Quantum-like Visual Recognition in Mental Degree 2.3 (the external everyday mind)

A sketch of a proposal, v2
Ian Thompson. June 26, 2020
Developed from "Perception in the Mental Degree 2.3"

Introduction

Here we describe the operation principles of vision in a mental degree 2.3 in ways that correspond as much as possible to those of the quantum-physics degree 3.3. The correspondences will result in both similarities and differences which need to be explained.

2.3: External Mind	2.3.1 Concepts of objects and actions	2.3.2 Conceiving of possible objects	2.3.3 Sensorimotor mind: specific seeing and acting
3.3: Quantum Physics	3.3.1 Hamiltonian: kinetic + potential energies	3.3.2 Quantum wave function	3.3.3 Actual selections Measurements etc.

The basic idea here is that there are *wavefunctions in the mind*. These however are not *physical* wavefunctions giving probabilities for the different outcomes of quantum propensities, but they are *epistemic* wavefunctions giving degrees of belief of specific visual contents. They thus give knowledge probabilities in the same manner as Bayesian inference methods. We will see some differences from Bayesian methods once we allow some non-commuting operators (a well-known feature of 'Generalized Quantum Theory' (Basieva, 2016)).

This wavefunction or field in the mind is thus not in the brain or a property of neural events. It is in fact a form of some mental propensity such as the love or desire for seeing. Forms of such a love or desire form their own discrete degree that is distinct from the brain, but the forms are still closely coupled to brain activity by the principles of influx and selection. The mind produces influx into the brain that sustains and creates neural activities there, and the specific resulting neural events (with further inputs from the body and its senses) selects further mental propensities from the initially available alternatives. The fields in the mind are not in physical space but are in their own space for possible actions.

I am going assume a specific mechanism for how the senses create the first sensory field by means of inputs via the visual cortex in the brain. It is described in Appendix A. My main focus now is how, given this 'sensory sheet', animals (from insects to humans) are able to recognize objects in 3D space and their actions and possible actions (affordances).

It is the process of *recognizing* that we need to understand. The ability to recognize an object still functions if the object is moved around, or rotated, some of many possible transformations such as

> *Geometric*: translate, rotate, dilate, accelerate, spin, (but not up-down reflections)
> Illumination: light intensity, light color, shadowing, occlusions.
> *Self-generated*: eye saccades, moving head, blinking, rotating body, walking, etc.

Computer vision programs have managed to produce some of these functions, and that might suggest that neural systems in the brain might also succeed. However, the simple invariance of object recognition with respect to rotations, inflations and translations requires convolutional integrals that are not available to neural networks without much more interconnection than is in the brain. The *speed* of animal object recognition is also a puzzle, given the millisecond time scales of neural propagations (compared to nanoseconds for computers). This article proposes to see whether an explicitly *mental* scheme could accomplish object recognition on the needed time scale and linked with a plausible complexity of the neural systems in the brain. To guide our theorizing, we look for a scheme using principles corresponding to quantum mechanics with wave-functions, operators (continuous, and projections), superpositions, and probabilities. Such a quantum dynamics has been called Generalized Quantum Theory by Atmanspacher (2006), Filk (2011) and others, though I am treating it not just as a model, but realistically as existing in some mental space.

Mathematical notation and interpretation

The basic idea of using quantum mechanics as a basis for describing seeing and beliefs in a mind, is to use a 'wave function' or 'state' Ψ to describe the degree of some belief. More specifically, wave functions are always a function of some coordinates as $\Psi(x_1, x_2, \dots)$, such that the degree of believing that the configuration $(x_1, x_2, \dots)$ is true or is existing, is $|\Psi(x_1, x_2, \dots)|^2$. That is the square modulus of the wave function, which allows that values of Ψ itself may be real positive, real positive or negative, or complex-valued. In quantum physics $|\Psi|^2$ gives a probability of some events, but here gives a degree of belief. The number and nature of the coordinates $(x_1, x_2, \dots x_n)$ need not be fixed, so they will depend on the complexity of whatever the belief is about. The total degree of belief may be an integral over all the coordinates of the degree of belief $N = \int |\Psi|^2 d^n x$. If the space S of coordinates is mutually exclusive and complete, then holding a belief about the true configuration should satisfy the normalization condition

$$\int_S |\Psi|^2 d^n x = 1,$$

for $N = 1$ corresponds to certainty. An integral of $N = 0$ corresponds to ignorance, from $\Psi = 0$ everywhere: a lack of belief altogether about what is true in the space S. Another kind of belief, slightly above ignorance, is that all places in S are equally likely. This would correspond to small but constant Ψ values everywhere in S, such that the integral was still unity rather than 0.

Sketch of a Method

We begin by taking the wave function ψ to describe the epistemic state of a mind at any specific time. If there is no previous knowledge of its visual surroundings, then it starts, when the eyes are opened, to have the value $\psi = \psi_s$, where ψ_s is the knowledge contained in the two-degree 'sensory sheet' that is produced by mental interaction with the visual cortex as described in Appendix A. The beliefs conveyed by this sensory sheet have coordinates $\psi_s(c, \theta, \chi)$, where c is the color index R,G or B, the angle θ is the horizontal pan angle ($\pm 105°$ for humans), and χ is the vertical angle ($\pm 75°$). In this method we assume this ψ_s is given and try to determine how 3D objects are recognized using that information.

Our task, therefore, is to find a fast and plausible method for finding patterns in the sensory sheet that can be recognized as 'objects' such 2D shapes, 3D objects, or as 4D (3D+time) processes and affordances. There are an enormously large number of such objects, and for each object there are a great many various translation, dilation and rotation (etc) operations by means of which every recognizable object may appear in the visual field. Some of those operations are necessary just for the eye, head and bodily movements that humans themselves initiate, but do not initially change the location of objects in the room that have been recognized. And objects seem to be recognized once they (or something similar) has appeared in past experience, even if the prior expectation of such objects are very small or practically zero. The reliability of the method does have to be 100%, as witnessed by ambiguities, misunderstandings, and known visual paradoxes. It is often capable of being distracted, as witnessed by 'gorilla in the room' experiments.

In physical quantum mechanics, the final mechanism of observations is the random selection of quantum alternatives $\phi_1, \phi_2, \ldots$. The result i of such selections comes back into the mathematics as a projection operator $P_i = |\phi_i\rangle\langle\phi_i|$, so that the initial wave function is changed as $\psi \leftarrow P_i\psi = c_i\phi_i$ for constant $c_i = N_i\langle\phi_i|\psi\rangle$ which is the degree of overlap between ϕ_i and ψ. The constant N_i has the value needed to normalize the new ψ to unity. We here expect something similar in mental quantum mechanics, whereby the mental state ψ is also changed by a projection operation when something like a ϕ_i is observed in the sensory sheet ψ_s. Indeed such seems plausible, since ψ is now purely ϕ_i and that should be the result of seeing the object i. However it must be more complicated than this. For multiple objects can be seen, and that would not be permitted by the above projection operation if the different ϕ_i have no

overlaps with each other: $\langle \phi_i | \phi_j \rangle = 0$. We rather believe that seeing successive objects *increments* our visual knowledge, not just changing it always to the most recent object seen. Furthermore, each new object seen has its own set of dimensions for rotation and translation: the whole mind is not rotated when imagining a new orientation of an object.

I therefore suggest a new kind of 'projection operator' appropriate for mental recognitions. When an object i is recognized, this new operator adds a new component (with its new dimensions) to the main mental state ψ. To do this we need a quantum formalism allowing for changing numbers of objects. The 'creation operators' a_i^+ of field theory are what we need. Using them, we can theorize that when a new object i is recognized during a visual fixation, the epistemic state ψ of the mind is incremented as

$$\psi \leftarrow \psi + |\phi_i\rangle\, a_i^+ \langle \phi_i | \psi = [1 + N_i\, |\phi_i\rangle\, a_i^+ \langle \phi_i |]\, \psi.$$

This is, therefore, to define new projection operators like $\mathcal{P}_i = 1 + N_i\, |\phi_i\rangle\, a_i^+ \langle \phi_i |$.

Such a 'fixation equation' is however still not quite right for vision. By the above equation, the projection operator $\mathcal{P}_i$ for producing knowledge state ϕ_i only functions when that knowledge is already present (to some non-zero extent) in the initial state of ψ. That is because it only has an effect if the overlap $\langle \phi_i | \psi \rangle \neq 0$. But it is *not* the final knowledge ϕ_i which is in ψ initially, only the appearances of objects from the sensory sheet ψ_s. It is those appearances which should be sought in the initial ψ. And the appearance of an object i depends on bodily translations, rotations, etc.

Let us define, therefore, for each object i its appearance $\Phi_i(\rho, \omega) = T(\rho)R(\omega)\phi_i$, given a specific translation vector ρ and a rotation vector ω. (By using 4-dimensional vectors, time shifts can be included in ρ, and with quaternions scale changes can be included in ω.) The recognition operator should now be something like $\mathcal{P}_i(\rho, \omega) = 1 + N_i |\phi_i\rangle\, a_i^+ \langle \Phi_i(\rho, \omega)|$. This could in the fixation equation as

$$\psi \leftarrow \mathcal{P}_i(\rho, \omega)\psi = \psi + N_i |\phi_i\rangle\, a_i^+ \langle \Phi_i(\rho, \omega)|\, \psi.$$

The recognition now correctly depends on the similarity of visual appearances according to the overlap integral $\langle \Phi_i(\rho, \omega) | \psi_s \rangle$ for the visual sheet component ψ_s. However, we have not yet explained what particular translations ρ or rotations ω enter into this equation. Recognition seems to occur for *any* translation or rotation (within some limits). How? Let us try to make an operator with a *superposition* of all possible rotations and translations, such as

$$\mathcal{P}_i = 1 + \int d\rho \int d\omega\, N_i |\phi_i\rangle\, a_i^+ \langle \Phi_i(\rho, \omega)|$$

This can be rewritten so that projection operator appears in the first part of the second term:

$$\mathcal{P}_i = 1 + N_i |\phi_i\rangle\, a_i^+ \langle \phi_i | \int d\rho \int d\omega\; T^+(\rho) R^+(\omega)$$

This operator effectively broadcasts the knowledge ϕ_i of object i to all possible translations and rotations (restricted by the limits of the two integrals, not yet specified). Then, whenever in a fixation one translation and rotation of the object produces an appearance that matches some part of the visual field, then the above 'composite projection operator' is enacted to produce new knowledge of the object in the sensory mind.

(more details to come).

Appendix A: Producing the Sensory Field

A scheme for how the visual field ψ_s in the sensory or corporeal mind could arise from a scheme of influx.

This scheme requires

1. that whenever any influx occurs, the upper source degree gains some awareness of how successful that influx was in achieving the desire that produced the influx.
2. So: the sensory mind starts of as a combination of all colors red, green, blue (RGB), as in white light from which all can be produced by modifications.
3. the sensory mind generates influx into the visual system (e.g. visual cortex) for each of the 3 colors: R, G and B.
4. the visual cortex accepts that influx according to its input from the color-sensitive cones in the retina. If red is being viewed, then influx of the R-kind is more successful, and so on.
5. So the sensory spiritual gets a lot of information about what combinations of R-influx, G-influx, and B-influx are successful, everywhere in the visual field to which attention is directed.
6. From that information of various 'success rates', the sensory spiritual generates in itself a mental image of the visual field as a function $\psi_s(c, \theta, \chi)$, where c is the color index R,G or B, the angle θ is the horizontal pan angle ($\pm 105°$ for humans), and χ is the vertical angle ($\pm 75°$).