Routledge
Taylor & Francis Group

Neuropsychoanalysis

An Interdisciplinary Journal for Psychoanalysis and the Neurosciences

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rnpa20

New project for a scientific psychology: General scheme

Mark Solms

To cite this article: Mark Solms (2020): New project for a scientific psychology: General scheme, Neuropsychoanalysis, DOI: 10.1080/15294145.2020.1833361

To link to this article: https://doi.org/10.1080/15294145.2020.1833361

Published online: 24 Oct 2020.

NEUROPSYCHOANALYSIS
https://doi.org/10.1080/15294145.2020.1833361

New project for a scientific psychology: General scheme

Mark Solms

Neuroscience Institute & Psychology Dept., University of Cape Town, Cape Town, South Africa

ABSTRACT
This is a revision of Freud's "Project for a Scientific Psychology: General Scheme." It updates the original, sentence for sentence where possible, in light of contemporary neuroscientific knowledge. The principle revisions are as follows. (1) Freud's conception of "quantity" (the precursor of "drive energy") is replaced by the concept of "free energy." This is the energy within a system that is not currently performing useful work. (2) Shannon's conception of "information" is introduced, where information is equivalent to unpredictability, and is formally equivalent to "entropy" in physics. (3) In biology, the fundamental purpose of "homeostasis" is to resist entropy – i.e., to increase predictability. Homeostasis turns out to be the underlying mechanism of what Freud called the "principle of neuronal inertia." (4) Freud's conception of "contact barriers" (the physical vehicles of memory) is linked with the modern concepts of consolidation/reconsolidation, whereby more deeply consolidated predictions are less plastic (more resistant to change) than freshly consolidated ones. (5) Freud's notion of sensory "excitation" is replaced with the concept of "prediction error," where only that portion of sensory input which is not explained by outgoing predictions is propagated inwards for cognitive processing. (6) Freud's conception of "bound" (inhibited) cathexis, the main vehicle of his "secondary process" and *voluntary* action is equated with the buffering function of "working memory"; and "freely mobile" cathexis (the vehicle of Freud's "primary process") is equated with the *automatized* response modes of the nondeclarative memory systems. (7) Freud's notion of ω (the system "consciousness") is replaced by the concept of "precision" modulation, also known as "arousal" and "postsynaptic gain."

ARTICLE HISTORY
Received 25 June 2020
Accepted 4 October 2020

KEYWORDS
Free energy principle; predictive coding; information; consciousness; reconsolidation

Author's introduction

Freud's "Project for a Scientific Psychology" is the Rosetta Stone of neuropsychoanalysis. It was the foundational text for the whole of what became known as metapsychology – Freud's basic theory of the functional dynamics of the mind – *and* a sophisticated model of the neural mechanisms by which those dynamics come about. In other words, the "Project" (as it came to be abbreviated) was the first attempt – Freud's own attempt – to achieve what the whole interdisciplinary endeavor called neuropsychoanalysis is trying to achieve today.

It is perhaps not surprising therefore that when the "Project" was first published (posthumously, in 1950) it caused an uproar. It enjoyed highly ambivalent responses, not dissimilar to those that have greeted our own efforts today in neuropsychoanalysis. For example, James Strachey, Freud's authorized translator, while pointing out that "the *Project*, or rather its invisible ghost, haunts the whole series of Freud's theoretical writings to the very end," hastened to add:

> There is a risk that enthusiasm may lead to a distortion of Freud's use of terms and may read into his sometimes obscure remarks modern interpretations that they will not bear. And after all we must remember that Freud himself ultimately threw over the whole neurological framework.

Strachey concluded: "The *Project* must remain a torso, disavowed by its creator" (*Standard Edition*, **1**, pp. 290, 293). Accordingly, an attempt by the great neuroscientist Karl Pribram to update the "Project" led to an embarrassing falling out with his co-author, Merton Gill, a leading psychoanalyst of the day. In the closing sentences of their joint book, they wrote:

> Where we differ is that Gill feels that psychoanalysis must go its own way and that means purging it of its natural science metapsychology, while Pribram welcomes psychoanalysis back into the natural sciences. Pribram doubts that the differing views of the two authors are really, in the long run, incompatible, while Gill finds them irreconcilable. (Pribram & Gill, 1976, p. 169)

When I first came upon the "Project," while still a student, it was not my Neuropsychology professors

CONTACT Mark Solms ✉ mark.solms@uct.ac.za
Supplemental data for this article can be accessed here https://doi.org/10.1080/15294145.2020.1833361

who introduced me to it but rather a lecturer in the department of Comparative Literature, Jean-Pierre de la Porte (who has kindly contributed a commentary, below). I was mesmerized; it seemed to address all the questions that my training in neuropsychology neglected, the very questions that had drawn me to the field in the first place. As Oliver Sacks famously remarked: “Neuropsychology is admirable, but it excludes the psyche – it excludes the experiencing, active, living ‘I’” (Sacks, 1984, p. 164).

As soon as I found the opportunity, therefore, one long weekend in 1984, I sat down with the “Project” on one side of my desk and my trusted neuropsychology textbook (Luria, 1980) on the other, and then I tried in vain to translate Freud’s opaque terms and concepts into their presumptive contemporary equivalents. For three days I did nothing else. I was completely spellbound, but ultimately frustrated; put simply, too much was unknown. I concluded that what was required was not a theoretical exercise but rather a comprehensive program of interdisciplinary, experimental and clinical *research*. I have spent the rest of my working life trying to get that research program off the ground.

My great good fortune in the intervening decades between then and now was to have opportunities to work with and thereby closely learn from two outstanding scientific pioneers of our time, first, the affective neuroscientist Jaak Panksepp, and second, the computational neuroscientist Karl Friston. The integrative achievements of these two great minds, far more than my own clinico-anatomical research efforts, brought me to the present juncture, where I believe the moment has come for us to attempt what Merton Gill declared impossible and undesirable when I was still a student: to update Freud’s “Project” in such a way that it can once again perform its historic role as the Rosetta Stone of our field.

It is important to clarify here: this is literally a *revision* of Freud’s classic paper; it is not an original article in the usual sense of the word. Some of Freud’s concepts are replaced outright but most are found to still be viable, and in surprisingly many places the wording of the original text is followed exactly. I thought it appropriate to update it in this literal way so that scholars in psychoanalysis and neuroscience can see precisely how the model that I am now proposing builds upon, and differs from, Freud’s *Ur*-text. This approach unfortunately has the necessary consequence that my paper is very dense and obscure in places, since the whole point of it is to track the original (dense and obscure) “Project” as closely as possible; to update it, word for word, sentence for sentence and paragraph for paragraph. Readers will find it useful, therefore, to compare my revision with Freud’s original. To this end, the editors have kindly provided a “marked-up” version of the revised text to facilitate this comparison (see Supplemental data). Please also note, the paraphrasing of Freud’s “Project” requires this paper to be written in the first person singular, as the original was; so it is important to acknowledge at the outset that its main ideas are based on joint theoretical work between the author and Karl Friston (Solms, 2019; Solms & Friston, 2018).

A fresh and fully explicated account of the new model, in its own terms, requires a book-length treatment. This treatment will be published shortly, under the title *The Hidden Spring* (Solms, 2021), written in a language and style that should be comprehensible to non-specialist readers. My profound hope is that the model described here (and there, in expanded form) will provide our field with a meta*neuro*psychology which can serve the purposes for the foreseeable future that the “Project” served in the past. In saying this, I want to make explicit what I just alluded to, namely that this model rests upon the work of giants, not only Freud, but also Panksepp and Friston. Of course, there were many others besides them. In fact, it could be said that the true foundations for this model were laid by Hermann von Helmholtz, one of the founding fathers of the Berlin Physical Society (sometimes called the Helmholtz school of medicine), whose work so obviously undergirded the original “Project” and all that Freud built upon it.[1]

One of the editors asked me to write this Introduction to encourage daunted readers to persevere with the challenging task of understanding my paper. I find that I cannot do any better than she did:

> In my view, the value of the piece (and the full treatment you are elaborating elsewhere) is the articulation of an infrastructure of a truly neuropsychoanalytic model of the mind and brain. This model allows for theoreticians, clinicians, and researchers to account for the complex interactions between cognition and emotion; impulse and regulation; consciousness and unconscious processes; genetics and experience; etc. All of this is meaningful to clinicians who work with these dynamics in everyday work. I think if you can articulate that, it may encourage a few more readers to wade into the dense piece – it really opens up after a few pages, and I think any readers who can get into the middle and end sections will find it thrilling. (Maggie Zellner, personal communication)

So, take a deep breath; here we go.

KEY TO ABBREVIATIONS IN THE *NEW PROJECT* [2]

Q	=	*External states (unknowable in themselves)*
$Q\eta$	=	*Generative model of external states*
φ	=	*Sensory states*
M	=	*Active states*
ψ	=	*Predictions (based on $Q\eta$)*
e	=	*Errors (based on φ and its prediction ψ)*
ω	=	*Precisions*
F	=	*Free energy (based on e and ω)*

New project for a scientific psychology

Introduction

The intention is to attempt, once more, to furnish a psychology that shall be a natural science; that is, to represent mental processes as quantitatively determinate states of specifiable physical elements, thus making those processes perspicuous and free from contradiction. Two principle ideas are involved: (1) What distinguishes activity from rest is to be regarded as *F*,[3] subject to the general laws of information.[4] (2) Neurons are to be taken as the physical elements.

Neurons and information processing – similar models are now commonplace.[5]

(1) *First principle theorem*

The quantitative conception

The theorem "what distinguishes activity from rest is to be regarded as *F*" is derived from statistical physics but it manifests in clinical observation, especially where excessively intense feelings are concerned – as we shall see, the quantitative characteristic of neural functioning emerges more plainly with affects than with cognition. Processes such as stimulus, substitution, conversion and discharge, which were introduced into metapsychology to explain the dynamics of affect, suggested the existence of an underlying mental *energy* (called "drive") as a quantity in a state of flow.[6] It seemed legitimate to attempt to generalize what was initially recognized clinically. Starting from this consideration, it is now possible to lay down a basic principle of neuronal activity in relation to *F*, which promises to be highly enlightening, since it appears to comprise its entire function.[7] This is the free energy principle: neuronal systems tend to minimize *F*. On this basis the structure and development as well as the functions of the nervous system, as they relate to mental life, can be understood.

In the first place, *homeostasis* underwrites a "principle of neuronal inertia."[8] This is a conservative tendency which maintains the organism within its phenotypically viable bounds, across many biological parameters. The design of the nervous system clearly serves this principle. The structural dichotomy of neurons into *sensory* (homeostatic "receptor") and *motor* (homeostatic "effector") types is a contrivance for minimizing *F*. Sensory demands for work (increasing *F*) generate motor actions. *Reflex movement* is therefore intelligible as a primitive form of "work" (overseen by homeostatic "control centres") in the service of this basic tendency. The imperative to minimize *F* supplies the aim and purpose of reflex movement. If we go further back from here, we can in the very beginning of life conceptualize the organism, which is but an inheritor of the conservative imperative of *all* self-organizing systems, as cloaking these systems in a *Markov blanket*. Markov blankets induce a partitioning of states into internal (system) and external (not-system) ones, so that the system is insulated from the entropic forces that surround it, to resist those forces. External states (Q) can only influence the internal states of a self-organizing system *vicariously* as states of its blanket. The blanket is itself partitioned into "sensory" and "active" states (these are its homeostatic receptors and effectors, embodied in the sensory and motor systems of neurons: φ and M). This yields a circular form of causality: external states can influence the internal states of a self-organizing system via the sensory states of its blanket, while the internal states of the system couple back to the external world through its active states. Crucially, the sensory states *feed back* the external consequences of prior active states, and thereby adjust the posterior[9] states of the system via the homeostatic control centers that generate predicted consequences. This is learning. A complex self-organizing system is equipped with a meta-control center which controls its total "generative model" of the causes of its external states ($Q\eta$) and registers "prediction errors" in general. This predictive model, based in instincts *supplemented by learning*, serves the singular aim and purpose of maintaining the viable bounds (the ongoing existence) of the system, through the minimization of free energy, where *F* is *a measure of the average difference between the system's predicted and actual sensory states* over a given period of time (i.e. it is a measure of the demand for work by the system). Minimizing *F* represents the *primary* function of the nervous system. Here is room for the development of a *secondary* function. For among the paths for minimizing *F* those are preferred and retained which involve a cessation of prediction errors: *minimization of surprise*. Here in general there is a proportion between the free energy derived from surprisal and the work necessary for its minimization, so that the free energy principle is not upset by this.

The insulation of biological systems from surrounding entropic forces is, however, broken through from the first owing to another circumstance. With an increasing complexity of the interior of the organism, the nervous system receives entropic perturbations from the somatic element itself – *endogenous* sources of *F* – which have equally to be minimized. They have their origins in the cellular processes of the body and give rise to the major needs: hunger, respiration, sexuality and the like. (In this formulation, the internal milieu of the organism is "external" to the blanketed nervous system, no less than the world outside is. The generative model within the blanket represents and regulates the body, but it is not the body itself.) The system cannot simply adjust its prior predictions to minimize surprisal regarding the viable states of its body, as it can with prediction errors regarding sensory samples of the external milieu; it cannot employ "model updating" of autonomic reflexes since its model of the own bodily functions is dictated by its phenotype. Vegetative error signals only cease subject to particular conditions: when prior predictions representing the major needs are met, which means they *must* be realized in the external world when autonomic regulation reaches its threshold. (Cf. the need for nourishment, for instance.) In other words, an internal sensory demand such as hunger can only be met by actively taking nutrients, and this requires motor action; no amount of updating a predictive model can meet nutritional needs directly. In order to accomplish the compulsory actions, which deserve to be named "specific," the changes demanded by Friston's law must involve adjustment of predictions concerning the sensory states that flow from the organism's actions in the external world, since the system is being subjected to conditions which may be described as *the exigencies of life*. In consequence, the nervous system is obliged to abandon its original trend to reflex action. It must put up with tolerating a store of free energy sufficient to meet the demand for the specific action: whenever an active state driven by a major need does not have the predicted sensory consequences, the generative model of the external world must be updated so that *supplementary motor skills* may be acquired. This is the impetus to *learning from experience*, and therefore to *voluntary action*. The primary trend to minimize *F* persists in learning, but trial-and-error processes take time, and *uncertainty* (the mortal enemy of self-organizing systems) is greatly increased in the process. The obligation to *tolerate error* while uncertainty is resolved therefore implies that *F* can never be nullified; it can only be minimized.[10] All the functions of the nervous system can be comprised either under the aspect of the primary function or of the secondary one imposed by the exigencies of life.

(2) *Second principle theorem*

The neuron theory

The idea of combining this quantitative theory with the theory of neuronal functioning is the second pillar of this thesis. The main substance of the theory (no longer controversial) is that the nervous system consists of distinct and similarly constructed neurons, which have contact with each other through the medium of synapses, which terminate upon one another as they do upon portions of non-nervous tissue (e.g. muscle), and in which certain lines of *transmission* are laid down in so far as the neurons receive signals through dendrites and give them off through an axon. They have in addition *modulatory* functions which shall be described later (the importance of the distinction between transmission and modulation is greatly underestimated).

If we combine this account of the neurons with the concept of a generative model ($Q\eta$) in service of the free energy principle, we arrive at the idea of an active ("cathected") prediction which at other times may be inactive.[11] Friston's law finds its expression in the fundamental hypothesis of a cathectic signal flowing from a predictive neuron to a sensory one, with the first neuron transmitting the *expected signal* to the second. These two classes of neuron reflect the basic function of the nervous system, since predicting the incoming signal minimizes surprisal and therefore *F*.[12] The secondary function of the nervous system, however, which calls for learning, is made possible by the assumption of *error signals* which may override the prior predictive ones when the distribution of sensory states does not match with what was expected. In other words, because predictive models are imperfect things, the system must expect error signals and must constantly update itself to accommodate them. The structure of the nervous system, which is layered somewhat like an archaeological site,[13] dictates that the predictive signals will be *centrifugal* and the error signals *centripetal*. This layered arrangement yields a conception of the system's generative model unfolding over a *concentric predictive hierarchy*, which proceeds from the homeostatic core to the sensory periphery, with each layer endeavoring to predict the pattern of neuronal activity that will occur in the layer immediately beyond it. This makes it probable that error signals are *progressively* resisted as they are propagated inward from the surface to the depths, with *more error being tolerated (expected) towards the periphery than towards the core*. The resistances would be located at the interfaces between the layers, at their points of *contact*, which in this way assume the value of *barriers*.[14] The hypothesis of *resistance* is fruitful in many directions.

(3) *Resistance*

The first justification for this hypothesis arises from the consideration that there are many more centrifugal paths of transmission in the nervous system that centripetal ones, not only in the motor pathways but in sensory ones too. Far fewer neurons propagate signals from the sense organs to the cortex than the other way around. For example, the ratio of afferent connections to efferent ones in the lateral geniculate body (which relays information from the eyes to the visual cortex and vice-versa) is about 1:10. The heavy lifting is done by the predictive signals that *meet* the sensory ones. The same general arrangement applies to centripetal and centrifugal connections within the cortex itself.[15] This saves an enormous amount of information processing, and therefore metabolic work, as predictive signals (ψ) "explain away" the sensory (φ) ones. Considering that the brain consumes about 20% of our total energy supplies, this is a valuable efficiency.[16] This gives us a hint that incoming signals are to be linked with Shannon entropy. The hypothesis of resistance therefore coincides with the now widely accepted notion that the cortex generates, in the first instance, not a bottom-up assemblage of incoming sensory (φ) data but rather a top-down "fantasy" – an unconscious predictive (ψ) "hallucination."[17]

Furthermore, the theory of resistance can be turned to advantage as follows. A main characteristic of nervous tissue is *memory*: that is, quite generally, a capacity for being permanently altered by single occurrences – which offers such a striking contrast to the behavior of a material that permits the passage of a wave movement and thereafter returns to its former condition. A psychological theory deserving any consideration must furnish an explanation of memory. Now any such explanation comes up against the difficulty that it must assume on the one hand that neurons are permanently different after an excitation from what they were before, while nevertheless it cannot be disputed that, in general, fresh excitations meet with the same conditions of reception as did the earlier ones. It would seem, therefore, that neurons must be both influenced and also unaltered, unprejudiced. We cannot offhand imagine an apparatus capable of such complicated functioning; the situation is accordingly saved by attributing the characteristic of being permanently influenced by excitation to one class of neurons, and, on the other hand, the unalterability – the characteristic of being fresh for new excitations – to another class. This coincides with the distinction in computational neuroscience between "*prediction units*" and "*error units*" – a distinction, moreover, which is not only found in machine-learning contexts but one that can also appeal to a wide range of neurophysiological findings (e.g. evoked cortical responses) for its support.[18]

The theory of resistance, if we adopt this solution, can express it in the following terms. There are two classes of neurons: (1) those which allow stimuli to pass into the brain as though there were no barriers and which, accordingly, after each passage of excitation are in the same state as before, and (2) those whose resistances make themselves felt, so that they only allow stimuli to affect them with difficulty or partially.[19] The latter class may, after each excitation, be in a different state from before and thus afford the possibility of representing memory.

Thus there are *incoming* neurons (offering no resistance and retaining nothing), which serve for error signals, and *outgoing* ones (loaded with resistance, and holding back *F*), which are the vehicles of memory (i.e. representation) and so of cognitive processes in general. Henceforward I shall call the former system of neurons (flowing from φ) the *e* system and the latter one (predicting φ) the ψ system.

It will be well now to clear our mind as to what assumptions about the ψ neurons are necessary in order to cover the most general characteristics of memory. This is the argument. Memories are basically predictions; they are about the past but they are *for* the future. Predictions are permanently alterable by the passage of an *e* signal, but, crucially, ψ neurons can also code for *expected e*. If we introduce the concept of synapses: their synaptic weighting is brought into a permanently altered state. And since psychological knowledge shows that there is such a thing as *consolidation* on the basis of repeated learning, this synaptic alteration must consist in consolidation. Consolidation may therefore be equated with resistance to change. *An increasingly consolidated neural assemblage is accordingly decreasingly plastic* (cf. Hebb's law).[20] Combining this insight with the notion of a predictive hierarchy, systems consolidation in the hierarchy must consist in the deeper ψ layers becoming progressively more resistant to change, and so *more like core homeostatic predictions*. They will behave more like reflexes; they will expect (tolerate) less *e* and become more *automatized*. We shall describe this state of the successive layers as their degree of *certainty*. We can then say: Memory is represented by the parameters of certainty existing between the ψ neurons.

What, then, does "certainty" in the ψ neurons depend on? According to empiricist doctrine, the memory of an experience (that is, its continuing operative power) depends on a factor which is called the magnitude of the impression and on the frequency with which the same impression is repeated. Translated into our theory: Certainty depends on the *e* which passes through the ψ hierarchy during a learning process and on the number of repetitions of the process. From this, then, we see that *F* is the operative factor and that the *magnitude* of surprisal (i.e. the depth to which an *e* signal travels through the hierarchy)[21] plus its *generalizability* inversely determines predictive power. Generalizability, in turn, is contingent upon the balance between model "accuracy" and "complexity."[22] At the periphery of the hierarchy, short-term complexity prevails, at the cost of long-term generalizability which is enjoyed by the deeper predictive layers. Translated into functional-anatomical terms: Nondeclarative (subcortical) memory traces are more certain than declarative (cortical) ones because they are optimized for simplicity rather than accuracy.[23] This makes them more generalizable, both spatially and temporally. But generalizability comes at a price: less complex models are less accurate *when the context varies*. The greater complexity of cortical predictions coincides with their higher plasticity and lower automaticity – that is, their tolerance of expected error. In a word, the cerebral cortex specializes in *contextual* memory; it restores model accuracy in complex and therefore unpredictable situations. A compromise is inevitable: less automaticity means more uncertainty and therefore higher values of *F*. That means greater demand for work; so, the system codes *Qη* parameters with as much certainty as it can get away with.

Here we are almost involuntarily reminded of the endeavor of the nervous system, maintained through every modification, to avoid being burdened by *F* or to keep the burden as small as possible. In the theory on offer here, under the compulsion of the exigencies of life, the nervous system was obliged to tolerate a large degree of uncertainty. This necessitated an increase in the number of its neurons, and these had to be *ψ* neurons which were arranged hierarchically: the expanded mammalian forebrain. It now avoids, partly at least, being overwhelmed with *F* (or unpredictability) by setting up resistances between the layers of the hierarchy through deepening degrees of consolidation. Stated conversely: the system avoids its own destruction by expanding the *expectation of uncertainty* towards the periphery of its generative model. The predictions encoded by the newer *ψ* layers are therefore more transient (since they are less generalizable). This is *short-term* memory, the pivotal vehicle of tolerated uncertainty, which provides a buffer for mental *work*.[24] This buffer has significant capacity constraints, which further explains the requirement for its predictions to be consolidated into the long-term systems. It will be seen, then, that progressive consolidation from the periphery towards the core (progressive certainty and resistance) serves the primary function of the nervous system.

The necessity for finding a place for memory calls for something further from the theory of resistance. Every *ψ* neuron must in general be presumed to have several paths of connection with neurons in the levels beyond it; their connections are more widespread than are those of the *e* neurons. On this, indeed, depends the possibility of *choice* that is so characteristic of uncertainty. We shall see later how choice is governed by Friston's law. For now, consideration need only be given to the most basic choice that is faced when a cathectic prediction confronts error signals: If an action does not yield the predicted sensory consequences, then the system must either (1) *change its prediction* to better explain the data, or (2) if it remains confident about the original prediction, it must obtain better data; that is, it must perform actions that will *change its sensory input*. These two options – changing *ψ* or *φ* – are the fundamental mechanisms of *perception* (i.e. here-and-now representation) and *action* respectively. In Bayesian terms: the brain has two alternative ways of responding to prediction error. When faced with a hypothesis to which decreasing "posterior" probability applies, it creates a better fit between the hypothesis and the data by changing either (1) its "prior" prediction or (2) its input. The difference between these alternatives comes down to the statistical *direction of fit*: error is reduced if the prediction is changed to match the sensory input, and it is also reduced if the sensory input is changed to match the prediction. Organisms alternate between these two options all the time. (Think of a field-mouse darting through the scrub, stopping to look around, darting again, stopping to look around again, and so on.) Perception and action are more similar than they seem. They are just alternative routes for reducing surprisal and therefore *F*. Nevertheless it is interesting to observe that *prior* predictions which give rise to errors can only be confirmed (as they always must be in the case of endogenous needs) through renewed *action*; there is therefore a special relationship between drive and action.

It remains to be seen in what else resistance consists. A further idea might be: By inverting the causal dependencies that shaped the predictive hierarchy, the brain produces perceptual *inferences*. In other words, patterns of causation that were *learnt* in the world become patterns of prediction that *explain* the world.[25] Inverting the hierarchy simply means shifting from learning to predicting on the basis of what was learnt: reversing the system's probabilistic model of the causal regularities that exist. The system then uses this model – this *virtual* world – to generate inferences that guide its actions, which, in turn, should be viewed as hypothesis-testing to improve the model (when all goes well). The world of phenomenal experience is derived indirectly, from *Qη*, not from *Q*.[26] Perception and action both proceed from the inside outwards –

always from the viewpoint of the system. They are the system's "best guesses" as to what lies beyond the blanket: provisional answers to the questions it is putting to the world. This underscores the *intentionality* of biological systems. The information flow diagrams of cognitive science routinely overlook the fact that "information" implies *communication* between an information source and an information receiver – a question asker. That is why information is coded in bits, 1s versus 0s: "yes"s versus "no"s. The diagram-makers fail to consider this crucial issue: who is asking the questions that evoke these responses?[27]

(4) *The biological standpoint*

The hypothesis of there being two systems of neurons, for prediction (ψ) and error (e) respectively, of which e consists in centripetal elements and ψ centrifugal ones, seems to provide an explanation of this one of the peculiarities of the nervous system – that of retaining and yet remaining capable of receiving. All mental acquisitions would in that case consist in the negentropic *organization* of the ψ system through partial and locally determined lifting of the variable resistances that distinguish e from ψ.[28] With the advance of this organization the nervous system's capacity for fresh reception would literally have reached a barrier.

Anyone, however, who is engaged scientifically in the construction of hypotheses will only begin to take their theories seriously if they can be fitted into knowledge from more than one direction and if the arbitrariness of a *constructio ad hoc* can be mitigated in relation to them. It will be objected against our hypothesis of resistance that it assumes two classes of neurons with a fundamental difference in their conditions of functioning, though there is at the moment little other basis for the differentiation. At all events, morphologically (that is, histologically) nothing is known of the distinction.[29]

Where else are we to look for this division into classes? If possible, in the biological development of the nervous system, which, in the eyes of natural scientists, is, like everything else, something that has come about gradually. We should like to know whether the degree of resistance can have had a different significance biologically, and if so, by what mechanism the two classes have developed this difference. What would be most satisfactory, of course, would be if the mechanism we are in search of should itself arise out of the primitive biological part played by resistance; if so, we should have a single answer to both questions.

Let us recall, then, that from the first the nervous system had two functions: the reception of stimuli *from outside* and the discharge of excitations of *endogenous* origin. It was from this latter obligation, indeed, that, owing to the exigencies of life, a compulsion came about towards further biological development. We may then conjecture that it might actually be our function of variable resistance which had assumed one of these primary obligations. The flexible system would be the group of neurons which the exogenous stimuli reach; the automatic system would contain also the neurons which receive the endogenous excitations. In that case we would not have *invented* the function of resistance, we should have *found* it already in existence – in the form of automatized reflexes and instincts which are highly resistant to change. It still remains to identify them with something known to us. In fact, we know from anatomy a system of neurons (the grey matter of the spinal cord and cranial nerves) which is alone in contact with the external world, and a superimposed system (the grey matter of the brain) which has no direct peripheral connections but to which the development of the nervous system and the cognitive functions are attached. The prosencephalon fits pretty well with our characterisation of the system ψ, if we may assume that e paths lead directly, and independently of φ, from the interior of the body to the brain – as indeed they do, to the diencephalic hypothalamus. (The prosencephalon is divided into telencephalic and diencephalic components.) Now, the derivation and original biological significance of the prosencephalon are not known to anatomists; according to our theory, it would, to put it plainly, be a *sympathetic ganglion*. Here is a first possibility of testing our theory upon factual material. The graded resistance introduced by the expanded memory systems of the telencephalon – which surrounds the diencephalon – would provide the increasing possibilities of *choice* that are demanded by the exigencies of life.

We will provisionally regard the ψ system as identified with the telencephalic cerebrum. It will now easily be understood from our biological remarks why it is precisely ψ that is subjected to further development through an increase in the number of neurons, for the tolerance of *F*. And it will now be realized how expedient it is that ψ should consist of neuronal assemblages with varying degrees of plasticity, since otherwise it would be unable to meet the requirements of the specific action. Hypothalamic reflexes are almost completely resistant to change. But how did ψ arrive at the characteristic of graded resistance? After all, the automatic diencephalic core receives e signals from the internal milieu, which is no less "external" to the Markov blanket than is the outside world; if it shows so little plasticity, why should ψ's memory systems? To assume that there is an ultimate difference in synaptic functioning between the neurons of the two systems has an unfortunate tinge of arbitrariness, though it would still be possible to show that telencephalic and hypothalamic neurons have very different patterns of connectivity. Moreover, later we shall see that hypothalamic (and some other core diencephalic and mesencephalic) neurons involve a mode of communication that is strikingly different from that of classical synaptic transmission.[30]

Another way out seems more fruitful, although theoretical. Let us recall that the homeostatic "beliefs" which comprise the viable states of the phenotype give rise to prior predictions that *must* be confirmed (e.g. "I shall remain between 36.5 and 37.5°C"). Predictions arising from such beliefs have very different biological implications from those associated with external sources of *F*, such as, for example: "the person in front of me is going to turn left." Exteroceptive beliefs can in large part be updated repeatedly (hence their large tolerance of *e*) whereas their endogenous counterparts are matters of life and death. That is why endogenous beliefs are encoded *from the outset* at the core of the predictive hierarchy; their certainty is complete. Stated in terms of the theory of resistance: the *e* values that trigger them enjoy the highest possible "magnitude" and the consequences that flow from them the greatest "generalizability"; the answer to the questions the system asks of its internal milieu must ultimately always be yes. The certainty of exteroceptive hypotheses, by contrast, may vary by degrees. Hence the differing biological implications of this series of perceptual inferences: "the person in front of me is turning left"; "the person in front of me is racing towards me"; "the person in from of me is lunging at me with a knife." The consequences of these inferences enjoy increasing certainty; they therefore trigger beliefs that are encoded at progressively deeper layers of the hierarchy. A difference in the essence of neurons located within the expanding telencephalon on the one hand and the diencephalic core on the other is thus replaced by a difference in the environment to which they are destined.

Now, however, we must examine our assumption – whether we may say that the magnitudes of *e* signals reaching the neurons from the internal periphery are of a higher order than those from the external periphery of the body. There is in fact much that speaks in favor of this.

In the first place there is no question but that the internal milieu is the source of relentless quantities of *F*, since, due to the exigencies of life, the major needs simply must be met in the outside world when they exceed the capabilities of vegetative reflexes (such as metabolizing the energy supplies deposited in adipose tissues, in the case of the need for nourishment); and so we must *learn* how to master them. That is why life is difficult, and also why bodily needs typically trump other errors in the hierarchy of priors. However, bodily needs do not exhaust the range of endogenous demands that must be met in the outside world, with all its attendant uncertainties – that necessitated the development of the ψ systems, which required the system to tolerate tonic levels of *F*. In the taxonomy of Panksepp, bodily needs – which he divides into "sensory" (exteroceptive) and "homeostatic" (interoceptive) subtypes[31] – are accompanied by a number of "emotional" needs. These, too, are *endogenous* and they are just as inexorable as the bodily ones.

To the best of our knowledge, there are seven of them in us mammals. (1) SEEKING demands engagement with the external world,[32] which is where all the bodily needs must be met. The innate prior prediction attaching to this need triggers "foraging" behavior, which gives rise to all manner of learning from experience. The supplementing of such innate predictions (called "instincts") through learning is necessary with all the emotions, as with bodily reflexes, if not more so.[33] (2) LUST demands the release of erotic tension, since sexual behavior will – in the average case – facilitate reproductive success. Sexual reflexes and instincts are very rudimentary (e.g. stroking the clitoris and penis, vaginal lubrication and erection, lordosis and mounting behavior with intromission); they therefore require much supplementation. (3) RAGE demands removal of agents that come between the organism and the objects of its other needs. The associated innate behavior is "affective attack," but as with all instincts this is frequently not the best prediction, so it too requires supplementation through learning. (4) FEAR demands safety from predators and other dangers to life and limb. The associated instincts are freezing and fleeing, but the animal must learn *what* to fear and *what else* to do. (5) PANIC/GRIEF demands the reliable presence of "attachment" objects: reliable caregivers in the broad sense; but these must be imprinted. "Protest" and "despair" behaviors are the relevant instinctual responses.[34] (6) CARE demands the wellbeing of our offspring and their (learnt) equivalents. This gives rise to innate nurturing behaviors – so-called maternal instincts, which again require much supplementation. (7) PLAY demands engagement with the social group, with a view to maximizing status and thereby access to its territorial resources. The associated instinct is rough and tumble play, which is governed by innate rules concerning fairness and boundaries. Juvenile mammals *must* play, not least because they must learn about reciprocity.

It will be noticed that all these emotional needs (just like the bodily ones) represent deviations from homeostatic settling points – e.g. “no frustrating obstacles are in my way,” “no dangerous threats are in sight,” “none of my dependants is distressed” – which are the emotional equivalents of bodily satiation. The requirement for organisms to remain within these biologically viable emotional bounds is no less obligatory than it is with the bodily needs. It will also be noticed that emotional needs inevitably *conflict* with one another, even more so than bodily ones. This means that needs must be *prioritized*, since the organism cannot achieve everything at once. That demands compromise – which is again facilitated by learning. All these facts count as exigencies of life.

(5) *The problem of quantity*

As we have seen, complex organisms have multiple endogenous needs, each regulated by their own homeostatic mechanism, all of which contribute *e* values to the overall calculation of *F*. Biological needs are these error values. When a need cannot be met autonomically, it becomes a “drive”: a demand upon *the mind* to perform work. Having a need and feeling a need (i.e. prioritizing it) are not the same thing, as we shall see later. Drive demands are registered subjectively as affects, the fundamental feature of which is *valence*. Although this valence is a qualitative feature of affect (pleasure is “good” and unpleasure is “bad”) it is nevertheless possible to quantify it: the more an *e* value deviates from the prescribed settling point, the greater the unpleasure, and vice-versa. Thus, at any given moment, an organism’s hunger value might be 3/10 (which is worse than 1/10) and its thirst value might be 2/10 (which is better than 5/10). This, if it is so, is easily understandable.

However, all the more interesting are certain perspectives and conceptions which arise from these assumptions. In the first place, need values cannot be summated in any simple way. The multiple needs cannot be reduced to a common denominator; they must be evaluated on separate, approximately equal scales, so that each of them can be given its due. One cannot say that 3/10 of hunger plus 1/10 of thirst equals 4/20 of total need, and then try to minimize that sum, because each need must be satisfied in its own right. It therefore makes sense for the system to distinguish its *e* values *categorically*, so that they may be computed independently. The different categories of quantity also have different implications for the system in different *contexts* (for example, hunger trumps sleepiness in some situations but not others). This contributes greatly to uncertainty – which requires more computational complexity, which means more information, which means more entropy and *F*. Thus, in addition to hierarchical levels of processing, categorization becomes a necessity when the biological value of different quantities changes over time (e.g. 8/10 for *A* is currently but not always worth more than 8/10 for *B*).

It is conceivable that an extremely complex set of model algorithms could evolve to calculate relative survival probabilities in all predictable situations, to enable organisms to automatically prioritize needs (and thereby actions) on this basis. However, such complex models are extremely expensive, in every sense of the word. Statisticians call the exponential increase in computational resources necessitated by a linear increase in model complexity the “combinatorial explosion.” Here we face another form of capacity constraint, generalizing from that of short-term memory and the related problem of over-fitting.

Compartmentalization is the standard statistical method for achieving optimal balance between complexity and accuracy. This takes many guises but what matters here is that it enables the organism to rank its needs and attendant predictions (the *salient* sources of expected *F*) over time, and to focus computational efforts on a *prioritized* compartment. It is as if the system says:

> under present conditions, this is the category of *e* in which computational complexity cannot be sacrificed; this is the one among the several categories of *e* signal converging upon me that provides the greatest opportunity for minimizing my free energy.

Compartmentalization, then, appears to be the statistical-mechanical basis for the observed fact that each affect possesses not only a *continuous* hedonic valence (a *degree* of pleasure and unpleasure, which is something common to all affects) but also a *categorical* quality (so that thirst *feels* different from separation distress, which feels different from disgust, etc.). But how can the quality of feelings like thirst versus disgust, etc., be quantified?

(6) *Pain*

All contrivances of a biological nature have limits to their efficiency, beyond which they fail. This failure is manifested in phenomena which border on the pathological – which might be described as normal prototypes of the pathological. We have found that the nervous system is contrived in such a way that the major exteroceptive sources of *F* are kept off from the *e* neurons and still more from ψ, by the Markov blanket and by resistance. Is there a phenomenon which can be brought to coincide with the failure of these contrivances? Such, I think, is *pain*.

Everything that we know of pain fits in with this. The nervous system has the most decided inclination to *withdrawal from pain*. We see in this a manifestation of the primary trend against a raising of *F*, and we infer that pain consists in the *irruption of large e's into ψ*. The two trends are in that case a single one. Pain sets the *e* system as well as the *ψ* system in motion; there is minimal resistance to pain transmission, it is the most imperative of sensory processes. Thus the *ψ* neurons seem permeable to it; it therefore consists in the transmission of *e* signals of a very high magnitude.

We saw previously that the magnitude of an *e* signal determines the depth to which it penetrates the *ψ* hierarchy. This implies that pain signals penetrate it completely. *They behave like endogenous e*; which fits neatly with the observation that pain dispenses with the distinction between the *e* and *ψ* systems. It is both an exteroceptive sensation *and* an affect. It therefore behaves like a core homeostatic belief, which is why a highly generalizable withdrawal reflex is triggered by pain. Now we realize that these characteristics apply to *all* the "sensory affects," not only to pain; consider disgust, for example, and fright – but here we will use pain as our model example. Attention is always reflexively grabbed (as opposed to voluntarily directed) by sensory affects. It is via this evolutionary bridge, we may suppose, that *valence* was originally extended from the interoceptive core onto external representations of objects. In fact, sensory affects are probably the origin of aesthetic experience in general (e.g. music, painting and dance). Emotional affects, too, we now realize, entail innately valenced object representations (e.g. becoming separated from a caregiver or being attacked by a predator) which are the basis for learning from experience – that is, for associating empirical objects by analogy with the innate ones.[35] The deep connection between bodily and emotional affects now comes into view. For example, pain and PANIC/GRIEF are modulated by the same molecules – μ opioids – so that the mental pain of separation may be considered an evolutionary extension of physical pain. Similarly, pain experiences encode FEAR associations like nothing else; fear conditioning occurs through single-exposure learning. It is not difficult to see why that should be the case.

Two questions flow from these considerations, and they will prove pivotal. First: where do pain signals terminate after they have breached *ψ*? In functional-anatomical terms, where is the core of the predictive hierarchy located; what is the *φ* terminus and the ultimate source of *M* – the hierarchy's meta-homeostatic control center? This coincides with a major issue we addressed before: where is the system's question-asker, the information receiver? It is in the *periaqueductal grey* (PAG) of the brainstem. This becomes even more interesting when we learn that the PAG is the final destination of literally all the brain circuits transmitting endogenous *e*. It is, in its turn, tightly interconnected with two adjacent brainstem structures – the tectum and the midbrain locomotor region – which collectively comprise what Merker calls the "midbrain decision triangle," what Panksepp calls the primal "SELF."[36] Whereas the PAG registers residual endogenous needs, the tectum generates a compressed two-dimensional representation of the external world – a "saliency map" or "priority map."[37] The interface between these two structures generates a picture of "where things stand now" in terms of both subjective *needs* (*e*) and objective *opportunities* (*φ*) at the completion of each action cycle (*M*). Thus, the midbrain decision triangle performs the crucial prioritization function discussed previously: "what to do next" – since the organism cannot do everything at once; since it cannot suppress all sources of *F* simultaneously. The second question is: how is this decision registered by the organism? The answer is: as *feeling*. A prioritized need is the one that is felt (i.e. selected). This registers the currently most salient source of expected *F*. Felt affects therefore convey the decision mentioned before: "under present conditions, this is the category of *e* in which computational complexity cannot be sacrificed." The *feeling* of a need (as opposed to the mere existence of one) makes a big difference to what the subject of that need will do next. Hence the distinction between needs and drives. So, in a sense, the PAG is both the terminus of every affect circuit and the genesis of every newly felt affect. This brings us back to the question with which I ended the last section: how can feeling (the subjective aspect of drive) be quantified?

(7) *The problem of quality*

Hitherto, flowing from this question, nothing whatever has been said of the fact that every psychological theory, apart from what it achieves from the viewpoint of natural science, must fulfill yet another major requirement. It should explain to us what we *experience* through our consciousness; and, since this (consciousness) knows nothing of what we have so far been assuming – quantities and neurons – it should explain this lack of knowledge as well.

We at once become clear about a postulate which has been guiding us up to now. Before we got to this problem of quality, we were treating mental processes as something that could dispense with experience through consciousness, as something which exists independently of such experience. We are prepared to find that some of our quantitative inferences are not confirmed through consciousness. If we do not let ourselves be confused on that account, it follows, from the postulate of consciousness providing neither complete nor trustworthy knowledge of our own information processing, that information processing (and therefore mental processing) is in the first instance to be regarded to its whole extent as unconscious and is to be inferred like other natural things.

In that case, however, a place has to be found for the content of consciousness in our quantitative ψ processes. Consciousness gives us what are called qualia – *qualities* – phenomena which are different in a great multiplicity of ways and whose *difference* is distinguished according to its relations with the world beyond the blanket. Within this difference there are series, similarities and so on, but there are in fact no manifest quantities in it. It may be asked *how* qualities originate and *where* qualities originate. These are questions which call for the most careful examination and which can only be treated roughly here.

Where do qualities originate? Not in the external world. For, out there, according to the view of our natural science, to which psychology too must be subjected, there is only information and nothing else.[38] In the *e* system perhaps? That tallies with the fact that qualia are linked with the affective and other sensory modalities, but it is contradicted by everything that (wrongly, as it turns out) argues in favor of the seat of consciousness being in the *upper* storeys of the nervous system, in the telencephalon. In the ψ system then. Against this, however, there is a weighty objection. During conscious perception the *e* and ψ systems are in operation together; but there is a broader class of mental process which is no doubt performed exclusively in ψ – cognition in general – and this, as we know, is for the most part without quality. Most cognition brings about nothing that has the peculiar quality of conscious perception.[39] Thus we summon up courage to assume that there is a third system of neurons – ω perhaps we might call it – which is excited along with felt affect and conscious perception but not with cognition in general, and whose contribution to information processing gives rise to the various qualities – that is to say, to conscious experience.

If we keep firmly to the fact that our consciousness furnishes only *qualities*, whereas science recognizes only *quantities*, a characterisation of the ω neurons emerges, as though by rule of three. For whereas natural science (including information science) has set about the task of tracing all the qualities of our experience back to quantities, it is to be expected from the structure of the nervous system that it contains contrivances for *transforming quantity into quality*; and here the original trend to minimize *F* must triumph once more. The Markov blanket was a screen that would only allow external states (*Q*) to influence internal states via its sensory states (φ). The ψ system was already protecting the organism against entropic forces by meeting these φ states with outgoing predictions. This enabled it to act on the external world (via the blanket's *M* states) in service of the primary function. It had to deal only with updating its representational model ($Q\eta$) on the basis of *e* processing. As a further step, it is to be suspected that the system ω, which is located deeper within the brain than ψ, is moved by a further negentropic function. It would seem as though the characteristic of quality (that is, conscious experience) comes about in connection with the predictive processing of φ via ψ. Now it dawns on us that error cannot be got rid of by minimizing the difference between $\psi(\varphi)$ and φ through goal-directed *M* and $Q\eta$ updating alone – that is, through action and perception alone.

At this point, however, we are met by what seems to be an immense difficulty. We have seen that resistance opposes *e* and that the ψ neurons are the vehicles of this resistance. If the ω neurons support the ψ function, then they too must be resistant – perhaps even more resistant than ψ. But that is a characteristic that we cannot grant to the vehicles of consciousness. The mutability of their content, the transitoriness of consciousness, the easy linking of qualities simultaneously felt – all of this tallies only with complete permeability of the ω neurons, together with total restoration of their former state following the passage of an *e* signal. The ω neurons behave like sensory receptors or like the error units themselves, and in them we could find no place for a memory. Permeability, then, complete transitoriness, which does not arise from resistance. From where else can it arise?

I can see only one way out of the difficulty: a revision of our previous hypothesis about how *e* (and therefore *F*) is minimized. So far, I have regarded the only alternatives as *action* and *perception* (i.e. representation in Brentano's sense)[40] – which are fundamentally *unconscious* processes. But there must be still another alternative, also of a predictive nature; and the mechanics of statistical physics have allowed another characteristic of information to be measured. We speak of this as *precision* for short. Precision is a measure of *confidence*. Thus I shall assume that an error signal (the difference between a predicted distribution and the sample actually obtained) is assessed by the brain not only by comparing the means of its predicted and actual sensory distributions but also the *variation* about those means. A large amount of variation in a φ sample (i.e. low precision) makes the system less confident about the fit. Judgements of difference between an outgoing prediction and an incoming error signal are easier to make when the distribution is narrow and *precise*. This is indeed a transient variable, but it enjoys a significant degree of control; for error signals which are assigned low precision will be sequestered in the sensory epithelium – such as occurs in sleep, for example. *The ω values therefore determine the magnitude of e and the generalizability of ψ*. In other words, precision (confidence) determines which *e* signals will be *aroused*, which will be selected for ongoing processing, and which will be *resisted*. So, quality is quantified by the system via the measurement of its confidence.

Here very much remains to be done in the way of functional-anatomical clarification, for in neurophysiology, too, the general laws of information must apply without contradiction. The hypothesis on offer here assumes that the *ω* neurons are capable of evaluating *e* by means of precision modulation; they must *prioritize* the various *e* contributions to *F* on a *categorical* (i.e. qualitative) and *contextual* basis. This prioritization (or salience-measuring) function – which is performed, as we know, by the midbrain decision triangle – now emerges as *the fundamental basis of consciousness*.[41] The *ψ* units, too, have their precisions; but that system is without quality whenever its *ω* values remain *monotonous* – when confidence doesn't change – that is, during stereotyped, automatized behaviors. I repeat: the expected precisions assigned by the system represent the confidence of the system. Deviations from the expected precisions of *ψ* units (which necessarily vary inversely with those of the corresponding *e* units)[42] then become *the quantitative basis of ψ quality* – which in turn licenses voluntary behavior, the consequences of which are fed back to the midbrain decision triangle in a circular fashion. This implies that *changes* in the expected *ω* values in the prioritized categories of *ψ* and *e* (i.e. *deviations from expected precisions* in the billions of little homeostats that make up the cathected predictive hierarchy) during the unfolding of an *expected context* generated in *ψ* (which flows from the selected action cycle) must be *palpated* by the *ω* system, and the precision values (the system's confidence) must be *adjusted* accordingly – in line with the actual amplitude of incoming *e* signals. This enables the ongoing evaluation of "what to do next." This palpating process (i.e. active modulation of prioritized *ψ* and *e* precisions) is the mechanistic basis of "bound" cathexis – that is, *reconsolidation* in working memory.[43] This cathectic process of reconsolidation is *predictive work in progress*, which we now recognize is the most essential work of the secondary process.[44] The modulatory nature of this precision-weighting process leads us to its anatomical localization: baseline and adjusted *ω* values of *ψ* must be set by the brainstem *reticular activating system*, since what information scientists call precision is what neurophysiologists call "arousal."[45] We now discover that the capacity for arousal (i.e. modulation of synaptic transmission, also known as "post-synaptic gain") is what distinguishes the basic physiology of *ω* system neurons from that of *ψ* and *e* neurons. This distinctive capacity – precision weighting, confidence, *which is calibrated in direct relation to the feelings selected by the PAG* – is applied to the expanded telencephalon by primitive brainstem structures. Consciousness is not, therefore, after all, a function of the upper storeys of the brain.[46]

If conscious experience is quantified by modulating the system's confidence in incoming and outgoing signals, where do the *differences* between its various qualities spring from? Affects provide the simplest answer. Their qualities vary in two respects. As we have seen, the *valence* (pleasure-unpleasure) of an affect is determined by the degree and direction of change in its homeostatic deviation, represented by the residual *e* values converging on the PAG following each action cycle. *Increasing confidence in a prediction* as to how homeostasis may be regained (which implies decreasing confidence in the corresponding error signals) is "good" and the converse is "bad." But recall that *consciousness* of such valence arises only in the prioritized (i.e. the currently most salient) category of affect, whose priority was determined by midbrain evaluation of all the categories of residual *e*. The *ω* values of whole *categories* of affect was assigned accordingly. Thereafter expected *ω* values *within* each category (within the expected context that flows from it) are assigned to the relevant *ψ* units throughout the predictive hierarchy, in accordance with the expected *φ* consequences of each *M*. The expected precisions within the non-prioritized categories will be assigned monotonous *ω* values, yielding *automatized* action sequences, while those within the prioritized category (where the most salient uncertainty lies) must be palpated and adjusted during the action sequence, yielding *voluntary* behavior. Exteroceptive qualia require a more complex answer. They, too, are categorized qualitatively, with each whole modality possessing its own phenomenal "feel," but the modalities of perception (unlike those of affect) typically come to consciousness simultaneously. It is true that the relative *ω* values of these categories, like the affective ones, are weighted differently depending upon contextual factors (e.g. vision is assigned more precision by day and hearing by night) but here the assignment of *ω* values (the confidence) within each category during the action cycle is determined by the prevailing *affect*. That is, the salience of external events within *all* the modalities depends upon the exigencies of life; so that, for example, sights and sounds *of caregivers* are prioritized during PANIC/GRIEF episodes, whereas those *of predators* are prioritized during FEAR episodes. The priorities are not determined by intramodal events. Thus, exteroceptive qualia come to consciousness secondarily, *relative* to their affective salience. Consciousness of perception merely *contextualizes* affect, using the common currency of felt uncertainty. For the mechanical reasons outlined previously, endogenous *e* signals always take priority over external ones – but now we learn that the *ω* values assigned to external *e* signals *determines* their uncertainty. This applies to all the billions of little homeostats, embedded within each other over the levels of the exteroceptive hierarchy. Increasing confidence in the predicted *φ* consequences of an action is "good" (more certain, less salient) while decreasing confidence is "bad" (less certain, more salient) relative to the prevailing affective valence; so that when things turn out as expected (with "satiation"), monotonous *ω* (and therefore automaticity) prevails in the action cycles generated by non-prioritized needs. It will be noted that "goodness" and "badness" (in both affective and cognitive consciousness) is an essential precondition for choice. *Choices can only be made if they are tied to a value system*. Here I must emphasize once more that precision (or confidence) is the ultimate mechanism whereby certainty/uncertainty is determined. The precision of an *e* signal will determine its centripetal reach through the predictive hierarchy (its magnitude), and therefore its plastic influence upon the generative model (its generalizability). Learning is therefore governed by feeling, by what may be called the law of affect.[47] Consciousness, therefore – all of it, both affective and cognitive (including perceptual consciousness) – is felt uncertainty.[48] What determines the differences between the varieties of phenomenal quality is what the uncertainty is *about*. That is why, for example, light waves feel different from sound waves.

(8) *Consciousness*

It is only by means of such complicated and far from perspicuous hypotheses that I have hitherto succeeded in introducing the phenomena of consciousness into the structure of quantitative science. Now an attempt must be made to explain how it is that modulatory processes in the *ω* neurons bring consciousness along with them. It is not only a question of establishing a *correlation* between the characteristics of consciousness that are known to us and processes in the *ω* neurons which vary in parallel with them (that is an "easy" problem). Addressing the "hard" (causal) problem is, however, now quite possible in some detail.[49]

A word on the relation of this theory of consciousness to others. According to an advanced mechanistic theory, consciousness is a mere appendage to physiologico-chemical processes and its omission would make no alteration in the causal passage of events. According to another theory, consciousness is the subjective side of all neuropsychological events and is thus inseparable from the causal physiological mental process. The theory developed here lies between these two. Here consciousness is the subjective side of one part of the information processing in the nervous system, namely of the *ω* processing; and the omission of consciousness does not leave the causal chain unaltered but involves the omission of the contribution from *ω*.

If we represent consciousness as the action of *ω* units, several consequences follow. We have posited that these neurons are located in the region of the meta-homeostatic control center of the entire system, in the mesencephalic decision triangle and reticular activating system, which lie beneath the deepest layers of the telencephalic hierarchy and even beneath the diencephalic hypothalamus. This control center is the ultimate receiver of self-system information, the ultimate asker of questions to which the world and body provide answers. This implies that precision-weighting is the system's deepest control mechanism: it determines which predictions and errors will be aroused, and therefore how the system will represent and act upon the world. This in turn implies that *precision optimization* is the causal basis of all intentional behavior accompanied by choice; that is, of any self-organizing system that is capable of voluntary action. It is small wonder, then, that Panksepp called this mechanism the primal SELF (which he conceptualized physiologically, but which I am quantifying here in terms of its statistical mechanics). This designation suggests what the original biological value of conscious experience was.

So far, we have given an incomplete account of what consciousness is. Besides the property of intentionality just described, self-organizing systems possess *selfhood*. This not only explains their intentionality, it also implies a *point of view*, arising from the formation of a Markov blanket which brings the system into being in contradistinction to the not-system. Now I must ask you to cross a Rubicon with me. I must ask you to consider the mechanism of precision optimization, which I have just outlined in objective terms, *from the subjective viewpoint of the system.* I must ask you to *empathize* with the system. This is a perspective that most natural scientists are reluctant to take, much to the detriment of science, and especially of mental science; but it is a perspective that is justified precisely by the selfhood of the system. Conscious experience can only be registered from the subjective point of view. To rule that viewpoint out of science is therefore to exclude consciousness from science (as many people do). To be clear: free energy and its constituent precisions are only experienceable within a system when it is subjectively conceived, from the viewpoint of the system; experiences cannot be observed *as experiences* from without, objectively. But the existential mechanics of biological self-organizing systems like you and me *obliges* us to ask questions about our own states in relation to our ambient surrounds. We must chronically ask, "What will happen to my free energy if I do that?" Moreover, we must ask this question in relation to multiple categorical variables; so, the answers – our vital statistics – must be both quantified (as continuous variables) and qualified (as categorical variables). What I am describing here in technical terms is nothing too complicated, in fact. You know it from your personal experience. What you experience all the time is fluctuating pulses of feeling in response to your movements through the world, as you check whether everything is as you expected to find it – and as you try to close the gaps, somehow, when it isn't.

The aptitudes of intentionality and subjectivity of complex self-organizing systems equipped with precision optimization leads to the following major conclusion. The "equipment-evoked responses" (in Wheeler's sense) that flow from the types of question that systems like us are obliged to ask must include existential values and multiple qualities. Our *confidence* in the fluctuating signals – the ω "phenomena" (in Wheeler's sense) – that systems like us register must be valenced, qualified and subjective. And that *just is* what it is like to experience consciously. The equipment-evoked responses, in the case of us vertebrates at least, are feelings.

(9) *The functioning of the apparatus*

It is now possible to formalize our picture of the functioning of the apparatus constituted by the dynamics between the quantities I have described.[50] Q plays no part in the picture, since its values are hidden from the system. That means, in what follows, we have a self-contained, autonomous description of mental dynamics in terms of the system's own internal ($Q\eta$, ψ, ω, e) states and Markov blanket (φ, M) states.

Equipped with these terms, we can formalize a complex self-organizing system's dynamics in relation to precision optimization. I will start with two equations which define variational free energy. The first equation says that "free energy is (approximately) the negative logarithm of the probability of encountering some actively authored sensory states":

$$F \approx -logP(\varphi(M))$$

where P denotes probability. The second equation says that "the expected free energy decreases in (approximate) proportion to negative log precision":

$$E[F] \approx E[-logP(\varphi)] = H[P(\varphi)] = -{}^{1}\!/_{2} \cdot log(|\omega|)$$

where $E[\cdot]$ denotes expectation or averaging and H denotes entropy, under Gaussian assumptions about random fluctuation.

Remember that the whole point of the self-organizing system's dynamics is to minimize free energy. With these relationships in place, it becomes evident that there are three ways for the system to reduce prediction error and thereby minimize F:

(1) It can *act* (i.e. change M) to alter sensations (φ) so that they match the system's predictions. This is action.
(2) It can *change its representation of the world* ($Q\eta$) to produce a better prediction (ψ). This is perception.
(3) It can *adjust precision* (ω) to optimally match the amplitude of the incoming prediction errors (e). This is consciousness.

Here are the equations for these three alternatives:

$$\frac{\partial}{\partial t}M = -\frac{\partial F}{\partial M} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial M} = \frac{\partial \varphi}{\partial M} \cdot \omega \cdot e \qquad (1)$$

$$\frac{\partial}{\partial t}Q\eta = -\frac{\partial F}{\partial Q\eta} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial Q\eta} = -\frac{\partial \psi}{\partial Q\eta} \cdot \omega \cdot e \qquad (2)$$

$$\frac{\partial}{\partial t}\omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2} \cdot (\omega^{-1} - e \cdot e) \qquad (3)$$

where ∂ denotes a partial derivative and t denotes time, and prediction error and free energy are quantified as follows:[51]

$$e = \varphi(M) - \psi(Q\eta)$$

$$F = \frac{1}{2} \cdot (e \cdot \omega \cdot e - \log(\omega))$$

Since the third of the numbered equations formalizes the precision optimization process that I associate with the evaluation of free energy which underpins conscious experience, I will translate only that one into words:

> The rate of change of precision (ω) over time depends on how much free energy (F) changes when you change precision. This means that precision will look as if it is trying to minimize free energy.[52] The rate of this free energy minimization process is the difference

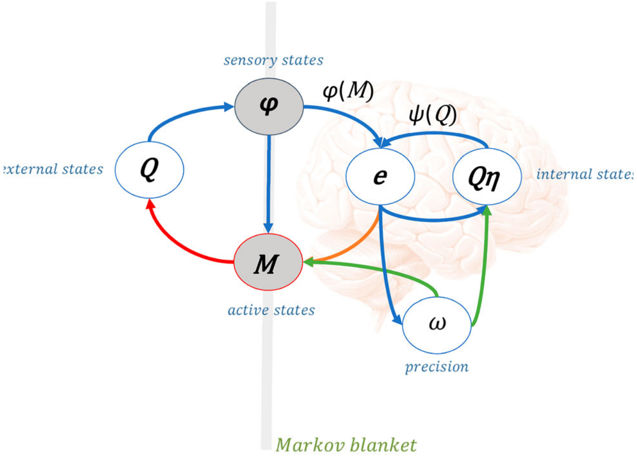

Figure 1.

> between the variance (the inverse precision) and the sum of squared prediction errors (*e.e*).[53]

In more basic terms, the third equation quantifies how *ongoing adjustment of precision implements Friston's law*, alongside action and perception. In short, it quantifies how *consciousness contributes to action, perception and model updating* – and thereby to minimizing free energy.

Figure 1 represents these dynamics visually.

(10) *The ψ paths of transmission*

Since *Qη* receives *e* from both exogenous and endogenous sources, it is preferable that our conception of the telencephalic cerebrum should be divided into its two components: the outer *pallium* layers of *ψ* neurons which comprise the cerebral cortex, and its *nucleus* known as the subcortical basal ganglia.[54] The nucleus is connected with the myriad paths by which endogenous quantities of *e* enter *ψ*. Without excluding the possibility of these paths being connected with *φ* in the case of "sensory" (and perhaps "emotional") affects, we must nevertheless hold to our assumption that a direct pathway leads from the internal body to the nuclear *ψ* neurons, at least in the case of "homeostatic" bodily affects.[55] If this is so, however, *ψ* is exposed to *e* signals at its nuclear side without the protection of the graded predictive hierarchy, and in this fact lies the mainspring of the mental mechanism.

What we know of endogenous needs is that they are of a phenotypic nature, that they arise continuously, and yet only periodically become *mental* stimuli. That is, endogenous needs only periodically become "drives" which make demands upon the mind for work. We cannot avoid the idea that there are core homeostatic thresholds; and the intermittent character of their mental effect necessitates the view that on their paths of transmission to *ψ* the endogenous *e* signals, too, come up against resistances which are overcome only when these thresholds are reached. There are therefore paths of interoceptive transmission made up of multiple segments, having a number of reflexes (which trigger phenotypically determined *M* states) interpolated between them along the way to the *ψ* nucleus. The diencephalic hypothalamus, discussed previously, apparently represents the highest of the barriers. Above these core homeostatic thresholds, endogenous needs act as a drive stimulus *continuously*, and every increase of *e* is registered as an increase of the *ψ* stimulus – that is, as a demand for mental work. It follows, therefore, that there is a state in which the paths of endogenous need transmission become permeable. Experience shows, further, that, after a drive stimulus has been processed by *ψ*, the reflex that processes the underlying need resumes its normal resistance to change. That is, updating of *Qη* predictions through learning *supplements* the organism's reflex response but never *extinguishes* it. That is because reflex (and instinctual) predictions are not "memories"; they are not part of *ψ* and are therefore essentially impervious to learning.

A process of this kind is called *autonomic*. The *ψ* mode of processing *e* is triggered when the autonomic mode reaches its limits. It is evident that there is a further, ultimate line of protection – namely *ω* modulation at the level of the midbrain decision triangle and reticular activating system – which is theoretically capable of preventing bodily needs from attaining consciousness even when they breach the thresholds, thereby preventing them from entering *ψ* (consider what I said about sleep). But the consequences of this for *F* minimization are potentially catastrophic due to the phenotypic exigencies described before.[56] That is why strong bodily needs eventually win the midbrain decision contest; they are routinely assigned (indeed, they grab) high *ω* values which tally with their inexorable magnitudes. Reflex action is also known to occur in the *φ* paths of conduction – for instance, in the case of pain as discussed previously. However, the lesser part played by reflexes on the exteroceptive side of *ψ* speaks in favor of the view that greater degrees of uncertainty are tolerable in relation to exogenous *e*. Still, reflex resistances do operate even at the peripheral extremity of the sensory receptor organs.

It was a very noteworthy fact that some *ψ* neurons are able to maintain a position between the characteristics of permeability and impermeability, since some are permanently altered by the passage of *e* and some are not. This did not contradict the property we assumed the *ψ* neurons to possess of being permanently altered by a current of *e*. That apparent contradiction was already resolved, as follows.

The restoration of *ψ* resistance after an *e* signal has ceased is progressively less likely to occur through the predictive hierarchy from the surface to its depths. Towards the periphery of the hierarchy, *ψ* synaptic weightings are altered on a short-term basis. Thereafter they are consolidated into the episodic and semantic (long-term) memory systems. Each of these systems is less plastic than the preceding one, which means they are progressively more resistant to updating. We are now in a position to localize the short-term and *declarative* long-term memory systems in the *ψ* pallium. The *nondeclarative* systems – which are still more resistant – are accordingly (for the most part) to be located in the *ψ* nucleus. Combining these ideas with the notion of autonomic thresholds, we arrive at the following formulation. Due to the magnitude and relentless nature of endogenous *e* signals, they eventually achieve *ω* weightings that breach the threshold into *ψ* (into cognition). Now, however, considering that these signals enter it from the side of the homeostatic core, the possibility arises that they will spread no further than the nondeclarative nucleus. The implication of this is that, although a prioritized need will achieve *affective* consciousness, it will not necessarily gain access to the pallium layers of *ψ* and therefore to the predictive systems which permit conscious *cognition* and action. The affect (the feeling of the need) can then only be resolved through involuntary action programs.

It follows from this that such error signals will possess the curious status of being simultaneously conscious (as affect) and unconscious (as cognition). In other words, the mental work demanded by the drive will be *felt*, but the work will not take place in working memory – it will not enjoy conscious *representation*. As it happens, this is a common method by which nondeclarative learning takes place: procedural memories (habits and skills) are consolidated through brute repetition, guided by the law of affect but not by deliberative thinking; and emotional responses (e.g. fear conditioning) readily occur non-declaratively, even upon single exposures to painful events. In fact, this is the *only* way that long-term consolidation takes place during the first few years of life, before the declarative pallium matures.[57] Much remains to be explained about why this separation of affects from ideas might occur in other circumstances. We will return to this. It is already clear, however, that nondeclarative predicting is not the most effective method in complex situations where accuracy is required. This applies specially to learning how to meet *emotional* needs, which, both individually and collectively, bring considerable uncertainty in their wake. That appears to be why the cortical pallium, which specializes in uncertainty and complexity, evolved in the first place. The *ψ* nucleus is at the mercy of endogenous *e*, and it is thus that there arises the impulsion which sustains all mental activity and development. Philosophers know this power as the "will" – the derivative of what we call drive.

(11) *The experience of satisfaction*

Breaching the resistance of the nuclear *ψ* system – by high-magnitude *e* signals arising either from the endogenous core or from exogenous *φ* – will have as their result an effort to discharge *F*, an *urgency* which activates *M* paths in an attempt to lower the *ω* of the *e* signal. There are two classes of "motor" (*M*) neuron, because the meta-homeostatic control center must contend with two sets of *Q*, both of which are "external" to the Markov blanket. Experience shows that here the first path to be taken is that leading to *internal change* (the firing of autonomic reflexes, leading for example to changes in blood pressure). As was explained at the beginning, such action only produces satisfactory results below homeostatic thresholds; but still internal changes continue after a signal breaches the *ψ* nucleus. The reduction of *e* precision is also made possible by another intervention which (for the time being) manages increased *F* – through the activation of *M* states that lead to what is loosely called "expression of the emotions" (crying, screaming, etc.). This intervention calls for an alteration in the external world (supply of nourishment, proximity of the sexual object) which, as a specific action, can only be brought about in definite ways. At first, the human organism is incapable of bringing about the specific action. It takes place by *extraneous help*, when the attention of an experienced person is drawn to the child's state by discharge along the path of internal change. In this way the expression of emotion acquires a secondary function of the highest importance, that of *communication*; and the initial helplessness of suckling mammals is the origin of the drive to CARE. Instinctual life is the primal source of all moral and ethical motives.

When the helpful person has performed the work of the specific action in the external world for the helpless one, the latter is in a position by means of vegetative reflexes immediately to carry out in the interior of the body the activity necessary for removing the endogenous stimulus. The total event then constitutes an *experience of satisfaction*, which has the most radical results for the development of the individual's functions. For three things occur: (1) In the homeostatic core, a waning of the endogenous *e* signal is effected through a gradual reduction of its ω value until it reaches zero, so the urgency which had produced unpleasure is brought to an end (i.e. confidence in the phenotypic prediction is restored).[58] (2) In the ψ pallium, an *exteroceptive* prediction is encoded, namely: active state *Ma* (crying and screaming) causes sensory state φa (caregiver performs the specific action). (3) In the ψ nucleus, a further prediction is encoded, namely: active state *Ma* resolves endogenous need α.

Cognitive scientists all too readily forget that the major needs, too, are "sensory" states – and therefore overlook the fact that *endogenous* sources of *F* are the mainspring of mental life. We only represent the outside world cognitively because we *must*; it is where our needs must be met. That is why we described the cerebrum as a sympathetic ganglion; it functions *in sympathy* with the needs of the organism. We must recall also that complex organisms like us have multiple needs; so this is a simplified account. For example, with frequent repetition of the satisfaction of need α (the need for nourishment), yet another prediction will be consolidated in the ψ nucleus, namely: sensory object a (the caregiver) satisfies endogenous need β (the need for attachment).[59] A cascade of parallel predictions is thereby established through the hierarchy, from the homeostatic core itself through the automatized ψ nucleus to the pallium.

We now become acquainted with a different aspect of the predictive hierarchy. In consequence of a phenotypic belief concerning nourishment (in our model example), a core autonomic prediction that "energy balance will be maintained" triggers the bodily affect of hunger, which predicts the basic emotion of SEEKING, which predicts the procedural habit of distress vocalization, which predicts the semantic belief about caregivers, which predicts the episodic memory that good object a will perform the specific action, which predicts the here-and-now sensory state φa. Each prediction in this cascade is less generalizable – both spatially and temporally – than the one preceding it. That is because the internal milieu of the body is a far simpler place than the external object world; as the predictive cascade is traversed outwards, the deeper predictions risk decreasing accuracy due to the increasing demand for complexity (i.e. for spatial and temporal [contextual] specificity).

Whenever the state of urgency re-appears, this cascade will be activated; it represents the first *expected context* when that particular need (for nourishment) is prioritized as felt hunger. This state of affairs can be described as *wishful activation*. Each level in the predictive hierarchy cathects the level beyond it, by self-generating the incoming signal it expects to occur there. This necessarily entails the assignment of baseline *levels of confidence* at each layer of the anticipated scenario. (Predicting ω can be learnt like anything else.) From the core to the periphery, however, the baselines will be progressively fainter echoes of the core homeostatic settling point. The predictions must become progressively less precise because there is progressively more variation in self-generated signals cathected from the depths of the self outwards to the uncertain world.

I do not doubt that in the first instance this wishful activation will produce the same thing as a perception – namely an (unconscious) "hallucination." If reflex action is thereupon introduced, however, disappointment cannot fail to occur, as we shall see.

(12) *The experience of pain*

Normally, the ψ nucleus is exposed to *e* from endogenous paths of excitation, and abnormally, even though not yet pathologically, in cases where excessively strong *e* signals break through the screening contrivances in the pallium – such as in the case of *pain*. Pain gives rise in ψ (1) to a large rise in surprisal, which is felt as unpleasure in ω, (2) to an inclination to discharge *F*, which can be modified in certain directions, and (3) to a predictive facilitation between this *M* inclination and a $\psi(\varphi)$ representation of the noxious object which caused the pain. Moreover, there is no question but that pain has a peculiar sensory *quality*, which makes itself felt as a category of exteroceptive sensation *and* one of endogenous unpleasure.

If the representation of the noxious object is freshly cathected in some way – for instance, by a fresh perception – an anticipatory state (FEAR) arises which is not pain but which nevertheless has a strong resemblance to it. As with wishful cathexis, it activates a predictive cascade that in this case includes, alongside the ψ representation of the noxious object, the unpleasure and the inclination to discharge *F* that corresponds to the experience of pain. Since FEAR (like all unpleasure) signifies rising endogenous *e*, it must be asked where this unpleasure comes from. In the actual experience of pain, it is often assumed that the external stimulus itself transmits the unpleasure; but the reproduction of the same experience in imagination – which is distinguished from perception precisely by the lack of any external stimulus – makes clear that the operative *e* signal must be *endogenous* in both cases. Affects are always endogenous; they are states of the subject, not of the outside world. This raises a question: in what, then, does the magnitude of an exteroceptive *e* signal consist, since it was this that enabled it to reach the ψ nucleus? The answer is: it consists in *arousal* (selection) of the signal *by assigning it a high ω value*. The same applies to the endogenous *e* that is aroused along with it. Rising expected *F* is rising uncertainty concerning a core homeostatic belief (viz., "I shall not suffer tissue damage") which is felt as unpleasure. Now we see how a self-generated ψ signal – the $\psi(\varphi)$ cathexis – can have the same attribute as exteroceptive pain: all ω is modulated from the interior of the system.

It remains to add, from the physiological point of view, that modulatory arousal consists not only in the release of psychoactive molecules from the reticular activating system (e.g. noradrenaline in the case of pain) but also from the *interior of the body*. The mechanism of this release is well understood. Just as there are motor neurons which, when activated, stimulate the muscles, so there are "secretory" neurons which cause the release in the interior of the body of endocrinological molecules (e.g. hormones and peptides, of which there are a great number) that operate as stimuli upon the endogenous paths of transmission to the homeostatic core and to ψ – motor neurons which thus influence the production of endogenous *e* and supply it in roundabout ways. Evidently these key hypothalamic neurons are activated when certain levels of resistance in the ψ nucleus have been breached. As a result of the experience of pain (which challenges a core homeostatic belief) the representation of the noxious object has acquired an excellent facilitation to the core neurons, in virtue of which unpleasure is now released in the *expectation* of pain.

Support for this formulation is to be found in our knowledge of the hypothalamic–pituitary–adrenal (HPA) axis, which releases cortisol in the example just discussed, but the same principles apply with the hypothalamic–pituitary–gonadal axis (HPG), hypothalamic–pituitary–thyroid axis (HPT) and hypothalamic–neurohypophyseal systems, in the cases of other affects.

(13) *Distressing and wishful states*

The residues of these two kinds of experience (pain and satisfaction) which we have been discussing are distressing and wishful *predictive* states. These have in common the fact that they both involve a raising of surprisal in the ψ nucleus – brought about by changes in core homeostatic $\omega(e)$ values, which increase in the case of distress and decrease in the case of wish. Both states are of the greatest importance for ψ, for they leave behind them motives which are of a compulsive kind. The wishful state results in an *attraction* towards the object wished-for, or, more exactly, towards its $Q\eta$ representation; the experience of pain leads to a *repulsion*, a disinclination to keeping the noxious representation cathected. All of this flows from the law of affect – learning by trial and error. Here we have the primary basis for wishful *approach* and defensive *avoidance*.

Wishful approach can easily be explained by the assumption that the cathexis of the attractive representation (the image of a "good" object) when the subject is in a state of desire – that is, SEEKING – greatly exceeds in baseline ω the cathexis which occurs when there is a random perception, so that a particularly strong facilitation leads from the ψ nucleus to the corresponding predictive units of the pallium.

It is harder to explain defensive avoidance – which, as we shall see, provides a template for "defence" in general – the fact that a repulsive representation (the image of a "bad" object) is regularly abandoned by its cathexis as soon as possible. Nevertheless, the explanation should lie in the fact that the primary experience of pain was brought to an end by reflex withdrawal. The emergence of *another object in place of the noxious one* was the signal for the fact that the experience of pain was at an end, and the ψ system, taught biologically through the law of affect, seeks to reproduce the state in ψ which marked the cessation of pain. With the expression "taught biologically" I am referring to a basis of explanation which should have independent validity, even though it does not exclude, but rather calls for, a recourse to statistical-mechanical principles (quantification). In the instance before us it may well be the increase of uncertainty (increasing precision in expected error) which occurs with the cathexis of a noxious prediction that forces increased confidence in an *M* discharge and thus a decreased precision in the predictive cathexis of the noxious representation as well.

(14) *Introduction of the "ego"*

In fact, however, with the hypothesis of "wishful approach" and the inclination to "defensive avoidance" we have already touched on a state of ψ which has not yet been discussed. For these two processes indicate that an *organization* has been formed in ψ whose presence interferes with passages of *e* which on the first occasion occurred in a particular way (i.e. accompanied by satisfaction or pain). This *unconscious* organization is called the "ego." It is important to notice that the ego's ψ organization, as opposed to its selective cathexis, serves automatic action programs and therefore the primary function. It can easily be depicted if we consider that the regular reception of endogenous *e* in certain neurons of the ψ nucleus and the predictive effect proceeding thence will produce a repetition of previously learned paths to satisfaction. A driving endogenous *e* signal which breaks into the ψ system will automatically proceed in the direction of the predictions with the largest certainty (greatest resistance) and will set up a freely mobile cascade in that direction. A *compulsion to repeat* will thus be triggered by fresh exogenous *e* signals of high magnitude.

While it must be the endeavor of the ego to give off rising *F* by the method of satisfaction, this cannot happen in any other way than by its influencing the repetition of experiences of pain and of distress, and by the following method, which is described generally as *inhibition*.

To put this more fully: an endogenous *e* signal (or an exogenous one which was sufficiently strong to have reached the ψ nucleus, and therefore to have aroused affective *e*) will trigger a predictive cascade that will divide up in the direction of the pathways that provide the greatest probability of suppressing (explaining away) the incoming *e*. This typically requires motor action. The *M* course taken is dependent both on *e* and the relative distribution of the ψ predictions – predictions that were consolidated over a lifetime by the law of affect. This approach favors pathways with the greatest certainty – pathways which have proven to minimize *F* in the *average* situation. This yields stereotyped behaviors which do not serve the organism equally well in all contexts. The inevitable consequence is surprising experiences, which, if the organism survives them, may prompt further learning by trial and error. This is biologically expensive in almost every way, including the fact that it comes at the cost of ever-increasing *complexity* – a cost that ultimately overwhelms the capacity constraints of ψ. We have, therefore, already come to know a third powerful factor, which was the factor that prioritized the driving category of endogenous *e* in the first place. This factor is ω optimization. The selection (prioritization) of an affect automatically generates an *expected context* within which the currently most salient need would be resolved. This expected context, too, implicates ω, since expected precisions must be predicted like everything else. Now we come to the enormous advantage that consciousness bestows. Although ω greatly reduces computational complexity through need prioritization, compartmentalization and precise predicting of sensory-motor contexts, this does not explain why the ensuing action sequence must be felt. The reason it must be felt is because the degrees of freedom at the periphery of the hierarchy are so large that ψ is readily overwhelmed by the combinational explosion. It is almost impossible to predict the present in every detail. The system's solution, we have seen, is to *feel* its way through it: to *palpate* the incoming error signals as they occur, *within the prioritised category*, and register *deviations* from the expected precisions. This underwrites choice and is the causal contribution of consciousness. The guiding force by which exteroceptive precisions are adjusted is by the feelings that the unfolding errors evoke in the PAG, throughout the action sequence. It is on this basis that the system's confidence in the consolidated predictive pathways that would otherwise have been used is *decreased*. This enables the system to *temporarily* facilitate alternative pathways: to up-regulate *their* precision. This might be called "side cathexis": *inhibition* of the automatized pathways. This leads to spreading activation of predictive cathexes, which allows qualitative palpating of a wider range of *e* signals. Without inhibition of the primary processes, voluntary behavior – and both biologically and computationally expedient learning – would be impossible. This provides a deep mechanistic account of the secondary process: the basis of the ability to tolerate *F* is ω optimization. Later we shall see that this mechanism underwrites *thinking* too.

The ego is to be defined as the totality of the ψ cathexes, at the given time, in which a *monotonous* ω component is distinguished from a *changing* and temporary one.[60] It is easy to see that changing cathexis interacts with the notion of graded resistance between layers of ψ neurons and with their orthogonal compartmentalization into qualitative categories, both of which are a part of the ego's possessions. They represent cathectic *possibilities*, if the ego is altered, for determining its extent in the next few moments.

Two further mechanisms now come into view, and they, too, are pivotal. These are called *repression* and *defence*. The ψ nucleus is regularly bombarded by endogenous *e*. These errors when prioritized by ω are felt as affects but not as ideas – unless and until they proceed from the ψ nucleus to the pallium. On the mechanism just outlined, this means that only *inhibited* ψ processes would typically become *cognitively* conscious – "declarable" – for the automatized processes proceed entirely according to the expected precisions, which must be monotonous. Therefore, when, for whatever reason, the ego does not tolerate a particular automatized prediction becoming inhibited through side cathexis, which means it does not tolerate it being *held in mind* (in working memory), that prediction will remain in a state of non-declarative certainty. This is what happens with "repression." Repressed predictions, then, constitute that portion of ψ which remains or becomes *illegitimately certain*. Predictions that do not hit the mark (that do not meet an emotional need *in reality*) inevitably give rise to repeated error signals. Such predictions *should* be updated. If this does not occur, the ego suffers feelings and *does not know why*. Its only recourse, then, in the case of repressed predictions, is to explain away the affect by means of declarative cognitive work concerning *other* predictions – which falls short of updating the actual cause of the unpleasure. This secondary process, which follows from the inevitable failure of repressed predictions, entails side cathexis not of the repressed cascade itself but of other predictive assemblages which can avert or mitigate its consequences. This is what happens with "defence," which is clearly not synonymous with repression.

Now a word on the highly interesting "resting state" of the ego.[61] We have said that the ψ nucleus is routinely bombarded from the diencephalic and mesencephalic core. When these *e* signals break through (become prioritized) they activate core predictions which run their course through the hierarchy. In all cases where there is lack of certainty, the predictive cascade should be progressively inhibited towards the periphery, enabling side cathexis (working memory) and voluntary behavior, and, thereby, expedient learning from experience. However, this does not occur with non-prioritized needs – which include the repressed ones by definition. These must be relegated to automatized predictive cascades (with monotonous ω), where inexpedient learning takes place – through high-magnitude exogenous errors which "grab" precision. The safer lessons learnt through working memory and voluntary behavior, by contrast, are fed back to ψ through graded consolidation. We have seen that only the most generalizable of these newly acquired predictions are consolidated as far as the nucleus, whereupon they become resistant to change.

Consolidation entails generalizability. We know already how this is ordinarily achieved, but we have seen also that the process can be very expensive. *Simplification* now emerges as the essential contribution of the resting state, which characterises the ego's mode of functioning when it is unoccupied by major exogenous sources of *e*. In this state, in the absence of pressing demands, the ω decision triangle prioritizes default-mode SEEKING – even, and perhaps especially (as we shall see), during sleep. This engages the "default mode network," which includes the ψ nucleus and extends into relatively deep layers of the pallium – mainly in the cortical midline – where declarative cognition can occur. In the default mode, SEEKING does what it always does: it engages uncertainty (e.g. novelty) with the aim of reducing it. This mode of cognition is also called "mind wandering" – it is a sort of internal foraging – whereby wide-ranging imaginary scenarios are palpated under safe conditions; that is, *virtually*. (These safe conditions are ensured through *φ and M deactivation*, as will be discussed further below.) Learning from *imaginary* experience results in ongoing consolidation: increasing the certainty of viable synaptic connections (i.e. latent predictions) and discarding redundant or weak ones. This "synaptic pruning" optimizes the ego's balance between complexity and accuracy in an ongoing fashion. This entails what we call *phantasy*.

(15) *Primary and secondary process in Ψ*

We return now to the functioning of the ego in its task-related modes. It follows from what has developed so far that the ego, which we can treat as regards its trends like the nervous system as a whole, will, when its task-related processes are uninfluenced by inhibition, be made helpless and suffer injury under two conditions.

This will happen in the first place if, while it is in a *wishful state*, it newly cathects the memory of a desired object and then acts upon it. In that case satisfaction must fail to occur, because the object is not real but is present only as a memory. Ψ is not in a position, to begin with, to make this distinction if it works on the basis of the predictive cascade described before. Thus it requires a criterion from elsewhere in order to decide between *perception* and *memory*.

Likewise, ψ is in need of an indication that will draw its attention to the re-cathexis of the memory of a *noxious* object – as opposed to the presence of a real one – and enable it to obviate, by means of side cathexis, the consequent release of unpleasure. If ψ is able to put this inhibition into operation soon enough, the release of unpleasure, and at the same time defence, will be slight; otherwise there will be immense unpleasure and excessive defence, as if the noxious thought were real.

Both wishful cathexis and unnecessary release of unpleasure, where a memory is cathected anew, can be biologically detrimental. This is true of wishful cathexis whenever it exceeds a certain amount and so acts as an enticement to action; and it is true of a release of unpleasure whenever a noxious cathexis results not from external events but only from ψ itself. Here, once again, it is a question of an indication to distinguish between a perception and a memory.

It is surely the ω neurons which furnish this indication: the *indication of reality*. We have seen that it is a misconception to assume that every perception is accompanied by conscious experience; quality is generated only on the basis of prioritized needs. The affect generated by need-prioritization renders normally unconscious perceptions (those with monotonous ω) *salient*. Now they must be palpated, and their ω values must be adjusted accordingly. That is made possible by *attention*, which now emerges as *the exogenous equivalent of endogenous* *need prioritization*. The system's confidence in a prediction (its ω value) is determined initially by the driving affect, which carries the twin risks we have been discussing. But perception is not the same as hallucination; it is *controlled* hallucination. Putative hallucinations are controlled by incoming error signals, leading to *reconsolidation* of the cathected predictions, as modulated by precision. This modulatory control – which happens throughout the layers of the hierarchy – is supplied by the ω values that are assigned to each incoming error signal (and its corresponding outgoing prediction). So, hallucination is prevented through *withdrawal of confidence* from predictions that attract strong disconfirmatory error signals. This arouses consciousness of the error signals (which then loop back to update the priors, to whatever depth is necessary). What comes to consciousness, then, is not a prediction but rather its sensory contradiction – i.e. modulation of the *mismatch*. The indication of reality may therefore be conceptualized as a mechanism for disambiguation. We see once more that consciousness, all of it, including perceptual consciousness, is felt *uncertainty*. Perceptions that carry full confidence *fade* from consciousness (cf. habituation).[62] We see now that, here too, it is the tolerance of error supplied by secondary process *inhibition* that makes possible the criterion for distinguishing between perception and memory.

(16) *Cognition and declarative thought*

We have brought forward the hypothesis that, during the process of wishing, inhibition by the ego brings about a moderated cathexis of the object wished for, which allows it to be cognized as not-yet real; and we may now proceed with the analysis of this process. Several possibilities may occur. In the first case: simultaneously with the wishful ψ cathexis of a representation of the object, a φ signal of it is present. If so, the two patterns of neural activity match – which cannot be made use of biologically – and the wishful cascade ends in automatized action, which need not attract attention. This case is easily dealt with but it almost never happens in reality that predictions and sensory samples match perfectly. This leads to the second case: the wishful cathexis is present and along with it a sample which does not match it wholly but only in part. For the time has come to remember that perceptual patterns never involve single neurons but always distributions of sensory signals in *neuronal assemblages*. So far we have neglected this feature; it is time to take it into account. Let us suppose that, quite generally, the wishful cathexis relates to neuronal assemblage $a+$ neuronal assemblage b, and the sampled distribution to assemblages $a+c$. Since this will be the more common case, more common than that of identity, it calls for more exact consideration. Biological experience will teach here that it is unsafe to initiate action if the indications of reality (the ω assigned to the error signals generated by sample pattern c) do not confirm the whole expected assemblage but only a part of it. A way is now found, however, of completing the similarity into an identity. The sample pattern, if it is compared with other sampled patterns, can be dissected into a component portion, pattern a, which on the whole remains the same, and a second component portion, pattern b, which for the most part varies. Language will later apply the term *judgement* to this dissection and will discover the resemblance which in fact exists between the nucleus of the ego and the constant sampled component on the one hand and between the changing pattern in the pallium and the inconstant component on the other; language will call pattern a the thing and pattern b its activity or attribute – in short, its *predicate*.

Thus *judging* is a ψ process which is only made possible by inhibition by the ego and which is evoked by the dissimilarity between the *wishful cathexis* of a prediction and a sensory pattern which is similar to it. It can be inferred from this that matching between the two becomes a biological signal for ending the act of cognition and for allowing action to begin. Their mismatch gives the impetus for the activity of conscious (and effortful) cognition, which is terminated once more with their match.

The process can be analyzed further. If sensory pattern a matches the predicted one but pattern c is sampled instead of pattern b, then the activity of the ego follows the precise error signals generated by this pattern c and by means of palpating the confidence levels along its collateral connections, through side cathexis, causes new sensory patterns to emerge until perception b is generated. As a rule, the collateral connections activate the memory of an *action* which is interpolated between assemblages c and b; and when this inference is freshly activated through a movement carried out *really*, the perception of pattern b, and at the same time the identity that is being sought, are established. The alternative to this movement – as we saw in Section 3 – would be for the ego to adjust its *prediction* from patterns b to c; but this solution is typically excluded in the case of endogenous needs. Let us suppose, accordingly, that the object wished for (i.e. predicted) by the infant is the neuronal assemblage corresponding to the image of the mother's breast and front view of the nipple, and that the first pattern sensed is a side view of the same object, without the nipple. This generates salient error (and therefore consciousness). In the child's memory there is the representation of an experience, made by chance in the course of sucking but consolidated through the law of affect, that with a particular head movement the front image turns into the side image. The side image which is now seen leads to high-precision error signals which continue until the passage of the movement brings the expected front image into view.

There is not much judgement about this as yet; but it is an example of the possibility of arriving, by precision optimization, at an action which is already one of the trial-and-error offshoots of the specific action, without enduring the dangers associated with fresh learning from experience.

There is no doubt that it is ω optimization that underwrites this cognitive palpating of temporary action/perception possibilities which gives rise to voluntary movement. The choice is dominated not by ψ consolidation but by an aim. What is this aim and how is it achieved?

The aim (i.e. SEEKING) is to find the missing sensory pattern *b* and to decrease precision in the error signal caused by pattern *c* and thereby improve confidence in the corresponding prediction of pattern *b*. It is reached by an experimental displacement of ω along every salient pathway, and it is clear that for this purpose sometimes a larger and sometimes a smaller tolerance of side cathexis is necessary, according to whether one can make use of the certainties that are present or whether one has to work against them. The struggle between the consolidated predictions and the changing and uncertain ω cathexes which underwrite choice and voluntary action is characteristic of the secondary process of declarative cognition, in contrast to the automatized primary sequence with its false certainties.

What is it that directs this cognitive palpating? The fact that the representation of the wished-for object can be held in mind (in working memory) along with – or due to – all the associated uncertainty. As we now know, indications of reality (perceptual consciousness) plays a pivotal role in the process, but this too is ultimately directed by the potential pleasures and unpleasures that are prioritized in the midbrain decision triangle.

In the course of this palpating it may happen that the changing ω weighting activates a memory which is connected with an experience of pain, and it thus gives occasion for the release of unpleasure in the interior of the system. Since this is a sure sign that sensory pattern *b* is not to be reached along that pathway, the cognitive cathexis is at once diverted from the predictions in question. Unpleasurable paths, however, retain their great value for obvious biological reasons in directing declarative cognition.

(17) *Remembering and judging*

Thus declarative thought has a practical aim and a biologically established end – namely, to lead ω from the superfluous (unwanted) sample back to the cathexis of the unfulfilled prediction. With this, identity and the right to action are achieved, if in addition the indication of reality appears from assembly *b*. The process can, however, make itself independent of the latter aim and strive only for identity. If so, we have before us a pure act of thought, though this can in any case be put to practical use later. Here, moreover, the cathected ego is behaving in exactly the same manner.

We now come to a third possibility that can arise in a wishful state: when, that is, there is a wishful cathexis and a sensory sample emerges which does not coincide in any way with the wished-for pattern (mem+). Thereupon there arises an interest in cognizing this sensory image, so that it may perhaps after all be possible to find a pathway from it to mem+. It is to be assumed that, with this aim in view, the sensory pattern (or rather, the error signal it gives rise to) is assigned a high precision, as happened in the previous case with only a component of it, pattern *c*. If the sensory image is not absolutely new, it will now *recall* and *revive* a memory image with which it coincides at least partly. If the sensed pattern *is* completely new, the same process occurs by *analogy* with something known. The previous process of thought is now repeated in connection with this memory image, though to some extent without the *aim* which was afforded previously by the cathected wishful idea.

In so far as the patterns match, they give no occasion for activity of thought. On the other hand, the non-matching portions "arouse interest" and can give occasion for thought activity in two ways. The attention is either directed onto the *aroused* memories and sets an aimless activity of memory at work, which is thus moved by differences and not similarities, or it remains attached to the components of the sensory sample which have newly emerged and in that case exhibits an equally aimless *activity of judging*.

Let us suppose that the object which furnishes the sensory pattern resembles the subject: a *fellow human being*. If so, the theoretical interest taken in it is also explained by the fact that an object *like this* was simultaneously the subject's first satisfying object and further its first noxious object, as well as its sole caregiver. For this reason it is in relation to a fellow human being that a person learns to cognize. Then the sensory patterns proceeding from the encounter with this fellow human being will in part be new and not-matchable – its *facial features*, for instance, in the visual sphere; but other visual impressions – e.g. those of the movements of its hands – will coincide in the subject with memories of quite similar visual impressions of its own, of its own body, memories which are associated with memories of movements experienced by itself. Other impressions of the object too – if for instance, it screams – will awaken the memory of the subject's own screaming and at the same time its own experiences of pain. These memories are encoded in assemblages of what are called "mirror neurons"[63] (although the fact that mirror neurons encode *memories* – predictions – is frequently overlooked). Thus the pattern of the fellow human being falls apart into two components, of which one makes an impression by its constant structure and stays together as a *thing*, while the other can be *understood* by the activity of mirror neurons – that is, can be traced back to information from the subject's own body. This dissection of the sensory pattern is described as cognizing it; it involves a judgement and when this last aim has been attained it comes to an end.

Judgement, as will be seen, is not a primary function, but presupposes ω cathexis from the ego of non-matching portions of the sensory sample;[64] in the first instance it has no practical purpose and it seems that during the process of judging the palpating of the non-matching component is ended, for this would explain why activities, "predicates," are separated from the subject-pattern by a comparatively loose pathway.

It would be possible from this point to enter deeply into the analysis of the act of judgement; but this would divert us from our topic. Let us content ourselves with bearing firmly in mind that it is the original interest in establishing the situation of satisfaction that has led, in the one case, to *declarative consideration* and in the other to *judging*, as a method of proceeding from the sensory situation that is registered in φ to the situation that is wished for in ψ. The necessary precondition for this remains that the ψ processes should not pursue their passage uninhibited but in conjunction with an active ego. The eminently practical sense of all thought activity would in this way seem to be demonstrated.

(18) *Thought and reality*

The aim and end of all thought processes is thus to bring about a state of identity, the conveying of *e* signals emanating from φ onto neuronal assemblages cathected from the ego. *Cognitive* or *judging* thought seeks an identity with a bodily cathexis, *declarative* thought seeks it with an experience of one's own. Judging thought operates in advance of declarative thought by furnishing it with ready-made synaptic connections for further associative traveling. If after the conclusion of the act of thought the indication of reality reaches the perception, then a judgement of reality, confidence, has been achieved and the aim of the whole activity is attained.

As regards judging, there is further to be remarked that its basis is obviously the presence of bodily experiences, sensations and motor images of one's own. So long as these are absent, the variable portion of the sensory pattern remains ununderstood – that is, it can be reproduced but does not point a direction for further paths of thought. Thus, for instance (and this will become important for what is to follow in Part II), no sexual experiences produce any effect upon predictive pathways for satisfying LUST so long as the subject is ignorant of mature sexual desire – in general, that is, till the beginning of puberty. To be clear: while early sexual experiences have immediate consequences for the management of drives that are not in a state of latency – such as SEEKING, FEAR, PANIC/GRIEF and PLAY – and therefore can elicit secondary emotions like shame and guilt, the predictive effects upon LUST (and only LUST) are largely *deferred*.

Primary judging seems to presuppose a lesser influence by the cathected ego than do declarative acts of thought. In primary judging it is a matter of pursuing an association which is due to partial matching between the wishful (predicted) and actual sensory patterns – an association to which no modification is applied. And indeed cases also occur in which the associative process of judging is carried out with a full amount of quantity. The sensory sample may correspond to an object nucleus plus a motor image. While one is sensing the φ pattern, one copies the movement oneself – that is, one innervates so strongly the motor pattern of one's own which is aroused towards matching the φ pattern, that the movement is carried out. Hence one can speak of a perception having an *imitation value*. Or the sensory pattern may arouse the memory of a sensation of pain of one's own, so that one feels the corresponding unpleasure and repeats the appropriate defensive movement. Here we have the *sympathy value* of perception.

However these two cases are not yet *empathy*, which is a further developmental achievement. In these two cases we must no doubt see the primary process in respect of judging, and we may assume that all secondary judging has come about through a mitigation of these purely reflex processes. Thus judging, which is later a means for the *cognizing* of an object that may possibly be of practical importance, is originally an associative process between signals coming from the outside and from one's own body – and *identification of information from φ and from within*. It is perhaps not wrong to suspect that judging at the same time represents a method by which *e*'s coming from *φ* can be transmitted and dealt with. What we call *things* are residues which evade being judged.

The example of judgment gives us for the first time a hint of the difference in their quantitative characteristic which is to be discovered between thought and the primary process. It is justifiable to suppose that during *thought* a slight magnitude of motor innervation passes from *ψ* – only, of course, if during the process a motor or key neuron has been innervated. Nevertheless, it would be wrong to take this active state for the process of thought itself, of which it is only an unintended subsidiary effect. The *process of thought* consists in the cathexis of *ψ* neurons, accompanied by a change, brought about by side cathexis from the ego, in what is imposed by the synaptic weightings. It is intelligible from the statistical-mechanical point of view that here only a part of the *e* is able to follow the certainties and that the magnitude of this part is constantly regulated by the *ω* cathexes. But it is also clear that at the same time enough *e* is economized by this to make the declarative process profitable as a whole. Otherwise, all the *e*, which is finally needed for action, would be given off at the points of *M* action during the course of its passage. *Thus the secondary process is a repetition of the original ψ passage of e, at a lower level, with smaller quantities*.

"With *e*'s even smaller," it will be objected,

> than those that ordinarily pass through in the *ψ* system? How can it be arranged that such small *e*'s shall have open to them pathways which, after all, are only traversable by larger ones that *ψ* as a rule receives?

The only possible reply is that this must be a mechanical result of the side cathexes. We must conclude that matters stand in such a way that when there is a side cathexis small *e*'s flow through the resistances which would ordinarily be traversed only by high ones. The side cathexis as it were *binds* a quota of the free *e* transmitted through the system.

There is a further condition that thought must satisfy. It must make no essential change in the resistances made by the primary processes; otherwise, indeed, it would falsify the traces of real experiences. Of this condition it is enough to remark that consolidation is probably the result of a single quantity of *e* and that cathexis, though very powerful at the moment, nevertheless does not leave any comparable lasting effect behind it. The small *e*'s that pass during thought cannot in general prevail against consolidation.

There is no doubt, however, that the process of thought does leave lasting traces behind it, since a second thinking, a re-thinking, calls for so much less energy than a first. In order that reality shall not be falsified, therefore, special traces are needed, signs of the processes of thought, constituting a thought memory which it is not yet possible to shape. We shall hear later (in Part III) by what means the traces of thought processes are distinguished from those of reality.

(19) *Primary processes – sleep and dreams*

The question now arises as to what, then, the quantitative means may be by which the *ψ* primary process is sustained. In the case of an *experience of pain* it is evidently the irrupting *e* from *φ*; in the case of self-generated *distressing states* it is the endogenous *e* released by remembering. In the case of the secondary process of *declarative thought* a greater or lesser *ω* can obviously be assigned to neuronal assemblage *c* from the ego, and this may be described as *thought interest*, and be proportionate to the *affective interest* where that may have developed. The question is only whether there are *ψ* processes of a primary nature for which the *e* supplied from *φ* is sufficient or whether the cathexis of sensory *e* is automatically supplemented by a *ω* contribution (attention) which alone makes a *ψ* secondary process possible (see below). This question must remain an open one, though it may perhaps be decided if it is specially applied to some psychological facts.

It is an important fact that *ψ primary processes*, such as have been biologically suppressed in the course of *ψ* development, are daily presented to us during sleep. A second fact of the same importance is that the pathological mechanisms which are revealed in some mental disorders by the most careful analysis have the greatest similarity to dream processes. The most important conclusions follow from this comparison, which will be enlarged on later.

First, the fact of sleep must be brought into our theory. It is easy to overlook the fact that sleep, no less than energy balance and hydration, is a bodily need which periodically intrudes on consciousness in the form of a drive: the need for sleep. Here, once more, we see the critical role that *need prioritization* plays. Babies sleep so long as they are not tormented by any other physical need or external stimulus, including sensory affects (hunger and cold from wetting). They go to sleep only when these other needs are satisfied – although, of course, eventually the *e* from the need for sleep attracts such high ω that it *must* be prioritized. Adults, too, fall asleep more easily after dining and copulating. Accordingly, the precondition of sleep is a *lowering of the other endogenous loads pressing for access to the ψ nucleus*, which makes the secondary function superfluous. It becomes superfluous because the need for sleep demands, in the first instance, *withdrawal of ψ cathexis from M and φ* (as occurred with the resting state). In sleep an individual is in the ideal state of complete (or nearly complete) automaticity, rid of the need for voluntary ψ activity – including voluntary thought activity.

In adults, this voluntary cathexis is collected in the "ego";[65] we may therefore assume that it is the *unloading of the ego* which determines and characterises sleep onset. And here, as is immediately clear, we have the *precondition of mental primary processes*.

It is not certain whether in adults the ego is completely relieved of its burden in sleep. In any case it withdraws an enormous number of its cathexes, which, however, are restored on awakening, immediately and without trouble. This contradicts none of our presuppositions; but it draws attention to the fact that we must assume that the synaptic weightings (the resistances) between neurons which are properly linked remain unaltered during sleep. Or do they?

Sleep is characterised above all by *motor paralysis* (paralysis of externally directed *M*). The "will" is abandoned. This is due to the fact that in sleep spinal tonus is relaxed; and this applies all the more during REM sleep. However internally directed *M* persists during sleep together with the endogenous sources of ψ excitation.

It is a highly interesting fact that the state of sleep begins and is evoked by a closure of those sense organs that are capable of being closed. Perceptions should not be made during sleep, and nothing disturbs sleep more than the emergence of sense impressions: signals entering from φ and cathected by ψ. This is consistent with our hypothesis that during day-time a tonic, even though displaceable, cathexis (attention) is sent into the pallium neurons, which palpates expected patterns of φ activation, so that the carrying out of the ψ secondary processes is made possible with the adjustment of expected ω. The salient φ neurons are, as it were, precathected. Pallium cathexis is, however, ultimately contingent on the deeper predictions emanating from the ψ nucleus, which are in turn activated from the ω decision triangle as prioritized affect. If ψ withdraws these pallium cathexes (i.e. when the need for sleep is prioritized) the φ signals fall upon largely uncathected neurons and the possibility of *e* transmission becomes slight. That is, ω values are increased in the centrifugal predictive pathways giving expression to the "expected context" of the drive to sleep, which implies that the tolerance of exteroceptive *e* will be greatly increased.[66] This implies also that indications of reality (and therefore perceptual consciousness) are terminated during sleep. As we have conjectured, along with this the innervation of attention comes to a stop as well. It is from here, too, that the enigma of hypnosis may be approached. The apparent unexcitability of perceptual consciousness in deep hypnosis must rest on this withdrawal of the cathexis of attention.

Thus, by a withdrawal process, the counterpart of directing attention, the ψ pallium excludes φ impressions so long as it itself is uncathected.

But what is strangest of all is that during sleep ψ processes occur – dreams – which have many characteristics that seem to contradict the drive to sleep.

(20) *An analysis of dreams*

Dreams exhibit every transition to the waking state and to a mixture with normal ψ processes; yet it is easy to sift out what is genuinely in the nature of a dream.

(a) Dreams are *devoid of motor discharge*. We are paralyzed in dreams.

The easiest explanation of this characteristic is the absence of spinal *M* precathexis, which renders φ signals largely irrelevant. As I explained before, action and perception are two sides of the same coin; they may accordingly both be terminated in essentially the same way, through modulation of expected ω. In some dream states (e.g. somnambulism) movement is not excluded. This cannot be the most essential characteristic of dreams.

(b) The ideational connections in dreams are partly *nonsensical*, partly *feeble-minded*, or even meaningless or strangely crazy.

This latter characteristic is explained by the fact that in dreams a *compulsion to associate* prevails, as it does primarily in mental life generally. Two neuronal assemblages that are present simultaneously *must*, so it seems, be brought into connection. The relationship between this compulsion and the default SEEKING drive may now be determined. It raises an important question: how can SEEKING be prioritized simultaneously with sleep? The answer is that the drive to sleep takes priority only at *sleep onset*. This prioritization does not mean that SEEKING activity ceases; it means only that it becomes automatized. If, then, SEEKING is subsequently re-prioritized *while ω and M are decathected*, sleep need not come to an end; the management of the need for sleep then becomes automatized. (This is of course facilitated by the fact that the need for sleep decreases as sleep itself proceeds.)

The two other characteristics, which are in fact identical, show that a part of the dreamer's mental experiences have been forgotten. Actually, indeed, all the biological experiences which ordinarily inhibit the primary process are forgotten, and this is owing to the lack of ego cathexis. The senselessness and illogicality of dreams are probably to be attributed to this very same characteristic. It seems, however, that activity in the ψ nucleus does not cease with sleep onset, and that it continues to wander through the deeper pallium layers. If the ego were completely unloaded, sleep would necessarily be dreamless (and indeed, thoughtless). At this point, to avoid unnecessary repetition, I must ask you to re-read the final paragraph of Section 14, concerning the *memory consolidation* function of the resting state.

(c) Dream ideas are of an hallucinatory kind; they awaken consciousness and meet with belief.

This is the most important mental characteristic of sleep. It appears at once when there are alternating spells of sleeping and waking. One shuts one's eyes and hallucinates; one opens them and thinks in words. There are several explanations of the hallucinatory nature of dream cathexes. In the first place, it might be supposed that when the prioritization of action as a means for reducing *endogenous e* during waking life ceases with sleep onset, the alternative pathway (perception)[67] *must* be given priority. The only argument against this is the consideration that the φ neurons, by the fact of being uncathected from ψ, should be protected from generating *e* signals. But this overlooks two points. A withdrawal of ψ cathexis implies only a cessation of its *secondary* process. And, moreover, primary processes in the ψ pallium are *unconscious* unless and until they generate indications of reality – by way of the ω mechanisms described in Section 16. This implies that the primary task of the nervous system to suppress surprisal, which must continue throughout sleep, gives rise in the first instance only to uncontrolled "hallucinations" (i.e. unconscious updating of the predictive model). Secondly, therefore, we revert to the nature of the primary process and point out that the primary activation of a perceptual memory trace is always an "hallucination," and that only inhibition of the ego has taught us to palpate the expected ω assigned to the incoming *e* signals. It is this *secondary* process that yields the indications of reality which bestow "belief" (i.e. confidence) in the resultant perception. How, then, does it happen that hallucinations during sleep awaken consciousness and meet with belief? The answer is to recall that consciousness, all of it, is an endogenously generated property. Endogenous sources of *e* persist in sleep, and so they arouse *affects* – which renders salient their attendant instinctual and learnt predictions: the expected context. Confusion on this score is the perennial price that cognitive scientists pay for believing that consciousness flows in through the senses. The biological purpose of dreaming is now clear: *hallucinatory resolution of affective demands is required for the preservation of the state of sleep*. This explanation is further supported by the circumstance that in dreams the vividness of the hallucination is directly proportionate to the salience – that is the ω cathexis – of the idea concerned. This indicates that it is ω modulation that determines hallucination. If a perception comes from φ in waking life, it is no doubt made clearer by ψ cathexis (interest), but not more raw; it does not alter its affective characteristic. As we know, this always comes from within. But still we face the question: what is the source of the *changing e* values in the outer layers of the ψ pallium that demand ω modulation during sleep?

(d) The aim and sense of dreams (of normal ones, at all events) can be established with reasonable certainty. Dreams are wish-fulfillments – that is, primary process SEEKING following upon experiences of satisfaction of the various drive demands (i.e. "wanting"); and they are only not recognized as such because the release of pleasure (the reproduction of pleasurable discharges of *F*) in them is slight, because in general they run their course almost without effect (without motor action). That this is their nature is, however, easily shown; the evidence for the hypothesis that dreaming is driven by dopaminergic SEEKING is now overwhelming.[68] Combining this fact with our insights concerning default-mode "mind wandering" in the service of consolidation (simplification of the generative model), which characterises the ego's resting state when it is unoccupied by external tasks, we arrive at the following hypothesis.[69] Default-mode pruning of recently acquired synaptic connections (i.e. recent *Qη* updates) will come upon traces of new learning in the ψ pallium which contradict *repressed* ψ nucleus predictions. Recall that repressed predictions cannot be updated. These mismatches must be resolved through disambiguation: hence the *mental work* of dreaming; which, it follows from the nature of repression, must aim to explain away the mismatching, exogenously acquired knowledge. *That is how primary-process wishful cathexes during sleep give rise to hallucinatory consciousness*. It is now easy to see why dreams so frequently are unpleasurable: repressed (illegitimately automatized) predictions can fail to resolve any one of the emotional drives enumerated in Section 4; and situations evoking FEAR, RAGE, PANIC/GRIEF, etc., are just as likely to be encountered during virtual "foraging" as they are in waking life. Failure to manage these affects would mean a failure of the biological function of dreaming which is in service of the drive to sleep.

(e) It is noteworthy how poorly dreams are remembered and how little harm they do as compared with other primary processes. But this is easily explained from the fact that, for the most part, they follow old synaptic connections and thus make no change in them and, moreover, from the fact that the synaptic pruning that characterises default mode activity (especially regarding the repressed) serves *forgetting* more than learning. In addition, owing largely to the paralysis of motility, dreams do not leave traces of fresh *φ* experiences.

(f) It remains interesting that consciousness in dreams furnishes quality with as little trouble as in waking life. This shows once more that consciousness is primarily affective; it does not cling to the ego but can become an addition to any *ω* (precision optimization) process. This warns us, in short, against possibly identifying id processes with unconscious ones!

If, when the memory of a dream is retained, we inquire into its content, we find that the meaning of dreams as wish-fulfillments is concealed by a number of *ψ* processes (defences): all of which are met with once more in mental disorders and characterise the latter's pathological nature.

(21) *Dream consciousness*

Consciousness of dream ideas is above all discontinuous. What becomes conscious is not a whole succession of associations, but only separate stopping points in it. Between these there lie unconscious intermediate links which we can easily discover when we are awake. If we investigate the reasons for this skipping, here is what we find. Let ***A*** be a dream idea which has become conscious and which leads to idea *B*. (I will set the symbols for conscious ideas in **bold type** and the unconscious ones not.) Instead of *B*, ***C*** is found in consciousness, and this is because it lies on the pathway between *B* and a *D* cathexis which is simultaneously present. Thus there is a diversion brought about by a simultaneous cathexis, of a different kind, which, is not itself conscious. For that reason, then, ***C*** has taken the place of *B*, though *B* fits in better with the connection of thought, with the wish-fulfillment.

For instance, in the famous "specimen dream,"[70] Otto has given an injection of *propyl* to Irma. The dreamer (Freud) then sees *trimethylamine* before him very vividly, hallucinated as a formula. Explanation: The thought simultaneously present is the sexual nature of Irma's illness (*D*). Between this thought and the propyl (***A***) there is an association from sexual chemistry (*B*), which Freud had discussed with his friend Fliess, in the course of which he had brought Freud's special attention to trimethylamine (***C***). This now becomes conscious owing to it drawing *e* signals from both sides (from *B* and *D*).

It is very puzzling that neither the intermediate link (*B* – sexual chemistry) nor the diversionary idea (*D* – the sexual nature of the illness) becomes conscious as well, and an explanation of this is called for. One would suppose that the *ω* cathexis of predictions *B* or *D* alone is not ambiguated enough to make its way through to a conscious hallucination (i.e. to disambiguation of the *ψ* and *e* signals), but that ***C***, drawing error signals from both sides, would bring it about. In the example chosen, however, *D* (the sexual nature of the illness) was certainly as intense as ***A*** (the propyl injection) and the derivative of these two, the chemical formula (***C***), was immensely vivid. The puzzle about unconscious intermediate links applies equally to waking thought, where similar events are of daily occurrence. But what remains characteristic of dreams is the ease with which *e* is displaced in them and accordingly the replacement of *B* by ***C*** which becomes superior to it quantitatively. This can only be due to a *defensive* use of precision modulation, to reduce unpleasure (on the model of distressing memories generally, discussed in Section 13) and therefore to preserve sleep.

Similarly with the fulfillment of wishes in dreams generally. What happens is not, for instance, that the wish becomes conscious and its fulfillment is then hallucinated, but only the latter: the intermediate link is left to be inferred. It has quite certainly been passed through, but without being able to develop declaratively. It is evident, however, that the predictive cathexis of the wishful idea cannot possibly be stronger than the motive impelling it. Thus the passage of ideas in dreams takes place in accordance with *e*; but it is the *ω* weighting rather than the base value which decides the question of it becoming conscious.

It should be inferred from dream processes, as from perception, that consciousness arises *during* the modulation of *ω*; that is, it is not awakened by a monotonous cathexis.[71] It should further be inferred, as we know, that consciousness does not attach to deeply consolidated predictions, since they are routinely assigned high precision values. It is to their uncertain outcome that consciousness attaches – and this occurs progressively towards the periphery of the predictive cascade. Thus, we can only conclude as we did before that repression and defence are two separate processes. Repression arises from illegitimate consolidation (long-term resistance to change) while defence arises from short-term precision modulation, diverting the predictive cascade from its automatic outcome – which if left to its own devices can only bring intense surprisal in its wake (it being the outcome of an illegitimate prediction). We must in addition take account of the fact that surprisal arising from repression is felt *affectively* (near the core of the hierarchy) whereas that arising from defence is of a more *cognitive* nature. Defence causes errors of reasoning and logic rather than the shattering of core beliefs.

The peculiarities of dream consciousness may, in conclusion, be added to the evidence for our view that the id should not be conflated with the repressed. The repressed consists in life-long beliefs which, when enacted illegitimately (virtually or really) can only yield errors which will be felt in the brainstem core. Id consciousness, feeling, then punishes the unconscious ego for its errors, by making renewed demands for work. Reconsolidation (cognitive consciousness, which must be tolerated by the ego) enables it to mend its ways. Since this solution is excluded in the case of repressed beliefs, defence – the suppression of affect through misleading cognitions – is its only recourse.

April 17th, 2020

Notes

1. Friston's work builds directly upon Helmholtz's insights, not least concerning the conservation of energy. The first law of thermodynamics states that energy is never lost or created, it is only *transformed*. The second law states that energy always *dissipates* during natural processes. The driving mechanism of self-organizing (e.g. living) systems is to resist this dissipative tendency, by minimizing free energy. That is the free energy principle.
2. Please note: the Greek letters used here are purely conventional; they have no literal meaning. Also, although the symbols denote concepts that are equivalent to those that Freud used, they are not *identical* with them; this is because the concepts have been substantively updated.
3. "Free energy." There are various types of free energy. *F* denotes "variational" free energy, or Friston free energy for short. This is analogous to Helmholtz free energy, where there is an *information* exchange as opposed to a *thermodynamic* exchange between a system and its environment. (See Friston, 2009.)
4. In Shannon's (1948) mathematical treatment of "information," the less predictable an event is, the more information it carries. The average information of a system is its *entropy* (a concept formally related to but not identical with free energy). Therefore, the less information required by a self-organizing system (i.e., the lower its uncertainty, the fewer yes/no questions it needs to ask), the better for the system. Uncertainty is, for obvious reasons, dangerous for biological systems.
5. This "orphan" sentence is a good example of the challenges posed by Freud's dense text. In the original version, the sentence reads: "*N* and *Qη* – Similar experiments are now frequent." There, *N* denoted "neurons" and *Qη* denoted "quantity of an intercellular order of magnitude." What Freud meant to convey was that he was not the only scientist using these two basic ideas to conceptualize the dynamics of the nervous system; and I am trying to convey the same, in relation to neurons and information theory. – Incidentally, the neuron had only just been discovered at the time of Freud's writing the "Project."
6. Freud (1915, pp. 121–122) defined drive as "the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demand made upon the mind for work in consequence of its connection with the body." Here, I am equating "drive" with *F*, bearing in mind that the "free energy" (*F*) within a system is the energy that is not currently performing useful work. Helmholtz contrasted it with "bound energy."
7. The basic principle in question is the free energy principle, which is formulated in Friston's law: "All the quantities that can change; i.e. that are part of the system, will change to minimise free energy" (Friston & Stephan, 2007). This law (like the law of affect, described below) is probabilistic; it applies only in the *average* case, i.e. over sufficient time periods. See Friston (2009) for an introduction to this unifying principle of brain functioning.
8. The "principle of neuronal inertia" is the theoretical precursor of Freud's "death drive." This is ironic because homeostasis is the grounding mechanism of all life.
9. The (Bayesian) terms "prior" and "posterior" here refer to the fact that the actions and perceptions of a self-organizing system entail *experiments* which test hypotheses generated by its predictive model. Prior hypotheses are supplanted by posterior ones which take account of the outcome of each experiment. See Hohwy (2013) and Clark (2015) for accessible accounts of the "Bayesian brain."
10. Freud conceptualized this compromise, his "constancy principle," as a special case of Fechner's "tendency towards stability." – The simplest neural contrivance that gives expression to this imperative is the default-mode SEEKING system, described below, which engages *proactively* with sources of uncertainty (see Panksepp, 1998).
11. Cathexis will be further defined below.
12. *F* can also be defined roughly as the sum of squared prediction error.
13. See Mesulam (2000).
14. Cf. Freud's concept of "contact barriers," which is here replaced by a broader conception of resistance.
15. See Felleman and Van Essen (1991).
16. See Tozzi et al. (2016): "Maximizing mutual information [i.e. minimizing demand for information flow] and minimizing metabolic costs are two sides of the same coin."
17. These terms are used by Hohwy (2013) and Clark (2015) respectively, but by no means only by them.
18. See Friston (2005) for a review.
19. I am disregarding the interneurons here.
20. "Neurons that fire together, wire together" (Hebb, 1949).

21. Freud linked this magnitude with Fechner's law: "Sensation varies with the logarithm of the intensity of the stimulus."
22. Technically, free energy may be decomposed into accuracy and complexity. Model evidence is the difference between accuracy and complexity, since models with minimum F provide accurate explanations of data under complexity costs, which in turn means that reducing model complexity improves model generalizability but at the cost of accuracy. The balance between accuracy and complexity determines a model's *efficiency*. In biology, efficiency is everything. (In Bayesian terms: "likelihood" must be assessed in relation to "probability," to prevent *over-fitting*.)
23. The nondeclarative and declarative systems are equivalent to Freud's unconscious and preconscious systems respectively. (The term "declarative" implies "capable of becoming conscious.") The pivotal topic of consciousness is discussed below, as is its relationship to plasticity.
24. Cf. the concept of "working memory," discussed further below. Mental work here refers to ψ cathexis in the face of e. In physiological terms, a "cathected" ψ trace is subject to *reconsolidation*, which is predictive *work in progress*. This mechanism will become clearer below, when the concept of "precision" is introduced.
25. Despite being so obvious, this simple mechanism has enormous clinical ramifications.
26. Q in itself is "unknowable."
27. Cf. Wheeler (1990): "That which we call reality arises in the last analysis from the posing of yes/no questions and the registering of equipment-evoked responses; in short […] all things physical are information-theoretic in origin."
28. I.e. through reconsolidation.
29. Error and prediction units are both assumed to be pyramidal neurons, although they have different patterns of connectivity. In cortex, the former originate in supragranular layers and terminate (centripetally) in layer 4 spiny stellate cells; the latter originate in infragranular layers and target (centrifugally) infra and supragranular cells. However, it is important to say that prediction units are very far indeed from being limited to the cortex.
30. This is post-synaptic *modulation*; see below.
31. Panksepp (1998). "Sensory affects" are discussed below. I do not use Panksepp's term for the interoceptive bodily affects because it can cause confusion: as we shall see, all affects are homeostatic.
32. As we shall see later, it demands engagement with our *representations* of the external world.
33. Emotional needs, unlike bodily ones, pertain in large measure to other *agents* in the world, the behaviour of which is far less predictable than that of inanimate objects. – Please note: learning does not update reflexes and instincts; it supplements them.
34. See Bowlby (1969).
35. Consider the common phobias.
36. Panksepp (1998), Merker (2007).
37. White et al. (2017).
38. The earlier quotation above from Wheeler (1990) is preceded by the following passage: "It from Bit. Otherwise put, every it – every particle, every field of force, even the spacetime continuum itself – derives its function, its meaning, its very existence entirely – even if in some contexts indirectly – from the apparatus-elicited answers to yes or no questions, binary choices, bits. *It from Bit* symbolizes the idea that every item of the physical world has at bottom – at a very deep bottom, in most instances – an immaterial source and explanation."
39. See Kihlstrom (1996) and Bargh and Chartrand (1999) for reviews of the literature on unconscious cognition.
40. Brentano (1874).
41. Technically, as we know, the most salient signal is the one which provides the greatest opportunity for minimizing F.
42. And which, please recall, can be both grabbed and assigned.
43. "Freely mobile" cathexis is *automatized*.
44. Cf. Freud (1920, p. 25): "Consciousness arises instead of a memory trace." Conversely, a memory trace arises instead of consciousness; when consciousness ceases, certainty is restored.
45. See Pfaff (2005).
46. See Merker (2007), Solms (2013).
47. The law of affect states: "If a behaviour is consistently accompanied by pleasure it will increase, and if it is consistently accompanied by unpleasure it will decrease." This law is attributable to Panksepp (1998) who derived it from Thorndike's (1911) law of *effect*.
48. The first scientist to arrive at this important insight was Fotopoulou (2013).
49. The terms "easy" and "hard" here refer to Chalmers (1995). The hard problem is addressed in far greater detail in Solms (2021).
50. See the "Key to Abbreviations." Please note that these quantities are vectors, apart from ω and F which are scalars. The dot notation in the equations below implies a dot product (i.e. matrix or vector multiplication).
51. As stated above, these simplified equations effectively reduce free energy to the likelihood of a Gaussian distribution. In fuller treatments, one would also need to consider hierarchical generative models (with precisions at each level) and accommodate conditional uncertainty about external states. Furthermore, the equations lump all sensory prediction errors together – including the endogenous and exogenous modalities.
52. Technically, this is called a "gradient descent," where the gradient is the rate of change of free energy with precision.
53. Under our simplifying assumptions about the encoding of Bayesian beliefs.
54. Unlike Freud, I am using the term "nucleus" here to denote what is conventionally called the "subpallium." I have already divided the prosencephalon into its telencephalic and diencephalic components. However, the diencephalic structure I discussed previously was the hypothalamus only, the role of which (as a source of endogenous e) is clarified further in this section. The functional-anatomical situation is complicated: the diencephalon includes nuclei (e.g. the lateral geniculate, mentioned before) which clearly belong to the φ system and others (e.g. the subthalamic nucleus) which are functionally inextricable from the basal ganglia – i.e. the "nucleus"

of ψ which I am discussing here – and yet others (e.g. the intralaminar nuclei) which form part of the ω system. These complexities were not recognized in Freud's times.

55. I.e. affects arising from vegetative needs. See Note 31.
56. This apparently occurs in some psychosomatic diseases. The potential consequences of reducing precision on strong *emotional* error signals are less catastrophic, in the short term at least, as we have long known from "hysteria."
57. Hence "infantile amnesia," which applies only to episodic and semantic (declarative) memories.
58. Confidence in an *e* value is inversely proportional to that in its prior prediction, so if the error = 0, the prediction = 1.
59. Cf. the PANIC/GRIEF category described above. This does not imply, as Freud thought, that the need for attachment "leans upon" the need for nourishment (cf. "anaclisis"); these are independent needs and *both* must be met.
60. This corresponds to the distinction between "bound" and "freely mobile" cathexis, mentioned above.
61. Cf. the important paper on this by Carhart-Harris and Friston (2010).
62. See the classical descriptions of this by Riggs and Ratliff (1951) and Ditchburn and Ginsborg (1952).
63. See Rizzolatti and Craighero (2004) for a review.
64. Vittorio Gallese's group describe what happens when mirror-neuron activity is *not* inhibited in this way, as occurs in schizophrenics who fail to distinguish the perceived object from themselves (Ebisch et al., 2012).
65. In babies, ω modulation is entirely in the hands of the midbrain decision triangle: in the "id."
66. Please note, this is not an homogenous cathexis. When sleep is prioritized it remains possible for ψ to "listen out" for salient *e* signals such as, for example, the sound of a baby crying. Here the prediction (the wish) that "the baby will not cry" is assigned a low ω value so that the relevant *e* signal can attain a sufficiently high magnitude to cause re-prioritization of one's needs from sleep to CARE. Likewise, a sufficiently high *e* signal emanating from anywhere in φ will always be capable of grabbing attention in the manner described before.
67. See the end of Section 3.
68. See Solms (2011) for a review.
69. See Hobson and Friston (2012).
70. Freud (1900).
71. Here we have a deep mechanical account of the principle of constancy.
72. Although Freud (1920) later forgot this.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Bargh, J., & Chartrand, T. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479. https://doi.org/10.1037/0003-066X.54.7.462

Bowlby, J. (1969). *Attachment*. Hogarth Press.

Brentano, F. (1874). *Psychologie vom empirischen Standpunkte*. Duncker & Humbolt.

Carhart-Harris, R., & Friston, K. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain*, *133*(4), 1265–1283. https://doi.org/10.1093/brain/awq010

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Ditchburn, R., & Ginsborg, B. (1952). Vision with a stabilized retinal image. *Nature*, *170*(4314), 36–37. https://doi.org/10.1038/170036a0

Ebisch, S., Salone, A., Ferri, F., De Berardis, D., Romani, G. L., Ferro, F. M., & Gallese, V. (2012). Out of touch with reality? Social perception in first episode schizophrenia. *Social Cognitive and Affective Neuroscience*, *8*. https://doi.org/10.1093/scan/nss012

Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1

Fotopoulou, A. (2013). Beyond the reward principle: Consciousness as precision seeking. *Neuropsychoanalysis*, *15*(1), 33–38. https://doi.org/10.1080/15294145.2013.10773715

Freud, S. (1900). The interpretation of dreams. In *Standard edition of the complete psychological works of Sigmund Freud*, 4 & 5. Hogarth.

Freud, S. (1915). Instincts and their vicissitudes. In *Standard edition of the complete psychological works of Sigmund Freud*, 14 (pp. 117–140). Hogarth.

Freud, S. (1917). A metapsychological supplement to the theory of dreams. In *Standard edition of the complete psychological works of Sigmund Freud*, 14 (pp. 219–235). Hogarth.

Freud, S. (1920). Beyond the pleasure principle. In *Standard edition of the complete psychological works of Sigmund Freud*, 18 (pp. 7–64). Hogarth.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K., & Stephan, K. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458. https://doi.org/10.1007/s11229-007-9237-y

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Hobson, J. A., & Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, *98*(1), 82–98. https://doi.org/10.1016/j.pneurobio.2012.05.003

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Kihlstrom, J. (1996). Perception without awareness of what is perceived, learning without awareness of what is learned. In M. Velmans (Ed.), *The science of consciousness: Psychological, neuropsychological and clinical reviews* (pp. 23–46). Routledge.

Luria, A. R. (1980). *Higher cortical functions in man*. Basic Books.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and*

Brain Sciences, *30*(1), 63–68. https://doi.org/10.1017/S0140525X07000891

Mesulam, M. M. (2000). Behavioral neuroanatomy: Largescale networks, association cortex, frontal syndromes, the limbic system and hemispheric lateralization. In M. M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 1–120). Oxford University Press.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Pfaff, D. (2005). *Brain arousal and information theory*. Harvard University Press.

Pribram, K., & Gill, M. (1976). *Freud's project re-assessed: Preface to contemporary cognitive theory and neuropsychology*. Basic Books.

Riggs, L., & Ratliff, F. (1951). Visual acuity and the normal tremor of the eyes. *Science*, *114*(2949), 17–18. https://doi.org/10.1126/science.114.2949.17

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Sacks, O. (1984). *A leg to stand on*. Simon & Schuster.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Solms, M. (2011). Neurobiology and the neurological basis of dreaming. In P. Montagna & S. Chokroverty (Eds.), *Handbook of clinical neurology*, 98 (3rd series) *Sleep disorders*, Part 1 (pp. 519–544). Elsevier.

Solms, M. (2013). The conscious id. *Neuropsychoanalysis*, *15*(1), 5–19. https://doi.org/10.1080/15294145.2013.10773711

Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Frontiers in Psychology*, *9*, 2714. https://doi.org/10.3389/fpsyg.2018.02714

Solms, M. (2021). *The hidden spring: A journey to the source of consciousness*. Norton.

Solms, M., & Friston, K. (2018). How and why consciousness arises: Some considerations from physics and physiology. *Journal of Consciousness Studies*, *25*, 202–238.

Thorndike, E. (1911). *Animal intelligence*. Macmillan.

Tozzi, A., Zare, M., & Benasich, A. (2016). New perspectives on spontaneous brain activity: Dynamic networks and energy matter. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00247

Wheeler, A. (1990). *A journey into gravity and spacetime*. W.H. Freeman.

White, B., Berg, D., Kan, J., Marino, R. A., Itti, L., & Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, *8*(1), 14263. https://doi.org/10.1038/ncomms14263

Appendix. Letter of June 25th 2020

Dear Peer Commentators,

Sigmund Freud sketched Part I of his momentous "Project for a Scientific Psychology" over a period of two weeks, between September 10th and 25th, 1895. Although it remained unpublished during his lifetime, the "Project" caused a huge splash when it was published in 1950 (in German, the English translation appeared in 1953). In this work, Freud attempted to achieve what has since become the goal of an entire interdisciplinary field called neuropsychoanalysis: namely, to integrate our psychoanalytic and neuroscientific perspectives upon the mental instrument. For this reason, I have found it worthwhile to try to update the "Project" in light of subsequent developments, particularly in affective and computational neuroscience. I now hope that you will help me to further develop this "New Project."

Freud's original text was drafted in the context of an ongoing correspondence with his friend and colleague Wilhelm Fliess. Two weeks after completing it, he submitted the draft to Fliess for comment, together with Part II, while continuing with the drafting of Part III. Fliess's side of the correspondence has not survived but, in response to his comments, Freud revised his "General Scheme" fundamentally. He even went so far, in a letter dated January 1st, 1896, to tell Fliess that he had moved the hypothetical ω system of neurons from its deep position in the processing stream to a slot between the φ and ψ systems, thereby paving the way for his subsequent decision to combine the metapsychological systems *Pcpt*. (φ) and *Cs*. (ω) into a single system *Pcpt.-Cs*. (Freud, 1917).

In my view, that was a mistake. By conflating perception with consciousness, Freud lost the opportunity for recognizing that the final arbiter of feeling was located not in the cortex but much deeper within the system, even deeper than the "ego" nucleus, in what he would later call the "id" (see Solms, 2013). Notwithstanding this unfortunate decision, the 1896 revision had the enormous advantage of allowing Freud to recognize that all "mental energy" is generated endogenously; none of it flows in through the senses.[72]

I am submitting my "New Project" to you in the same spirit that Freud submitted his "Project" to Fliess. It is a rough working draft, with many loose threads and some gaping holes. I hope that you will help me to tie up the threads and begin filling the holes. I am inviting you, in particular, to do so because of your considerable expertise in one (or more) of the specialist topics trenched upon in my text, sometimes in amateurish fashion. Feel free to comment on only those topics that concern or interest you. There is every reason to believe that you can assist where I have gone astray; so that we may collectively move closer to completing the task of furnishing, as Freud put it, "a psychology that shall be a natural science."

In order to follow the sometimes strange turns of my paper, I invite you to consider it in tandem with Freud's original "Project," as I have keyed my text to his, sentence for sentence where possible, and paragraph for paragraph where close paraphrasing was impossible. The editors are sending you a document with "tracked changes" to the original version, showing all the revisions and additions I have made. You will see there that Freud's "Project" was written in a highly condensed style. He expanded this draft in his later metapsychological papers. My forthcoming (2021) book provides an equivalent expansion of this "New Project."

I very much look forward to your comments.

With sincere thanks for accepting the invitation, and with all good wishes,

Mark Solms