mental frame to another, I exercise my mental abilities and am thereby the thinking subject. Besides that, there is no extra entity, no mental self, that is susceptible to introspection or other kind of perception.

Together, the first-person *I* and the third-person *One* constitute the *existential self*, our full sense of subject and our consciousness as personal identity. The existential self is not an entity that exists before encountering objects in the world, for the possibility of objective experiences, as represented by the transformation schema, is the very minimal structure of *I*. Thus I am aware of myself only as I make intelligible the world and find my place in it. This is similar to Kant's highest principles of all objective knowledge. Kant (1781, A158/B197) wrote: "The conditions of the *possibility of experience* in general are likewise conditions of the *possibility of the objects of experience*." One commentator explained: "There can be no *it* unless there is an *I*," another added: "Not only does the *it* entail an *I*, but conversely the *I* entails a *it*." Zhuang Zhou expressed it best more than two thousand years earlier in nine words: "Without *it* [there is] no *I*, without *I* nothing is intelligible."

References

- Auyang, S. Y. (1995). *How Is Quantum Field Theory Possible?* New York: Oxford University Press.
- Auyang, S. Y. (1998). *Foundations of Complex-system Theories: in Economics, Evolutionary Biology, Statistical Physics*. New York: Cambridge University Press.
- Auyang, S. Y. (2000). *Emergence of the Open Mind: Mind in Everyday Life and Cognitive Science*. Cambridge, MA: MIT Press.
- Carruthers, P. and Smith, P. K., eds. (1996). *Theories of Theories of Mind*. New York: Cambridge University Press.
- Feynman, R. (1965). *The Character of Physical Law*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1981). *Representation*. Cambridge, MA: MIT Press.
- Hood, B. M., Willen, D., and Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, **9**, 131-4.
- Horst, S. (1999). Symbol and Computation: A Critique of the Computational Theory of Mind. *Mind and Machine*, **9**, 347-81.
- Kant, I. (1781). *Critique of Pure Reason*. New York: St. Martin's Press, (1965).
- Lillard, A. S. (1997). Other Folk's Theories of Mind and Behavior. *Psychological Science*, **8**, 268-74.
- Locke, J. (1690). *An Essay Concerning Human Understanding*. New York: Oxford University Press (1975).

- Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Robinson, J. A. (1996). Perspective, Meaning, and Remembering. In *Remembering our Past*, ed. by D. C. Rubin, New York: Cambridge University Press (1996), pp. 199-217.
- Searle, J. R. (1983). *Intentionality*. New York: Cambridge University Press.
- Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Talk presented in Boston Colloquium for Philosophy of Science
Boston University
November 2, 1999