Contemporary Neuropsychiatry

Contemporary Neuropsychiatry:

Implications from Cognitive Neuroscience

Edited by

Drozdstoy Stoyanov

Cambridge Scholars Publishing

Contemporary Neuropsychiatry:
Implications from Cognitive Neuroscience

Edited by Drozdstoy Stoyanov

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-9455-6
ISBN (13): 978-1-5275-9455-5

TABLE OF CONTENTS

PREFACE

Contemporary neuropsychiatry is largely informed by the methods and sources of cognitive neuroscience and phenomenology.

This book attempts to consolidate some of the recent developments in the field.

The first part covers topics from phenomenology and clinical psychopathology.

Phenomenology and neurobiology of well-being set the stage in the opening chapter to formulate a comprehensive understanding of the limits of mental health and the premises of psychopathology. This fundamental chapter is delivered by David Foreman.

Phenomenological discourse is adopted in further investigation of the embodied attention by Francesca Brencio.

Formal thought disorder is explored by Dornelles and Telles-Correia in view of its pathogenetic mechanisms. Depression is reviewed from the perspective of developmental neuro-psychology in the next chapter by Matanova and Kostova.

Okan Caliyurt focuses on the meeting of descriptive psychopathology, inherited from the XX century tradition with post-modern trends in artificial intelligence approaches to construction of psychiatric diagnosis and treatment.

The second part includes topics explaining the mechanisms of mental disorders.

Chronobiological factors are largely underestimated as explanatory models in psychiatry. The contribution of Nadejda Madjirova is a tour-de-force introduction to this field and represents her legacy generated for decades of intensive research commitment.

Neurogenesis, more specifically stress-related alterations in hippocampus are presented in the chapter to follow by Peshev and Milanova.

Cerebellar mechanisms, ataxia and movement regulation are implicated in most of the mental disorders. The work of Haralanova and associates summarizes an effort to link neurological methods to psychiatry, with an emphasis on the potential application of original quantitative methods for evaluation of the movement pattern disorders in schizophrenia.

Finally, Basli and Bounaas present innovative treatment strategies in Alzheimer's disease, driven by the advances of molecular neuroscience.

This book is produced by an international group of established scholars with the intention to serve as state-of-the-art update about applications of neuroscience, cognitive science and phenomenology in the transition of psychiatry from a hybrid proto-science to a medical discipline.

It is dedicated to the loving memory of our colleague Prof. Dr Nadejda P. Madjirova (1945-2021), who authored one of the chapters in it.

Plovdiv
October 27th 2022
Drozdstoy Stoyanov

CHAPTER ONE

WELL-BEING THEORY: USING THE FREE ENERGY PRINCIPLE AND ACTIVE INFERENCE TO CONNECT PHENOMENOLOGY AND PSYCHOPATHOLOGY TO BIOLOGY THROUGH WELL-BEING

DAVID FOREMAN
KING'S COLLEGE & ROYAL COLLEGE OF PSYCHIATRISTS, LONDON

Abstract

Explaining the relationship between biology and subjectivity is challenging. Well-Being Theory uses the Free Energy Principle and Active Inference Model to formally connect objective biology to subjective phenomenology and psychopathology through well-being. The first section introduces the concepts of Markov blankets as necessary for an organism's survival. Generative modelling of informational free energy up to a bound is necessary for the organism to maintain adaptation to its environment without breaching its Markov blanket. Well-being becomes an organism's subjective measure of free energy, extending from simple to hierarchical organisms such as us, and is shown to be equivalent to ordinary measures of well-being, such as health-related quality of life. However, we must redefine our concept of our everyday selves as a set of nested environments, with upper and lower scale boundaries defined by embodiments of information-theoretic estimations. The second section shows how the embodiment of well-being across scale creates the unscalable subjective dimensions of our minds, while we can deduce both ordinary and pathological psychological phenomena from the necessary properties of our generative models. Well-Being Theory can thus be a useful tool for

modelling general principles of phenomenology and psychopathology in terms of plausible biology, thus replacing the biopsychosocial model, which is shown to be inadequate to model well-being across scale.

Keywords: well-being; phenomenology; free energy principle

Introduction

It is current orthodoxy that the domains of genetics, neurobiology, economics, and sociocultural factors all contribute to disturbances in our mental health, expressed as psychopathology. However, our theories about how this happens are extremely vague. The face validity of concepts such as the biopsychosocial model have not been matched by advances in construct or predictive validity (e.g., Ghaemi, 2011), and quantification remains elusive.

Enormous strides have been made in methodology, reflected in large and increasing numbers of valuable, insightful publications. However, the practical obstacles to a purely empirical search strategy for finding efficient causes are insuperable. For example, in psychiatric genetics & genomics, sample sizes are now tens of thousands and still inadequate for complex interactions (e.g., Sullivan et al., 2018). In neuroimaging, the small sample sizes that the cost of the technologies impose, together with the number of investigator-sensitive parameters each methodology possesses, have generated a literature full of noise (Botvinik-Nezer et al., 2020; Marek et al., 2020, 2022). At the other end of the scale, indubitable societal influences on mental health, such as poverty, are also noisy, being only weakly predictive of mental health difficulties at the individual level (Das, Do, Friedman, McKenzie, & Scott, 2007; Ridley, Rao, Schilbach, & Patel, 2020). Practically, this means that advances in mental health therapy remain largely serendipitous, while biological and sociocultural psychopathologists propose competing potential explanations that are necessarily incomplete.

This failure may reflect the "curse of dimensionality" (Spruyt, 2014) that challenges artificial intelligence and machine learning (AI/ML): here, the dimensions arise from the progress of empirical research in discovering ever more parameters—with their associated errors in measurement—to be considered. These difficulties are compounded as the parameters have different scales, affecting how the errors in their measurement can be managed in respect of each other. In AI/ML, the solution is some form of weighting, where previous experience is abstracted to produce a set of rules which guide the search for the best solutions. Theories about mental health could provide such weighting in determining research choices, but currently, those available are either highly localised, frequently compete, or, if high-level, are vague, and the relationship between different theories remains undefined.

The introduction of Research Domain Criteria (RDoC) (NIMH, 2016) attempted to provide an empirically based framework that would scale sufficiently for cross-domain research. However, Patrick and Hajcak (2016) observed that further progress would depend upon developing

> "… a systematic measurement-oriented approach to clarifying relations between psychological phenomena and physiological variables in the service of understanding clinical problems." (Patrick & Hajcak, 2016)

In short, RDoC will fail without some measure-related theory to link those parameters. Unfortunately, Carcone & Ruocco (2017) found that research using the RDoC paradigm tended to sit within specific RDoC domains rather than form bridges between them and empirical attempts to link to other measurements struggle (Funkhouser et al., 2021). Within biology, successful theories are scalable; Darwinian evolution's random mutation and selection are good examples. In contrast, psychopathology, which dominates our attempts to understand mental disorders, is extremely resistant to scaling, as it is based on mapping humans 'subjectivities using principles derived from Husserl's phenomenology (Jaspers, 1963). Viewing well-being as psychopathology's inverse can resolve these scaling difficulties as it is not limited to subjectivity, and treating it as "model evidence" for free energy minimisation allows its embodiment within a scalable theory of biological cognition, called here the Free Energy Principle (FEP) (Kirchhoff, Parr, Palacios, Friston, & Kiverstein, 2018) which may give an account of the genesis of well-being and psychopathology (Corcoran & Hohwy, 2017; Miller, Rietveld, & Kiverstein, 2021). Here, the approach is used to develop an embodied general theory of phenomenology to help empirical research link biology with subjectivity.

Imagining a Biological Organism from the Perspective of Well-Being

Imagine an organism O living in an environment $\boldsymbol{E}$. We are interested in how O can persist in $\boldsymbol{E}$, given that only a subset of $\boldsymbol{E}$, E, will permit this, and $\boldsymbol{E}$ is subject to random variation. We consider two points of view. There is the observer's perspective, but also the perspective of the organism itself, and the two will not always coincide, e.g., when O makes a mistake. Formulae reflecting an observer's point of view will be written in black, but those from the organism's point of view in red to aid clarity. Table 1 provides a list of symbols.

Symbol	Interpretation
Red, Black	Organism's (subjective) perspective observer's (objective) perspective
O	Biological organism
$\mathbb{O}$	Observer
$\boldsymbol{E}$	General specifier of environment
E	Subset of environment which can sustain the organism
V	Generic random variable
$\subset$	Subset
$\sqcup$	Disjoint union
$\perp$	Conditional independence
$\setminus$	Disjoint sets
V_O	Environmental variable values that support the organism
$V_{\mathbb{O}}$	Environmental variable set of an arbitrary observer
$D_{KL}(X \parallel Y)$	Kullback-Liebler divergence (between X and Y)
$\boldsymbol{S}$	Entropy
$\mathbb{S}$	Relative Informational Free Energy between X and Y
$\mapsto$	X maps to Y
δ	Modifier indicating a small increment or variation to a max/min value.

Symbol	Interpretation
$F or F$	Objective or Subjective Well-being
$\mathcal{L}, \mathcal{L}$	Laplace transformation
s	The frequency parameter of the Laplace transformation
C	Constant of integration
$\mathfrak{S} or \mathfrak{S}$	Surprisal (satisficing is $-\mathfrak{S}$)
m	Model used to generate predictions (generative model)
Θ	Parameters of generative model
p, q	Generic probability distributions
iid	Variable(s) drawn from identical & identically distributed populations
R, R	The organism's biology, understood from its own perspective
L	Lagrangian
T	Total kinetic energy in a system
V	Total potential energy in a system
S	Action in a Lagrangian system
U	Economic utility
$\succ$	Succeeds (also with reference to scale)
W	Embodied subjective well-being
Ψ, Ψ	O's communication, command and control system (C3)
ψ	That part of C3 phenomenologically represented as "mind"
$\hat{F}, \hat{\mathbb{S}}$	Imputed well-being or relative informational free energy

Symbol	Interpretation
$\mathcal{F}$	Subjective embodiment of $\mathcal{L}$; dimension of phenomenological space
$\tilde{\mathcal{F}}$	Psychopathological dimension

Table 1: List of Symbols (in order of use)

The Markov Blanket

Defining $\boldsymbol{E}$ as generally as possible, we can say that it is a set of random variables $\{V_1 \dots V_n\}$. An organism may be present, but, as observers, we do not yet know. We guess that $O \subset \boldsymbol{E}$, as we assume O will include at most $\{V_{a\geq 1} \dots V_{b<n}\}$, for any O will need some separable environment. We also note that $O \sqcup \boldsymbol{E}$ so it can sustain itself in its environment and not dissolve. A Markov blanket $O \perp \boldsymbol{E} \backslash V_O | V_O$ can enable this for O, as it defines a boundary between O and $\boldsymbol{E}$. It means that we, or O, can deduce its existence from observing V_O. It follows that $V_O = E$. We should note that if either we or O do not observe V_O (E) we will not conclude an organism is present, and O must be able to self-identity, as otherwise it will be destroyed when it meets values of $\{V_1 \dots V_n\}$ that are not compatible with its continued existence. In symbols, this is $O \sqcup \boldsymbol{E} | E \subseteq V_{\mathbb{O}}$. Observers have Markov blankets, too. We shall see that $E \subseteq E$.

Deriving Well-Being from Informational Free Energy

The problem of how O can do this without breaching its Markov blanket (and so invite dissolution) is solved by invoking the Free Energy Principle (FEP). In general, informational free energy is defined as $\boldsymbol{S}(Q|\{X_i\}) = -K \sum_i p_i \log p_i$, where $\boldsymbol{S}$ is entropy, Q the quantity being considered, $\{X_i\}$ is a set of variables and p_i their probability distributions. Free energy is high when there are many equivalently possible outcomes; it decreases as the number of possibilities falls.

Provided it has some prior information, O can make a reasonably informed guess about what is out there without breaching its Markov blanket. In our notation, this is E. If the guess is correct, all is well, so it needs to know how different its guess is from what is *really* out there. Then, it can adapt. The Kullback-Liebler divergence between predictions E and reality E $D_{KL}(E \parallel E)$ will capture this, which is also the relative informational free energy of that difference, $\mathbb{S}$ (Buckley, Kim, McGregor, & Seth, 2017). The asymmetry of D_{KL} makes space for subjectivity, implicit in E. For O, The difference in terms of free energy is $\mathbb{S}(O \sqcup \boldsymbol{E}|E) = -K\sum_O p_O \ln p_O$.

However, our organism still must get from E to E without breaking its Markov blanket, as what it needs to survive is $\mathbb{S}(O \sqcup \boldsymbol{E}|E) = -K\sum_O p_O \ln p_O$. It can do this by acting as if E were true, monitoring its consequential experiences (which are subjective) and plugging them into a loss function, something that reaches a minimum (or equivalent maximum) when $\boldsymbol{E} = E$. This is equivalent to trying to minimise $\mathbb{S}$ through its actions, symbolised as $\lim_{\mathbb{S}\to 0} f : \nabla\mathbb{S} \mapsto \nabla\mathbb{S}$; the red indicates that O needs to map the real gradient change into subjectivity to act upon it. O needs time $t + \delta t$ to enact its prediction, so the quantity of interest to O is actually $\lim_{\mathbb{S}\to 0} \int \mathbb{S}dt \sim \lim_{\mathbb{S}\to 0} \int \mathbb{S}dt|O$. A Laplace transform can estimate this as an expected value for a probability distribution (here shown for the gradient $\nabla\mathbb{S}$)

$$\mathcal{L}\{f : \nabla\mathbb{S}(t)\} = \int_0^\infty e^{-st} f : \nabla\mathbb{S}(t)dt$$

This maximises O's model evidence for its prediction E, getting its predictions E into an *objective* comparison with $\boldsymbol{E}$ so that it can get to $\mathbb{S}$. If we think of O's behaviour as modelling its predictions, we can understand that this quantity is the total change in evidence for its world model (hence "model evidence" above) following its action. This model is frequently called a "generative model" as it generates the evidence the organism relies on to decide its next actions. We will denote it as m; note that while it exists objectively, there is no need for O to be aware of more enough to enable appropriate actions, so $E \subseteq E$.

This model evidence is F or F, depending on the point of view. For the time being, let us assume O is getting it right, so $F = F$. If O has leached every possible benefit from its environment by its action, then $F = 1$, but with no environmental support $F = 0$, and O dies. It now makes sense to call this model evidence “well-being” but note that “well-being” inverts the value of $\mathbb{S}$: well-being F is maximised when $\mathbb{S}$ is minimised.

Surprisal

Compare a beloved goldfish in its aquarium with a mackerel being grilled for tea. In terms of our model, we can hope, for the goldfish, that $F \approx F \approx 1$. For the grilling mackerel, it is clear $F = 0$, but F simply does not exist, as there is no longer a functioning mackerel to model it. As the illustration shows, organisms can only minimise $\mathbb{S}$ up to some bound; what lies beyond that is called “surprisal.” Surprisal is why severe burns may be less painful than minor ones: with pain and temperature receptors gone, we can no longer model what has happened to us. Mathematically, surprisal is $\mathfrak{S} = -log(prob\boldsymbol{E} = E)$ so it grows as the environment becomes more hostile, but it can never be incorporated into F, as that refers to *model* evidence, and surprisal, by this definition, cannot be modelled.

Limits to modelling

There are more limits to O than E. Implicit in minimising $\mathbb{S}$ is O updating itself accurately in the light of new information. This involves the prior information O has, the structure of the model m it uses, and the parameters Θ in that model, which define the dependencies between the model’s variables. The number of variables in $\boldsymbol{E}$ is typically enormous, and the parameters of all possible interactions $\Theta \sim {}^{\boldsymbol{E}!}/_{E!\,(\boldsymbol{E} - E)!}\,|O$ potentially much larger still, so $\int \Theta$, necessary to determine $\mathbb{S}$, is intractable. However, we have seen O only needs to model up to a boundary, which permits a variational approach to updating. The Laplace approximation to a Gaussian distribution enables an approximate factorisation of the distribution of interest $q(E, \Theta) \approx q_E(E)q_\Theta(\Theta)$. If E is iid, all E can be treated as mutually independent, while $N(\Theta_E) = N(E)$. The price paid is ignoring how fluctuations in E require fluctuations in Θ to accurately model $q(E, \Theta)$. Laplace’s approximation also greatly simplifies finding integrals—here $q_E(E)$—by allowing variation distant from a global maximum to be ignored. This lets a (computationally tractable) variational Bayesian approach find $-\mathbb{S} = F$ (Buckley et al., 2017), but the maximum F identified may not be global. So, these approximations and assumptions also contribute to surprisal. As this surprisal reflects computational limits rather than environmental incompatibility, it is subjective & denoted as $\mathfrak{S}$. Despite being subjective, it is still outside F. It is usually (imprecisely) referred to by its negative, i.e., what we can comprehend given our cognitive limits—satisficing, denoted as $-\mathfrak{S}$.

Scaling Well-Being

Organisms function at several scales. We, for example, are composed of organs, which are composed of cells, which have their own organelles, each with its individual biochemistry. Well-being must be expressed across all those scales, and each level must make its separate contribution to total well-being F, independently and interactively. The generative model m is therefore hierarchical. From the above, we can write

$$F = \lim_{\mathbb{S} \to min} \left(\prod_{1}^{m} F_i \to 1 \right)$$

For any individual level

$$F_i = E_i R_i F_{i-1}$$

To understand the new term R, consider the relevant Laplace transform $\mathcal{L}_i\{f: \nabla\mathbb{S}_i(t)\} = \int_0^\infty e^{-st_i} f: \nabla\mathbb{S}_i(t)dt$. Remember that $\lim_{\mathbb{S} \to min} \sim F|O$ and F is being measured at its maximal, steady state. R_i is, therefore, the integral of e^{-st_i} with respect to time, $e^{-st_i}/s + C$. $1/_S$ is not level specific, so it may be applied recursively across all levels. The constant of integration, C, can be interpreted as representing all the unknown structures and processes O uses to generate e^{-st_i}, allowing C to be proportionally scaled as 1. R will also behave as a normalising constant, as the probability for F at steady state is also (approximately) 1. Therefore, it represents the organism's biology from the organism's perspective and sets its upper and lower scale boundaries. Beyond the upper boundary lies the organism's environment, and beneath the lower is the chemistry from which the organism arises. We can think of these boundaries as phase transitions in how informational free energy is embodied across scale.

A problem estimating well-being arises at the lower boundary $m = 1$, as $F_i = F_0$ does not exist, stopping evaluation of F_1 and thus the whole system.

Recalling $R_i = e^{-st_i}/s + C$ and taking logs, we can write (assuming $C = 1$ as above)

$$\lim_{F\to 1}\left(\sum_{1}^{m} \log E_i + \log F_{i-1} - st_i - \log s \to 0\right)$$

For $m = 1$, we can write $\lim_{F\to 1}(\log E_1 = st_1 - \log s)$. So, not only may an organism estimate well-being even at its lowest level from its biology, but organisms without internal hierarchy are possible, too, while s is a frequency parameter, so it can coordinate levels in m across temporal and spatial scales.

Environment, Individual Identity, and Action

Our consideration of how we need to measure our well-being, especially across scale, leads us to a novel interpretation of what it means to be ourselves. We are not material individuals in an environment, but a nested set of environments, structured and coordinated by some shared rules, which also define the everyday boundary between ourselves and our world, and a less familiar boundary with the small-scale world of biochemistry. Each nested environment, while following the recursive rules expressed in s and embodied in C, functions to a different timescale t_i. The ability to process information to ensure survival is not just a function of our nervous system but is a fundamental property of any living system, so implies similar nesting of our subjectivity, which is no longer coterminous with our awareness.

All this sounds strange and alien, but we have an everyday illustration available in our own blood. At our scale, it looks like a thickish liquid that comes in various shades of red and brownish purple. Put a drop under a microscope, and we see a liquid environment containing white and red blood cells. Some of the former seem to have, at the very least, more awareness than the latter, displaying what looks like purposeful intention as they engulf bacteria. Drilling down further, we see sets of chemical reactions organised by a micro-architecture that seems purposefully arranged. While it is easy to observe the difference in timescale here—e.g., macrophages engulfing bacteria takes seconds to minutes, but a successful immune response may take days to complete—more is needed to understand how it happens.

From a purely physical perspective, an organism, like its environment, is just a bunch of stuff moving around. So, like all other stuff in our universe, a mechanical model may describe its behaviour, and the Lagrangian one is helpful here. Especially, its "principle of least action" is a thermodynamic equivalent to the loss function discussed above, which enables the descent/ascent (equivalent here) of *informational* gradients to be accurately embodied by *thermodynamic* ones.

The basic Lagrangian formalism is very simple, with a numerical summary of the total dynamics of a system, called the Lagrangian (L) being the difference between the system's total kinetic energy (T) and potential energy (V); $L = T - V$. As T includes velocity and therefore implies motion, V can be conceived as a gradient (slope or potential) in any space (or field) necessary for motion (or its equivalent, e.g., current) to occur. If the beginning and end of the motion are points in time, the path between them is called an action (S). There are many possible paths connecting two time points. The path of least action is the most direct one. We can thus conceive of O's continued existence as an action and its environment as a kind of landscape, where it needs to find the most direct route to each and every future point that best sustains its existence. We can formalise this path as $S = \int_{t_1}^{t_2} L dt$, where $\delta L = 0$. O finds it through embodying the equivalent informational gradient— $\min_{F \to 1}(\mathbb{S} \approx \mathbb{S}) \equiv (\delta L \to 0)$ —as discussed above. When, like us, O is hierarchical, $S = \prod_m S_i$, with each integral nested up to m. So, superordinate integrals, being integrals of integrals with respect to time, will define longer timescales. If we now focus on the kinetic energy T associated with each action S, the longer timescales of superordinate actions will make them seem like gradients V to their subordinate components. So hierarchical development allows more sophisticated actions by letting O manipulate both T and V at multiple scales, using its objective biology R.

Rhythm, Control, and the Non-Equilibrium Steady State

Organisms are one example of Non-Equilibrium Steady States (NESS), systems that maintain themselves by constant correction, though remaining far from their environmental equilibrium. If we think about the Lagrangian associated with that, it will need to follow a closed route, which, extended in time, will form a sinusoidal path as it drifts away from and then moves back towards optimal conditions to maintain the NESS. Breathing is one example. Embodying S provides adjustable frequency parameters to deliver this. In reality, δL represents random perturbations, not a regular rhythm. However, as the Fourier transform demonstrates, any irregular waveform can be represented by a superposition of enough regular ones, and in hierarchical organisms, m's nested levels can provide these (Rue, Martino, & Chopin, 2009).

Empirical Assessment of Well-Being

In everyday language, "well-being" denotes individuals flourishing (or not) in their environments. As defined by our model, embodied well-being represents us as the uppermost level in the nested environments of O. Let us denote this level as I (for individuals) and its associated well-being as F_I. There must be an F_{I-1}, which fits our understanding of "health" as it estimates the well-being of those nested environments that, together and connected by R, make up our bodies.

From the arguments above, F estimates the usable fraction (E) of $\boldsymbol{E}$ available to O, so it describes a utility U. In health economics, the health-related version is Health-Related Quality of Life (HRQOL), which is measured by differences in health-state preference, with a presumption that health will be preferred. It may be directly estimated by evaluating how much loss would be equivalent to a particular suboptimal health state, which is also $\max_{\boldsymbol{S}}(\boldsymbol{S}(Q|\{X_{loss}, X_{health-state}\}) = -K\sum_i p_i \ln p_i) = -F_{max}$ as there is, at estimation, no information to allow preference of one over the other. This utility is scaled to 1 being perfect health and 0 death (Payakachat, Murawski, & Summers, 2009). We can therefore state $F_{I-1} \propto U_{HRQOL}$, as they are derived equivalently. Defining well-being as embodied model evidence has turned out to include everyday health-related well-being, which latter requires no reinterpretation.

HRQOL is not considered qualitatively distinct from any other utility considered in economics (Payakachat et al., 2009). We also have an everyday word for the set of affordances we need to survive as individuals; they constitute our welfare. If we accept that HRQOL is what we understand as "being healthy" in everyday language (which implicitly sets $m = I$), we can therefore say $Well-being \sim (Welfare \times Health)|R$, or, forgetting R as everyday language is wont to do $Well-being = Welfare \times Health$. In its complete form, we can express this as previously

$$\lim_{F \to max} \left(\prod_{Health}^{Welfare} E_{Welfare} R F_{Health} \to 1 \right)$$

We can see that our model does not undermine or redefine our current measures of well-being. Instead, it clarifies the otherwise vague concepts, such as "welfare", "health", or "utility", which underpin them, in terms of our necessary biology.

From Well-Being to Psychopathology

Death, Disease and Disadvantage

It is essential to check if our imagined model can cope with these if it is to include psychopathology. From the above, we can write, from the perspective of an observer only

$$F_{death} = U_E = 0$$

As O is dead, there is no perspective to create R. We can say that the environment provides no utility to our dead O, but we cannot observe E either, because the organism's conditional independence from its environment is gone—the dead mackerel cannot resist being grilled. The multiplication rules say that any $U_{E_i} = 0$ is sufficient to determine death. An everyday example is $U_{E_{I-1}} = 0$, which is the same as saying $HRQOL = 0$. Death can therefore be defined as an instance of a level-specific environment (or more) being unable to provide sufficient utility to $\prod_1^m O_i = O$, which matches everyday experience.

It makes sense to say that disease is something intermediate between life and death, as modelled by HRQOL. Our model can cope with this, by setting at least one $U_{E_i} \ll 1$. Impaired welfare, or disadvantage, now becomes a special case of disease, where $U_{E_i} = U_{E_{I+1}}$, the external environment to O. It is "special" because, being outside R, our ability to regulate V (the gradients in our environment) can only be achieved by actions expressed through T_I. From this, we may draw the unsurprising conclusion that there is no clear boundary between ordinary ($U_E < 1$) and impaired ($U_E \ll 1$) health or well-being. Our model is consistent with our ordinary intuitions about these. However, there is a caveat, discussed next.

The Biopsychosocial Model Fails for Well-Being

The biopsychosocial model remains the most popular way to understand mental health. Because m_i is associated with a unique U_{E_i}, $F_{death} = U_E = 0$ can test whether a proposed level m_i is valid. The model can be formalised as

$$F_I \sim \left(\prod_{1}^{I+1} m_i \, , m = \{ m_{social} \succ m_{psychological} \succ m_{physical} \} \right) | R$$

The physical level precedes, so is nested within the psychological, which is nested within the social level. We situate our identity in the middle layer. The biopsychosocial model makes no distinction between subjective and objective perspectives, so we treat it as objective (from the observer's perspective). From this, we deduce a set of objective level-specific environmental utilities $\{U_{E_{physical}}, U_{E_{psychological}}, U_{E_{social}}\}$. We do not have to consider ordering, as any product including a zero will result in death, so disagreements over the order of nesting make no difference. Setting $U_{E_{physical}} = 0$ will lead to death, as that is HRQOL. However, coma or anaesthesia set $U_{E_{psychological}} = 0$, without death being in any way inevitable. So, it seems that $m_{psychological}$ is not a valid level in our model. Instead, we might want to say $m_{psychological} \subset m_i$, though we cannot yet say what the whole of m_i encompasses. A similar problem arises with and $U_{E_{social}} = 0$ though the latter is beyond this chapter's scope. We consider alternatives to $m_{psychological}$ next.

Imagining a Biologically Valid Psychology

The Dimensions of Subjectivity

We have already seen that subjectivity is the inevitable consequence of O needing to traverse its environment while maintaining its Markov blanket. We have acknowledged that F is information-theoretic, so it needs embodiment. We define our assessment of our well-being as $\{W\}$ as there are many ways that we can monitor our well-being. A helpful analogy is to think of a tool. If we liken F_i to a ruler, then W_i is the ruler we happen to own, with all its virtues and flaws. We can state $\{W\} = \{\mathcal{F}Fe\}$, where $\mathcal{F}$

denotes the transformation our measurement tools perform, so is the subjective embodiment of $\mathcal{L}$, and e is the error they introduce. As $F = \prod_1^m F_i$, we will need at least m subjective scales W_i to evaluate F. By substitution, we can say $W_i = \mathcal{F}_i F_i e_i$ as each W_i is likely to need a different transformative function and, therefore, will have a different error. We can therefore write, for a simple 2-level case, such as welfare and health, and substituting utilities for "objective" well-being to improve clarity

$$W = \prod_{Health}^{Welfare} W_i = \{W_{Welfare}, W_{Health}\} = \prod_{Health}^{Welfare} \mathcal{F}_i U_i \, e_i$$

We may scale W between 0 and 1, but $\max_{F \to 1} W < F$ because of e_i, and $\mathcal{F}_i$ is not recursive: the rules that embody R could, and indeed need to, produce many different forms of $\mathcal{F}$. This is consistent with our intuition—we use different vocabularies to discuss each, and our subjective awareness of our health is quite different from our welfare. This is despite measuring level-specific utilities U_i, which are simply components of the total utility U_E that O needs to survive. While this is "error" in terms of measuring U_E it is not mistaken, as it needs to match the E_i it is monitoring. For example, we cannot use our ruler to measure the distance between stars, but using a telescope instead gives output that is qualitatively distinct from that of the ruler, despite referring to the same metric of length.

We can now write, again substituting utilities

$$\begin{aligned} W_I \sim & (\mathcal{F}_{Welfare} U_{\boldsymbol{Welfare}} e_{Welfare})(\mathcal{F}_{Health} U_{\boldsymbol{Health}} e_{Health}) | R \\ & = \mathcal{F}_{Welfare} \mathcal{F}_{Health} \left(U_{\boldsymbol{Welfare}} e_{Welfare} U_{\boldsymbol{Health}} e_{Health} \right) | R \end{aligned}$$

The qualitative differences necessary to measure utility across our nested environments have created a multidimensional (in this case, the two dimensions are $\mathcal{F}_{Welfare}\mathcal{F}_{Health}$) space within which the total utility, and hence well-being, can be measured. Notice, as discussed above, that the multidimensionality of W_I encodes the transformation introduced by $\{\mathcal{F}_i\}$, so does not apply to U_I. Recall that monitoring our well-being is to shape our behaviour appropriately. We generalise $\mathcal{F}_{Welfare}\mathcal{F}_{Health}$ to $\prod_1^m \mathcal{F}_i$, as we might well be monitoring more levels of which we are not subjectively aware. $\prod_1^M \mathcal{F}_i$ is, therefore, a space in which well-being can be represented subjectively: Ψ. This gives us $W \sim U\Psi \prod_1^m e_i \,|R$ Think back to our ruler and telescope analogy. The distances they measure comprise U, encoded as F (remember, obtaining utility U from environment E is why organisms need to measure F in the first place). The difference in processing needed to obtain each measurement generates the qualitative differences that set the dimensions of Ψ, while the total within-measurement error is $\prod_1^m e_i$.

The subjective space Ψ we have just outlined is a superset of what is usually considered "psychological" in the biopsychosocial model. This becomes clearer if we consider that this definition of subjectivity extends to things that we would not typically attach subjectivity to, due to scaling. Consider a white blood cell. In its environment, it faces the same task of self-evidencing as we do in ours, and, as we have seen, subjectivity to some degree is an inevitable consequence of Active Inference to maintain a NESS Markov blanket. While this may feel odd, it fits with modern philosophical developments in intentionality (van Hateren, 2020). A helpful analogy is that of "Communication, Command and Control" (C3) in military and cybernetic studies. We can think of the Central Nervous System (CNS) as just one part of a more extensive C3 system, perhaps including endocrine, metabolic, and immune control systems. If we consider $m_{\Psi\equiv\Psi}$, it is evident that depriving that level of what it needs ($U_{\Psi\equiv\Psi} \to 0$) will result in death, as organised function will be lost, leading to failure to maintain O's Markov blankets. So, while $O_{psychological}$ is not a valid level in our model, O_Ψ is.

The Properties of Psychological Subjectivity

Let us concentrate on well-being estimates we evaluate psychologically. First, we define $\psi \subset \Psi$ to represent this subset, making ψ everyday subjectivity. We can already deduce some properties of ψ.

1. The fabric of ψ is dimensional.
2. The whole of ψ—both dimensions and the objects they constitute—is subject to updating, so it will change with time.
3. The data that ψ encodes is composited, with both a predictive component generated subjectively, and an error component generated comparatively.
4. All evaluations will include a preference dimension that will reflect their utility.

This sets out the formal properties of ψ, including ascribing a common nature to its contents. Note that O_ψ is no longer commensurate with other levels of O due to $\{\mathcal{F}\}$ so it cannot scale to them. Such a space is apt for phenomenology and fits our conception and experience of "mind".

Defining the Contents of Mind

W estimates F, but we have not yet discussed how that is achieved, beyond specifying some function $\mathcal{F}$. As $F = -\mathbb{S} \approx -\mathbb{S}$, all our estimates are (minimised) predicted differences. However, maximising well-being may entail either approaching or avoiding some state of affairs, and these will be instantiated differently. Imagine O is confronted by a dangerous fire. Putting the fire out and running away from it both minimise $\mathbb{S} = D_{KL}(E \parallel E)$ but the first modifies this through acting on $\boldsymbol{E}$ and the second through acting on O. However, internal sensory information is generated in anticipation of experience and can be detected through interoception (Barrett & Simmons, 2015), which imputes a potential *E*, not a real *E*. Empirical demonstration of this has been achieved for hunger and thirst (Livneh et al., 2020). This difference is $\hat{F} = -\hat{\mathbb{S}}$, the circumflex indicating the estimated prediction error has arisen entirely from O, without E. Consider "Willingness To Pay" (WTP) an amount of money for a particular good: two items with the same WTP would be considered to contribute equally to well-being. This technique is used to evaluate HRQOL, for example, as discussed above. However, in HRQOL estimation, WTP is often to avoid a sickness, disability, or similar. If the person has experienced the illness, witnessed it, or even read a description, it is probably safe to assume $\mathbb{S} = \hat{\mathbb{S}}$ and this presumption underlies much WTP usage in practice as a tool to evaluate consumer preference for new goods (e.g., Taylor & Armour, 2002). But, without external input from *E*, $\hat{\mathbb{S}}$ can go wrong, as the story below illustrates.

> A man was seen covering his garden with rolled-up newspaper. When his neighbours asked him why he replied, "It's to keep the tigers away." Puzzled, his neighbours assured him there were no tigers where they lived. "That's right," he replied, "It's really effective."

We may see how this arises in Ψ by considering $\{W\} = \{\mathcal{F}Ue\}$. Our man and his neighbours have similar well-being estimates $\{W\}$; both were quite content in their lives until the conversation in the story. Similarly, there is no appreciable difference in $\{e\}$; all agree on the absence of tigers and the presence of a garden covered with rolled-up newspaper. To explain the difference, we must presume that our man additionally possesses some functions $\{\mathcal{F}_{Man}\}$ to allow him to deduce a necessary and sufficient connection between his behaviour and the local absence of tigers. We also need to distinguish the different classes of evaluation discussed above. For $F = -\mathbb{S}$ the term "direct" emphasises that it has been evaluated with direct reference to E. For $\hat{F} = -\hat{\mathbb{S}}$ the term "imputed" implies E has been predicted from interoceptive anticipation, so distinguishing a current situation we predict from some postulated state we seek.

The Roots of Emotional and Imaginary Thinking

Let us assume $\{W_i\}$ generates a set of estimators $\{\omega_j\}$. The use of Greek lowercase reminds us that we will experience these estimators within Ψ, and we assume $i \leq j$. Some of these estimators will be of $\hat{\mathbb{S}}$, where the action minimises *imputed* free energy. Subjectively, W will scale to unity, as we have seen above. We can therefore write $\{\omega_j\} = \{\omega_{j-k}, 1 - \omega_k\}$, where $\{\omega_{j-k}\}$ are direct estimators, and $\{1 - \omega_k\}$ are imputed estimators. However, this leaves the unsatisfactory position that sometimes we estimate directly and sometimes by imputation, without explaining how this can happen or how we can tell which is which.

The goal of either estimation is to identify the utilities $\{U_E\}$ we need from our environment. Put another way, the environmental variables providing those utilities $\{V_O\} = E$ are salient to us, and to evaluate them, we must express preferences. From this perspective, O is behaving as an investor in its environment; it is betting its predictions against what its environment will afford, e.g., oxygen when it respires or nourishment when it eats. Economists observe that these preferences can arise either from rational deduction or emotional state, which they call "sentiment", and its separate existence as a subjective predictor of value may be reliably deduced (e.g., Ryu, Kim, & Yang, 2017). The equivalent term in psychology is "valence" (cf. Shuman, Sander, & Scherer, 2013). Active Inference can successfully model valence, at least in silico (Hesp et al., 2019). A full discussion of that model is inappropriate here, but the functions which deliver valence are bounded within specific hierarchical levels, while Shuman et al. (2013) attribute a hierarchical structure to valence itself, according to its level of complexity. From our principle that levels in O are represented as dimensions in Ψ, evaluations via sentiment will be represented by a different order of dimensions to cognitive evaluations. As Hesp et al. (2019) and Ryu et al. (2017) point out, valence/sentiment is expressed as confidence in belief. At the neurocognitive level, this is the precision of the predictive model generated; at the economic level, it is expressed by several functions of the volume and share closing prices with respect to time. We notice that at either level, valence/sentiment can generate values for $\hat{\mathbb{S}}$ though, of course, that value may be mistaken. I suggest that the imputed estimators $\{1 - \omega_k\}$ provide the *affective* dimensions of our subjectivity, which we typically experience differently from subjectivity's cognitive components. This is consistent with empirical work on the relationship between emotion and interoception. For example, the absence of the ability to express emotion, alexithymia, has been related to interoceptive failure (Murphy, 2017). Interoception also links emotion, time (which we have discussed above), and music, which are mutually related to normal and psychopathological states (Droit-Volet, 2013; Juslin & Sloboda, 2013; Vicario, Nitsche, Salehinejad, Avanzino, & Martino, 2020).

We are conscious of at least some of ψ. Emotion is conscious sentiment or affect, whose creation we have discussed above. Being conscious of at least some subjective predictions without engagement with error-correction from sensation also allows for a class of *imaginary* objects in our subjectivity. Unlike emotions, they are unconnected with sense-data or proxies, so no prediction of $\hat{\mathbb{S}}$ will be made, i.e., they do not imply action. We may deduce four classes of construction in our subjectivities.

a) Dimensions. These encode the estimators of our level-specific environments.
b) "Real" objects. These arise from imaginary objects composited with sense data.
c) "Imaginary" objects. These are predictions resulting from our generative models, which have not been composited.
d) Emotions. These are estimates of the utility of potential states created by imputation.

In short, we have been able to deduce the familiar contents of our everyday subjectivity from applying FEP to well-being.

From Errors in Self-Evidencing to Psychopathology

Let us return to our man and his unorthodox approach to tiger prevention. It follows from the above that our psychopathological interest centres on $\{\mathcal{F}_{Man}\}$. We assume, of course, that he is wrong, so $\{\mathcal{F}_{Man}\} = False$. This lets us imagine a superset of faulty functions generating false dimensions $\{\tilde{\mathcal{F}}_{Man}\} \subset \{\tilde{\mathcal{F}}\}$. Can $\{\mathcal{F}\}$ account for the psychopathology we observe in reality?

Consider a single errant function $\tilde{\mathcal{F}}_i$. Psychopathology refers to some aspect of subjectivity that is not part of our shared reality. So, it follows that it will be informed by (at least) one incorrectly imputed estimator $\tilde{\omega}_{k=i}$. The effort wasted generating this imposes a net loss to U_{E_O}, so psychopathology will always reduce F and be harmful. In our imaginary example, our man has spent useless time, money, and effort accumulating and spreading rolled-up newspaper around his garden, so incurring, at the very least, an opportunity cost. Were his neighbours unable to persuade him otherwise, it would be reasonable for us to assert that he is deluded. The difficulty we face is that we have defined imputed evaluations as emotions, and delusions are not understood as such. However, this objection can be overcome.

Knowledge is classically defined as "justified true belief". We need to be clear that the man's mistake is believing there is a causal connection between paper (presence) and tiger (absence). His justification is plausible, just incredibly unlikely compared to the alternative: it is the plausibility that adds the humour. The falsity is because the man's level of belief has been set at "certainty", and certainty is an emotion. So, delusions are inappropriate compositions with certainty. We can predict a subclassification depending on whether the composition was with real, imaginary, or emotional objects, and this can be found. Delusional perceptions are typically "two-limbed"; an actual occurrence leads to a delusional deduction, e.g., personal messages are perceived in public radio messages. There are emotional delusions of love (de Clerambault's Syndrome) or jealousy (Othello Syndrome). Paranoid delusions create purely imaginary webs of conspiracy. As they share a common misattribution of certainty, it is no surprise that they can co-occur. Objectively, FEP manages this well. We have seen above that emotion sets the precision associated with the predictions arising from FEP's generative models. Inappropriate certainty could either arise from over-precise predictions or impaired error correction—the latter is implicit in inappropriate composition. Paulus et al. (2019) both describe supporting evidence and explain how, once established, such errors do not automatically correct themselves.

The idea of inappropriate composition can be extended to explain hallucinations. The classic definition of a "sensory perception occurring without external stimulus" strongly suggests an imputation being inappropriately applied to an imaginary prediction. Schizophrenia is the obvious empirical model for hallucinations and delusions, as both psychopathological phenomena are found in it, and there is empirical support. The insula coordinates cognitive, emotional and somatosensory processing. Sheffield et al. (2020) describe reduced differentiation in resting-state functional connectivity within the insula in a large sample of schizophrenic patients, which is consistent with a hypothesis of increased error in composition.

While not an exhaustive list of all psychopathology, hallucinations and delusions are the prototypical exemplars of disordered perception and disordered beliefs, which covers much. In this model, both can be explained as the product of mistaken compositions of imputed error correction and subjective predictions in the ongoing process of self-evidencing by Active Inference at specific levels of biological organisation. I have not explicitly discussed the psychopathology of emotion, but I have already referenced relevant accounts framed within FEP (e.g., Paulus et al., 2019) and my account of well-being is no more than a subjective application of this theory. Therefore, it seems likely that the mechanism of inappropriate composition (by either presence or absence) will generalise across psychopathology.

Psychopathology, Surprisal and Satisficing

Astute readers may have noticed that the discussion of psychology above did not mention surprisal. This is because our psychology is a subjective expression of model evidence, and surprisal lies outside this. However, we can now understand psychopathology as a level-specific, deleterious change in our ability to generate accurate model evidence, and so signals an unwelcome increase in surprisal. All the complexities just described are to optimise $\mathcal{O}$'s path of least action $S = \int_{t_i}^{t_2} L dt, \delta L = 0$ to maintain its integrity, and so survive. The effect of psychopathology is to reduce fit or increase bias, so that $\mathbb{S} = \delta L > 0$. Because surprisal is also level-specific, the same surprisal may be represented as either $\mathfrak{S}$ (when V is an environmental property relative to level m_i), or $\mathfrak{S}$ (when the limit is computational). Recalling that psychopathology reflects a space $\{\tilde{\mathcal{F}}\}$ enables us to distinguish it from both $\mathfrak{S}$, e.g., a sound that destroys our ears cannot be heard & $-\mathfrak{S}$, which is evidenced through phenomena such as optical illusions. Objectively, $f\colon \left\{\tilde{\mathcal{F}}\right\} \mapsto \mathfrak{S} + \delta\mathfrak{S}$, so increasing the risk of both disability and death. This makes psychopathology qualitatively distinct from psychology, not in terms of its content, but in its forms, as described by Jaspers (1963).

Conclusions

It has been said that the greatest creation of the human brain is the real world. However, the world it builds for us is to aid our survival, and our awareness of it is only a means to that end, so it is not surprising to find that ideas based on our subjectivities alone, such as classical psychopathology, can be misleading. The biopsychosocial model has turned out to be, quite literally, a figment of our imaginations, as the separate orders it proposes result from necessary but illusory encoding of levels of biological organisation. Apparently obvious distinctions, such as those between "thought", "emotion", "reality", and "imagination" are anything but, and psychopathology signals when the processes which construct them break down. Our individual identities and environments are matters of perspective, while our "constitution" is an embodiment of information theory applied to survival. We are to our kidneys what our world is to us, while the social science of Macro-Economics stops at our households, as we do at our genetics, and we could, theoretically, also exist in any computer which could re-create all our Markov blankets in silico. We have even found that we may have over-privileged our brains as commanders of our lives, and other C3 systems need to be included in our understanding. However, with those corrections made, we have also found it possible to explicitly link our biological processes and the subjectivity we recognise in both its normal and pathological expressions. Further work elucidating that link will be challenging, but at least it need be mysterious no longer.

References

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950