Architects of Our Own Experience:
The Structure of Human Freedom

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Arts in Philosophy
at
The University of Waikato
by
JAY EVANS

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Abstract

The picture of a deterministic universe is difficult to reconcile with the phenomenology of agency, deliberation, and choice. While everyday conceptions of individual freedom and responsibility depend on the pre-theoretical assumption that we have meaningful control over our decisions and volitional actions, if determinism is true, the outcome of any deliberative process is always the only outcome available; the inevitable result of a reasoning sequence that could not have proceeded any other way. However compelling our ordinary sense of agency and control is, determinism gives us reason to doubt that we have the kind of freedom that we intuitively take ourselves to have.

The debate over whether free will is compatible with determinism is a long-running one, traditionally motivated by, and entangled in further disputes about what kind of freedom or control is required for moral responsibility. In a promising and refreshing turn, recent work by philosopher of physics, Jennan Ismael, has significantly advanced the debate. Leaving aside questions about moral responsibility, Ismael's naturalistic account shows that the freedom we ordinarily take ourselves to have is not threatened by deterministic physics; when determinism is understood on its own terms, stripped of imposed notions of compulsion and necessity, the purported tension between freedom and determinism is resolved. Furthermore, Ismael argues, biological systems with the capacity for representational thought have evolved unified, emergent 'selves', and the capacity for self-governance, behavioural flexibility and deliberative control.

Following Ismael's methodology of naturalistic pragmatism, I aim to enrich her contextualised picture of human agency with a precise, unified account of the nature of agency itself, and to further develop this integrated account of freedom and the structures that underpin freedom. Building on Ismael's account of the self-governing free agent, I aim to fill out some of the details of the cognitive structures that underpin our first-person experiences of choice and endogenous control of action, and discuss how those structures arise in the natural order. Using the predictive processing framework to explore the structures of agency, choice, and autonomy that underpin self-governance, I argue that our potential for freedom – freedom in its most fully developed manifestation – lies in our capacity for autonomy, which is realized when we act deliberately to modify our own cognitive processes and transform our own minds.

Acknowledgments

I am immensely grateful to my supervisors, Joe Ulatowski and Justine Kingsbury for their brilliance, insight, and warm encouragement. Their guidance, feedback, and support over the course of this project and in their classes have been enormously valuable, and it has been a privilege and a joy to have been able to learn from them. My appreciation also goes to the faculty of the Philosophy department at the University of Waikato, whose stimulating courses and expert teaching laid the ground for me to work on this project. Many thanks also, to the University of Waikato for providing financial support for this project, in the form of a Research Masters Scholarship. Finally, my personal thanks go to Sarah, whose ongoing love and support make all the difference.

Table of Contents

1. Introduction

A sense of agency and volitional control over our actions is built into ordinary, everyday human experience. From this perspective, the distinction between voluntary and involuntary actions is intuitively obvious. Actions we recognise as voluntary are accompanied by a familiar sense of endorsement, intention or desire that involuntary actions lack; we recognise immediately, for example, the difference between running and falling, between listening and hearing, between actively holding our breath and our unmodified breath pattern. The sense of internal control we experience over the movements of our own bodies extends to complex sequences of thought and action. We have a strong intuitive awareness of ourselves as agents, of being (or perhaps of having) a locus of decision-making and control – a psychological standpoint from which we deliberate, choose, and initiate actions. When we deliberate, it is with the implicit understanding that we can, from this psychological standpoint, exert meaningful influence over the course of events – that our choices and actions have effects in the physical world, and those choices are genuinely up to us. Pretheoretically, we accept the experience of deliberation and choice at face value; we take it for granted that the alternatives we conceive for action are (within the constraints of our social and physical environments) realisable, and that we are genuinely free to make the judgements and choices that guide our actions.

However, this everyday experience of freedom appears to conflict with the facts about physical determinism[1], according to which the state of the world at any given time, together with the laws of nature, necessitates the state of the world at any other point in time (Hoefer, 2023; Mele, 2013). Determinism underpins the assumption of *causal completeness* – the understanding that every event has a complete causal explanation – a sequence of antecedent causes that fully accounts for it (Hoefer, 2002). If this is the case, then all events, including human choices and actions, arise as the inevitable result of a chain of prior causes. In this picture, every event is necessitated by its causal antecedents, there is no possibility of acting outside of necessity, and the course of events is fixed in an unalterable sequence that extends

---

[1] Before the discovery of quantum mechanics, it was thought that the universe is entirely deterministic. This is no longer the case; the currently known laws of nature are deterministic, along with some quantum randomness. While it may eventually be shown that the universe is ruled by indeterminism, quantum mechanics is not yet fully understood, and may itself turn out to be deterministic (Hossenfelder, 2022). As it stands, the project of reconciling freedom and determinism remains relevant and worthwhile, and I will assume determinism in the discussion that follows.

back to the origin of the universe. There is no place for freedom in such a universe – to act freely would be to insert a supernatural event into the causal chain. The phenomenology of deliberation and choice seems to be in stark contradiction to a scientifically grounded view of reality.

If human choices and actions are necessary consequences of the universe unfolding according to the laws of nature, traceable, in principle, to the initial conditions of the universe, then our default experience of agency and volitional control over our own actions is illusory. Although we might accept that deliberated choices causally contribute to action, those actions and choices are not in any sense *free*; the outcome of any given deliberative process is always the only outcome available – the inevitable result of a reasoning sequence that could not have proceeded in any other way. If this is the case, although volitional and non-volitional actions are phenomenologically distinct – that is, they present differently in conscious awareness – the difference between them is not a matter of freedom. When we weigh our options, whether for straightforward, trivial decisions or weighty, complicated ones, there is nothing at stake – we are simply bound to deliberate, choose, and act exactly as we do.

The implications of determination for free will are articulated succinctly by Peter van Inwagen's Consequence Argument[2]. Jennan Ismael's formal reconstruction of the argument is as follows:

1. Dynamical determinism entails that the facts of the past, in conjunction with the laws of nature, entail every truth about the future.
2. The past is not under our control.
3. Laws of nature are not under our control.
4. Our actions are entailed by past and laws of nature.

Hence,

5. Our actions are not under our control. (Ismael, 2016, p. 88)

---

[2] Informally, van Inwagen (1983) states the argument as follows: "If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us" (p. 16).

Jennan Ismael’s (2016) response to the consequence argument, and her naturalistic account of free will, provide the starting point for the discussion that follows. Ismael’s work represents a radical advance in the project of reconciling free will with determinism, and is a significant departure from traditional approaches.

Much of the discussion about free will is underpinned by disputes about moral responsibility. Both compatibilists and incompatibilists tend to take the assumption of moral responsibility as a starting point, and take free will to involve the kind of control over our choices and actions that is necessary for moral responsibility. The dispute gives rise to two different requirements for free will: an alternative possibilities requirement, and a sourcehood requirement. According to the alternative possibilities requirement, freedom requires that the agent be facing genuinely realisable alternatives, and have the ability to do otherwise. As incompatibilists point out, this is difficult to reconcile with the deterministic picture in which there is only one possible course of events. The sourcehood requirement merely requires that the agent is the source of their own actions – that is, that their actions are uncoerced and connected in the right way to their decisions (O’Connor & Franklin, 2022).

In my view, while questions about moral responsibility are interesting and worthwhile in their own right, this approach seems to have things the wrong way around; we will be better positioned to answer questions about moral responsibility if we first understand the nature of freedom itself. In what follows, I will not have anything to say about moral responsibility, or the attribution of praise or blame, pride or shame. Rather than locating an ability or set of conditions that justify our moral intuitions about praiseworthiness and blameworthiness, I aim to outline a naturalistic account of the nature and scope of human freedom.

Nevertheless, both the sourcehood and alternative possibilities requirements do closely align with our pretheoretical intuitions about freedom of choice and action. In the common sense view, freedom requires that we ourselves are the source of our actions and choices; the free agent must be genuinely capable of acting on their own behalf, and not merely be an instrument of environmental forces. Moreover, freedom requires that our actions and choices are not predetermined by realities that we did not choose – there must be a meaningful sense in which there are available and realisable alternatives for action. Together, these two conditions require an agent who is capable of decision making (non-mechanical influence over its own behaviour) and a context in which there are multiple realisable possibilities. That is, the possibility of freedom is dependent on the nature of the agent, and the nature of the broader contextual reality of which the agent is a part. Both of these aspects – agent and world – are addressed in Ismael's (2016) naturalistic account of free will.

In *How Physics Makes Us Free*, Ismael (2016) does not offer a conceptual analysis of free will, but instead lays out a vivid account of the structures in nature that support our everyday experiences of agency, volitional choice, and free action. Ismael argues that our default experience of ourselves as free agents making self-determined judgements and choices is fundamentally compatible with deterministic physics[3] and refutes the idea that physics reduces human action to mechanistic workings of particles and processes. Ismael systematically debunks the assumptions that underpin the consequence argument, disentangles the notions of causality and determinism, and demonstrates that the laws of nature are not "iron rails" (p.111) that predetermine our fate, but emergent patterns of regularity to which our actions contribute. Furthermore, Ismael argues, self-governing systems with the capacity for representational thought have evolved the ability to deliberate and act freely. Self-governance decouples our behaviours from the environment, and enables deliberative control over our judgments and actions.

Ismael (2021) describes her approach to ontological questions as a kind of naturalistic pragmatism. For example, she describes her enquiry into causation in the following way:

> Ask how causal thinking and the structures that support causal thinking arise, and then assemble the pieces of what we've learned […] into an outline of an answer to those questions. [I suggest that] once those questions are answered, there is no specifically metaphysical question about causation that remains". (1:01:06)

Building on Ismael's account of free agency, and following her methodology, I aim to fill out some of the details of the self-governing agent, offer an account of the cognitive structures that underpin our first-person experiences of choice and endogenous control of action, and discuss how those structures arise. Using the predictive processing framework to explore the structures of agency, choice, and autonomy that underpin self-governance, I will argue that our potential for freedom – freedom in its most fully developed manifestation – lies in our

---

[3] Ismael notes that although the current picture offered by physics is more complicated than classical determinism suggests (and is far from settled) her discussion takes place entirely within a classical paradigm since it is in this setting where the problem is supposed to arise (p.233). Properly understood, she argues, deterministic physics is not a threat to freedom.

capacity for autonomy, which is realized when we act deliberately to modify our own cognitive processes and transform our own minds.

The discussion that follows is arranged in four chapters. In Chapter Two, I outline key arguments and explanations from Jennan Ismael's (2016) naturalistic argument for free will. I present Ismael's view that, properly understood, deterministic physics is not a threat to freedom, and does not enchain us to a string of causes which necessitate our actions. I explain how Ismael disentangles the notions of causality with determinism, how she accounts for the epistemic and practical asymmetries that characterise our experience at the macroscopic level of reality, and describe how she positions emergent, human 'selves' and free action within the natural order. In Chapter Three, I introduce the predictive processing paradigm as a unifying theory of brain function with extreme explanatory power, and explain how the paradigm adds depth and detail to Ismael's conception of the self-governing system as free agent. I describe Karl Friston's (2009, 2010) account of action from first principles (the Free Energy Principle) which enables a richer account of how choice and action, and the structures that support choice and action arise. I discuss how the predictive processing framework deals with choice, actions and action policies, and selves. In Chapter Four, I draw on the insights of the predictive processing framework to discuss the status of the self, the nature and boundaries of the agent, and the scope of freedom available to the agent. I discuss the structures and processes that underpin choice and action selection, and the way in which we exert top-down influence over those structures and processes. I describe the critical role of attention and information in defining the boundaries within which we can act, and discuss autonomy as a process of integration and active modification of our own cognitive environment. In Chapter Five, I 'assemble the pieces together', and offer an expansive view of freedom which extends from our basic position as adaptive agents situated in a universe which is open to intervention, to the cognitive structures that support volitional choice and action, to the possibility of constructing ourselves as fully developed free agents. Finally, I discuss how we might act to expand the parameters of our freedom by deliberately shaping the deep mind from which our actions and decisions emerge.

2. The Physics of Freedom – Ismael's solution

This chapter outlines key arguments and explanations contained in Jennan Ismael's (2016) naturalistic account of free will, and describes how Ismael reconciles free will with physical determinism. It explains Ismael's view that properly understood, deterministic physics is not a threat to freedom, shows how Ismael refutes the consequence argument, and describes how Ismael positions emergent, human *selves* and free action within the natural order. Ismael shows that the freedom we ordinarily take ourselves to have is not threatened by fundamental dynamical laws, the causal structure of the macroscopic world, or emergent global regularities. Ismael disentangles the concepts of causality and determinism, and strips imposed notions of necessity from microphysical laws, causal relations, and emergent global laws. Ismael argues convincingly that our common sense conception of ourselves as self-governing free agents is accommodated within a rigorously scientific worldview.

2.1. Determinism as a Logical Rather than Causal Relation

The fundamentalist causal picture implicit in the consequence argument is that events at the macroscopic level unfold as the inevitable result of the blind behaviour of particles at the microscopic level. However, according to Ismael, the idea that determinism at the microphysical level involves a kind of pushy, compulsive necessity is a notion transposed onto physics from outside; a misunderstanding which arises from the deep-rooted causal concepts that play a foundational role in our understanding of the world. Causal relations, felt exchanges of force, are fundamentally built into our neuromuscular interactions with our environment, from our earliest experiences of pushing, pulling, and exploring the world. But although causal thinking is an integral part of structured experience, it would be a mistake to think that the asymmetric causal pathways that are built into experience at the macroscopic level are a reflection of, or result of causal interactions at the microphysical level. Instead, at the microscopic level of physical description, forces, asymmetry, and causality disappear.

Ismael draws on Bertrand Russell's (1912) argument for rejecting a causal interpretation of fundamental laws. Russell argued that the asymmetries we associate with causation are absent from the dynamical laws that describe the fundamental nature of reality. While causation and causal thinking are inherently directional in nature, dynamical laws have no intrinsic direction. The fundamental laws of physics describe relationships between functionally interdependent variables; the equations allow us to calculate the value of variables based on the values of related variables, but there is no direction of determination. The equations that describe the most fundamental level of physical reality are temporally symmetrical, that is, they are invariant under time reversal, and do not tell us anything about causal necessity or direction of influence. Without a direction of influence, the claim that antecedents determine actions is no more legitimate than the claim that actions determine antecedents. Thus, at the microscopic level at least, the idea that facts about the past causally necessitate our actions in the present is not supported by physics. Determinism at the micro-level is a logical rather than causal relation.

Ismael's approach to disentangling the concepts of causation and determinism is closely aligned with that of Carl Hoefer (2002). Hoefer argues that the notion of causal determination is an erroneous conflation of two distinct concepts, and attributes the confusion to the mismatch between the way we experience time and the way science understands time – in Ismael’s language, the difference between the temporally embedded, versus temporally transcendent view. According to Hoefer, a proper understanding of time, which clearly distinguishes these two conceptions of time, resolves the purported threat that determinism poses to freedom.

Our experience is rooted in A-series time, in which temporal events are ordered according to what happened before, what is happening now, and what will happen in the future. We experience ourselves as moving through time towards the future, acting in the present, and leaving more and more of the past behind. Our experience of time, characterised by epistemic and practical asymmetry, is dominated by causal and explanatory thinking, which we tend to bring into our thinking about determinism. The resulting metaphysical picture, according to Hoefer, is an “unholy marriage” between deterministic physics and our A-Series view of time (p208).

In contrast, determinism belongs to B-series time, the scientific conception of time, in which events are ordered according to their temporal distance from each other. Properties change at certain points, but there is no ontological separation between past and future, no privileged now moment, or past-to-future explanatory relation. In B-Series time, deterministic laws express temporally symmetrical, logical relations of interconnected variables. Because deterministic laws are temporally symmetrical, events at any given moment in time have logical implications for the past and the future, but those implications are not causally explanatory. It is equally legitimate to say that past events determine future ones, future events determine past ones, or some event in the middle determines past and future (Hoefer, 2002).

The idea, implicit in the consequence argument, that necessity or compulsion is somehow built into the fabric of the universe, is an illusion based on a misunderstanding of physics. As Ismael and Hoefer demonstrate, there is no inherent tension between freedom and determinism. The apparent conflict arises out of the erroneous understanding of the deterministic relation, which is an artefact of the way we process information, experience the flow of time, and identify causal structures in macro-level reality. However, when physics is stripped of imposed notions of causal necessity at the micro-level, it is clear that there is nothing at the micro-level that rules out free will. The deterministic laws that underpin physical reality are logical relationships that describe the relationships between functionally dependent variables.

Causal information goes beyond information contained in dynamical laws and arises, Ismael argues, with the capacity for intervention. While dynamical laws are exceptionless, apply to the whole system, and tell us about how the system behaves if left to its own devices, causal information is contingent, localised, and tells us how the system would be affected if we intervened.

2.2. Interventionalism: Causation as a Consequence of Freedom

Although causal information is not contained in the fundamental laws of physical reality, causal thinking plays a pivotal role in understanding the world and guiding our behaviour. Ismael describes the dilemma articulated by Nancy Cartwright [4]in response to Russell's argument that dynamical determinism requires us to dispense with causal laws. In practical reasoning, Cartwright noted, causal information is indispensable. To navigate our environment effectively, it is not enough to identify correlations and patterns of regularity, we depend on accurate causal judgements. For example, there are correlations between rates of smoking, bad breath, and lung cancer, but rates of lung cancer cannot be decreased by giving people breath mints. Effective, practical strategies depend on clearly identifying causal relations between variables.

---

[4] Cartwright, (1979) Causal Laws and Effective Strategies

There is an important difference between functionally interdependent variables and causal dependence. Ismael illustrates the distinction using Boyle’s Law, a simple, three-variable formula that describes the interdependent relationship between the variables of pressure, volume, and temperature. Boyle's law states that the pressure and volume of a gas are inversely proportional, providing the temperature of a gas is held constant. So, if we know the values of two of those variables, we can determine the value of the third. But causation or direction of influence plays no part in the formal description of this functionally interdependent three-variable system. Causation only enters the picture when we intervene, and causal direction depends on which variables we hold fixed and which we modify. If, for example, we hold temperature fixed and decrease volume, pressure is increased; the direction of influence flows from the variable of volume. If we hold temperature fixed and decrease pressure, volume is increased; direction of influence flows from the variable of pressure.

In more complex, mechanical systems, the same principle applies: causal information cannot be extracted from the global laws of the system. Global laws, Ismael argues, tell us how a system evolves over time if left to its own devices, while causal information tells us what happens if we intervene in that system, and sever a variable from its antecedents. If we know the global laws of an open system – Ismael offers examples of a simple mechanical device, such as a lever, pulley and rope system, or an automated pool cleaner – we can predict the behaviour of the overall system. Global laws enable us to predict the configuration or position of a system at one time from its configuration or position at another time and to understand what sorts of possibilities exist for the system as a whole. The global laws of a mechanical system might tell us that when Part A is in Position 1, and Part B is in Position 2, then Part C is in Position 3. But we cannot extract modal information – information of what is possible, contingent, or necessary under counterfactual circumstances – from global laws. To understand how components within the system affect each other – whether Part A drives Part B, Part B drives Part A, or Part C drives both Part A and Part B requires intervention and hypothesis testing. Intervention allows us to isolate some collection of variables, reach into the system and manipulate some part, and trace the downstream effects. As with the example of Boyle’s Law, the direction of influence will depend on what we take to be the exogenous variable (the intervention, or independent variable), what we take to be the endogenous (the dependent variables) and what background factors we hold fixed. As such, knowledge of causal structure – the sort of knowledge that is needed for deliberation and effective, practical action – arises with the capacity for intervention.

In everyday causal reasoning, we make implicit, context-dependent choices about which variables are being considered (the network of variables in the system), which are held fixed (endogenous variables), and which are modified (exogenous variables). These selections, Ismael explains, are made explicit in Judea Pearl's Structural Causal Modelling (SCM), a formal, scientifically grounded framework which models the causal substructure of a network using Directed Acyclic Graphs (DAGs). DAGs display causal information – complex contingencies, direct and indirect causal relationships between selected variables in a system – in a mathematically precise way that supports hypothetical reasoning. SCM has provided a scientifically robust framework for causal thinking, and has resolved the dilemma highlighted by Cartwright and Russell's disagreement. Crucially, according to Ismael, SCM shows that the causal substructure of the world is not a threat to freedom, but an enabler of it.

Causal sub-structures in macroscopic reality are emergent patterns of regularity that appropriately situated agents can strategically exploit for their own ends. Causation is not fundamental, and does not entail necessity, but arises as a result of the capacity for intervention. While fundamental laws are symmetrical, and describe constrained relationships between interdependent variables, emergent, causal relationships are asymmetrical, and pick out connections which can be manipulated. The asymmetry of causal relationships arises as a result of our choice of variables in a network, where we draw the boundary between endogenous and exogenous variables, and the point at which we intervene (hypothetically, or in practice). These selections are pragmatic choices which depend on our purpose and interests. Thus, as Ismael argues, causal pathways do not enchain us but function as strategic routes towards our own ends.

2.3. The Self-Governing System: Inserting Selves into the Causal Chain

Pre-philosophical conceptions of free action are rooted in the phenomenology of *selves* as agents; we tend to hold ourselves and each other responsible for our actions because we take our *selves* to be the doers and deciders. A critical challenge for a naturalistic account of free will, then, is to coherently insert *selves* into the causal chain, and to provide a scientifically grounded account of the *self* as agent that avoids positing any kind of immaterial essence or inner homunculus. Rejecting both Cartesian, dualist conceptions of self as an immaterial substance, and what she terms, "nolipsism" (p.9) – the view that there is no self – Ismael argues for a conception of self which accords with both a naturalistic worldview and our first-person experience. In Ismael's view, *self* as an emergent structure within the natural order; a virtual object, or unified psychological viewpoint, which arises as a consequence of the way the brain processes information. According to Ismael, the *self* functions as an emergent, unified centre of experience, judgement, and action. *Selves* arise in nature in the evolutionary transition from self-organisation to self-governance.

A self-organised system is a functional unit comprised of a dynamic grouping of component parts, in which information and control is distributed throughout the system and order emerges spontaneously. The overall structure and organisation of a self-organised system is not achieved by top-down coordination but emerges from the interconnection and interaction between its component parts. Nature abounds with self-organised systems, loosely organised systems in which feedback loops amplify patterns, with no centralised control coordinating collective activity. Examples include the collective movement of schools of fish and flocks of birds, the building of ant colonies, and many systems in our own bodies and social structures. In self-organised systems, local-level interactions contain information about, and are regulated by collective variables, and the global configuration is comprised by, and modified by, information contained at the local level. The coordinated behaviour of the overall system is the result of information distributed throughout its component parts but – in contrast to systems which are self-governing – there is no centralised control that orchestrates collective activity. The difference between a self-organised system and a self-governing one is the difference between a free market economy and a state-run economy: self-governance adds a layer of centralised processing that enables top-down control.

The behaviour of self-governing systems goes beyond the spontaneous, emergent order that characterises systems that are merely self-organising. Self-governing systems are equipped with centralised representational structures that enable distributed information to be unified in a single epistemic and deliberative viewpoint. In such systems, information is continuously transferred between the unified, collective viewpoint and its component parts, in an ongoing cycle of reciprocal causation. The behaviour of component parts is regulated and controlled by information embodied in the global configuration, which is itself generated by the activity and information received from its parts.

With self-governance arises the unified, emergent *self,* a synthesised, high-level point of view, in which distributed information is amalgamated into a single, reflective standpoint. The *self*, as Ismael conceives it, fulfils three functions. First, it synthesises streams of sensory information into a single frame of reference or representative model. Second, it operates as a collective voice for the semi-autonomous subsystems it subsumes. Third, it operates as a unified centre of deliberative control, with the ability to make judgements and act in the interests of the system as a whole. Ismael's *self* does not pre-exist these functions, or occupy a space which acquires these functions, nor is the *self* exclusively the domain of linguistically mediated thought. Rather, the self is the emergent, conscious unity of the overall system, capable of making all-things-considered judgements and acting on behalf of the entire system. This scientifically grounded conception of human beings as unified self-governing systems with emergent, unified selves sits easily with the ordinary sense of ownership and control we associate with deliberation and volitional action.

In Ismael's framework, self-governance equips the organism with pivotal control over its own behaviour, decoupling the system from its environment and making genuinely free action possible. For any open system[5], behavioural output is a consequence of the state of the external variables impinging on the system combined with the state of the internal variables. The internal structures which control non-choice governed behaviour, such as the patellar reflex in humans and other animals, or the tongue-snapping response in frogs, are designed (by natural selection) to enable stabilised responses to environmental variables. But the structures that enable choice-governed behaviour support enormous variability in responses to environmental variables. As self-governing systems, the way that we process information – that is, our ability to alter the state of the internal variables and mediate the connection between stimulus and response with deliberation – gives us pivotal control over our actions. Voluntary behaviour in self-governing systems is not constrained to patterns of conditioned response, but is mediated by reflection and deliberation, and guided by higher order values, beliefs, and conceptual frameworks extracted from personal experience.

---

[5] An open system is one which exchanges energy and information with the environment (that is, between internal and external states). All biological organisms are open systems. The only truly closed system is the universe itself, which has no external states, and consequently does not exchange energy and information with the environment.

Ismael demonstrates that the self, as an emergent locus of experience and control, can be coherently inserted into the causal chain. Self-governance, which arises with the capacity for representational thought, enables the system to regulate its own behaviour, and make unified, all-things-considered judgements on its own behalf. First order claims on behaviour – drives, dispositions, and preferences – are resolved by reflective, deliberative processes, and informed by higher order values and goals, which are drawn out of a lifetime of direct and reprocessed experience. As self-governing sub-systems within a greater whole, we have the capacity for intervention – that is, we have the ability to sever a variable from its antecedents, and to mediate the link between environmental stimulus and behavioural response, with deliberative control. Thus, Ismael argues, the ordinary sense of freedom and control that we associate with deliberative choice and action, is not illusory.

2.4. Freedom to Defy Prediction

Even if actions can be caused by deliberated choices, one lingering concern is that determinism means that the outcome of any given deliberative process is preordained by nomological necessity. The Consequence Argument (outlined in the Introduction) states that all events are the consequence of the past and the laws of nature. If, as the argument implicitly claims, the laws are in place prior to our actions, then long before we begin to deliberate over some action, the outcome of that deliberation is already fixed. The sense of being free to choose is an illusion due to our own epistemic limitations. However, Ismael argues that determinism does not predetermine our actions, and there are no laws in place before we act that make our actions inevitable. To show that this way of thinking about laws is mistaken, Ismael introduces Michael Scrivens' thought experiment, the Paradox of Predictability[6].

The set-up goes like this: In a deterministic universe, there must be a deterministic relation that connects the state of the world at one time to its state at another time, in a unique manner. The universe would be immanently[7] predictable if a perfect predictor was embedded

---

[6] Scrivens first presented The Paradox of Predictability in a 1965 paper entitled An Essential Unpredictability in Human Behavior.

[7] Ismael distinguishes immanent predictability from transcendent predictability. The universe is not immanently predictable (that is, it is not predictable from the point of view of an intelligence within the system) because initial conditions together with local laws are insufficient to determine the future. At the same time, the universe could be transcendently predictable if,

in this universe: an all-knowing Laplacean intelligence would be able to predict events for any time in the future. But that embedded predictor could be undermined by the presence of a counterpredictive device – a device programmed to generate behavioural output opposite to any revealed prediction. A simple version of this sort of device could be easily programmed without breaking any deterministic laws, but in practice, we ourselves could operate as counterpredictive devices; we are free to invalidate any revealed prediction of our behaviour, just as we would ordinarily think. Ismael argues that the Paradox of Predictability not only affirms our practical ability to undermine any prediction of our actions, but also uncovers a subtle misunderstanding of laws, that underpins the belief that determinism is a threat to freedom.

In a deterministic universe of the type implicit in the consequence argument, the laws function to map the state of the world at one time onto its state at another time. In principle, if there was a Laplacean intelligence who had complete knowledge of those laws (along with knowledge of initial conditions) then that intelligence would be able to flawlessly predict any event before it occurred. The absence of epistemic limitations would render the behaviour of any system utterly predictable. But the two conditions together – the existence of an all-knowing predictor and a counterpredictive device – create a situation in which that information cannot be available to generate a prediction without also being available as a causal link to undermine that prediction. Therefore, there can be no set of information – no set of facts or laws – that fix our decisions before we make them. We can simply introduce an additional cause to the causal chain that changes the outcome. The upshot, according to Ismael, is that nothing from the past can nomologically fix our actions in the future.

The puzzle inherent in the Paradox of Predictability rests on the assumption that laws are in place before we act that determine our actions in advance. If this assumption about laws is removed, the expectation of being able to predict behaviour on their basis also disappears, and the puzzle is resolved. Ismael argues that thinking of laws as in place before we act gets things precisely the wrong way around. Laws are characteristic patterns which play out over time, and which are generated, in part, by our actions. Ismael uses the analogy of Shakespeare's tragedy, Hamlet, to illustrate the distinction. The play becomes a tragedy

---

in addition to initial conditions and local laws, there is an additional specification that there are no exogenous influences on the system. This additional specification is information that is available only within a transcendent perspective (p.178n7).

because of the character's decisions and actions throughout the play. But the property of being a tragedy does not *pre-exist* the actions of Hamlet, the character, functioning to constrain his actions and the unfolding of the plot. The property of being a tragedy comes about as a result of the unfolding of the plot. Likewise, Ismael argues, we should understand facts as first order information about events in space and time, and laws to be derivative, higher-order, global patterns which are constituted by and emerge out of those events.

Having exposed the error of reifying laws, and dispelled the idea that laws operate as 'iron rails' that predetermine our actions and life trajectory, Ismael delineates a layered reality in which freedom is completely compatible with classical determinism. At the fundamental level, particles respond to their environment in predetermined ways according to the laws of classical physics. At the macroscopic level, open subsystems interact with each other, and other aspects of the environment, in complex ways. Here, self-governing systems, which are themselves built out of deterministic components, regulate their own responses to the external world, and have the final word over their choices and actions. At the global level, patterns emerge across the entire system as it evolves over time. The global laws of temporal evolution are spread out over History[8], and are not logically entailed by classical microphysical laws that govern the behaviour of particles. From the point of view of an embedded participant in History, these laws are not in place before we act but emerge, in part, from our own actions and the complexities of macro-level interactions. There is, Ismael argues, a real sense in which deliberating human beings are ultimately free.

2.5. The Open Future: The Physics of Practical and Epistemic Asymmetry

The Paradox of Predictability demonstrates that there is a concrete, practical sense in which our moment-to-moment choices and actions are not fixed by anything in the past; from the perspective of the embedded agent, we are free to overturn any revealed prediction of our decisions, and our decisions become part of the overall pattern of events as the system evolves over time. But from the temporally transcendent view, the categories of past and future break down. The manifold of spacetime itself, does not evolve and from this point of view, there is no unfolding sequence of events within it. So we might think that the kind of openness we assume about the future is illusory. If future events are in some sense fixed already, the sense of forward-looking openness that we take for granted when we deliberate is

---

[8] Ismael capitalises History to refer to the entire sweep of history, encompassing past, present, and future (p. 237).

the result of epistemic limitation rather than a reflection of actually realisable possibilities. The vision of the block universe seems to preclude a genuinely open future. Nevertheless, Ismael shows that we can recover this openness by understanding the physics behind the epistemic and practical asymmetries that characterise our macroscopic-level experience, and clarifying the nature of openness from both the perspective of the embedded agent, and within the transcendent vision of spacetime as a whole.

Because deterministic laws, at the microscopic level, are symmetrical under time reversal, interventions at the microscopic level have microscopic effects that run into the past and future. But at the macroscopic level – the level of physical reality to which our senses are coupled – interventions propagate asymmetrically into the past and future. Making a change to the macroscopic environment, alters the future and leaves the past unaffected. Ismael explains the asymmetry of determination that arises at the macroscopic level, following the framework developed by David Albert (2000).

The asymmetry and direction of determination that arise at the macroscopic level of reality is attributable to three principles: (a) the temporally symmetrical Newtonian Laws of Motion, (b) a statistical postulate which assigns a probability distribution of system being in a particular microstate given its macrostate, and (c) the Past Hypothesis which states that the universe started in a state of very low entropy, with no corresponding boundary condition in the future. Ismael describes how these principles work together to bring about macroscopic asymmetry of determination: If we hold fixed all the information in a classical, Newtonian environment and make some macroscopic change to that environment, then apply the probability postulate to obtain a probability over a possible microstate of the world, and conditionalize that on a low-entropy boundary condition in the past, the probability distribution is altered in an asymmetrical way; the probability of past events remain unaltered while the probability of particular future events is raised. This accords with the way that we, as creatures coupled to the macroscopic level of reality, experience practical asymmetry in the world. Macroscopic-level interventions propagate asymmetrically into the past and future; current actions change the future but leave the past unchanged. When we act, we reorder the environment, creating information that is carried forward with the thermodynamic gradient, imprinting the future with records of current interventions. As human beings with complex representational structures, we also make psychological records, organising and reorganising them into themes, timelines and narratives. Practical asymmetry at the macroscopic level is the basis of causal thinking, which we use to guide our behaviour.

The flipside of practical asymmetry is epistemic asymmetry; we can have knowledge of the past, but not of the future. The thermodynamic gradient allows information about past states to be carried forward by leaving records in the environment. Our experience of the world is characterised by epistemic asymmetry because macroscopic reality is rich with environmental records of the past, which enables us to make inferences, and stabilise our beliefs about the past. But there are no corresponding environmental or psychological records of the future. Our lack of knowledge about the future is a consequence of the possibility of taking strategic action. Beliefs about the future are conditional on the decisions we make in the present, so we cannot stabilise our beliefs about the future until we stabilise our beliefs about our decisions, which we do moment-to-moment as we choose and act. The course of future events depends on our current choices and actions – we create information that is carried into the future with the thermodynamic gradient.

From the point of view of the embedded participant, the future is open, and the course of events unfolds as we experience it. In the moment of decision, there is a tangible sense in which there are multiple realisable possibilities. But from the temporally transcendent perspective this kind of openness does not make sense. In the entirety of spacetime, all events, and the relations between events, are already represented. The universe itself is not coming into being as we experience it, there is no movement *through* time as an exogenous parameter, and no ontological separation between past and future. From this viewpoint, concepts such as past, future, openness, and even *exist* – concepts in which the embedded perspective of time is built-in – start to break down.

From the temporally transcendent viewpoint, openness, like the categories of past and future, can only be a relational property; a property that one event has in relation to some other events, but not in relation to others. The property of being in the future, for example, is analogous to the property of being a nearby location. These are not intrinsic properties applicable to specific events in space or time, but apply to a particular relation between two points. Openness, from the transcendent view, works like this too. Internal to the entire spacetime manifold, openness is a property which events have in virtue of their relation to some events and not to others, a property an event has before it happens, but not after.

The openness we associate with the future is not merely the result of epistemic limitation. The future is not fixed independently of our choices, and not there anyway, regardless of what we do or think we can do. It is a misinterpretation of physics to think of the future as ‘there anyway’, independently of our choices. Our decisions are part of what shapes reality, our actions are the parts of the universe that we make actual.

2.6. Self-Constitution: the Active, Emergent Self as Agent

According to Ismael, our lives are partly constituted by our choices; at least to some extent, we are free to exercise creative power over our life course and decisions. But some philosophers argue that freedom would require the ability to create ourselves from the ground up. Galen Strawson (2010), for example, argues that we cannot be responsible for our actions because our actions are the result of how we are, and how we are is determined by genes and past experience. Further, we cannot acquire responsibility by changing how we are, because the ways in which we try to change are also determined by genes and past experience. For Strawson, for us to have any sort of genuine freedom would require self-constitution all the way down – our genes and environment would have to be genuinely up to us. In Ismael’s view, philosophers who set such an impossible standard for freedom and responsibility have been captured by a magical ideal of what freedom is. Although the scope within which individuals can choose is constrained by circumstance, we actively construct ourselves as believers and agents, and this is enough for freedom.

The self, Ismael argues, is a unified locus of judgement and agency. The nature of the unified self is constructed from the raw material of sensory experience in both active and passive ways. At the preconscious, sub-personal level, the construction is immediate and automatic; the self is stabilised as a unified sensorimotor presence distinct from the environment, sensory input is passively selected, and information is stored as episodic memories. At the conscious, personal level, however, the process is active; we are agents who actively select, reprocess, and construct information. The self makes high level decisions about what to believe and what to do, constructs an identity, and organises episodic information into an autobiographical narrative; we actively interpret and extract meaning from our life experiences. The construction of self is not just something that happens to us or within us, but something we actively do.

The self is the entire stream of thoughts, impressions, and feelings spread out over a lifetime. We make choices about what to believe and what to do, and we make choices about the values, goals and priorities that we use to guide those decisions, becoming increasingly self-created as we do. We constitute ourselves as believers by developing a rationally coherent framework in which we evaluate and decide what to believe. We constitute ourselves as practical agents by making self-conscious, top-down, all-things-considered decisions about what do. Our minds are not merely products of our environment. We make choices about who we are and what we care about, and those choices inform our actions. We actively reflect on and reprocess experience, creating our own minds as we choose, reason, and evaluate the contents of our experience.

The capacity for choice and self-governance is a given. We cannot avoid choosing. Some of our actions will necessarily be the result of reflective, deliberative choice. Although individuals vary in how careful, skilful, or motivated they are to self-govern well – and we can fail to self-govern if we are manipulated or at the mercy of drives and appetites, for example – we constantly encounter choice points that demand a reflective, conscious response. In making those choices, we produce information that shapes the future, and ourselves.

In summary, Ismael lays out a compelling, scientifically grounded account of the structures that support our everyday experience of freedom, choice, and agency, and shows that our pre-philosophical conception of ourselves as free agents faced with genuine choices and possibilities for free action is not illusory. Determinism does not entangle us in an unalterable sequence of causes that predetermine our course, and the process of deliberation and choice is not predetermined by factors out of our control. As self-governing systems, we are capable of non-mechanical control over our own behaviour.

3. Predictive Minds

Ismael's account of emergent, self-governing agents leaves many details to be filled in; a more complete picture of freedom requires an account of the cognitive structures that underpin agency and choice. In this chapter, I aim to enrich Ismael's contextualised picture of human agency with a precise, unified account of the nature of agency itself, before discussing, in subsequent chapters, the implications of this account for freedom. To develop this account, I turn to the predictive processing (or Active Inference) framework, an elegant, unified theory of brain function, which offers promising insights for understanding the nature of agency and choice. Predictive processing – an increasingly well-established theory in neuroscience, cognitive science, and philosophy of mind (Hohwy, 2020; Smith et al., 2024; Walsh et al., 2020) – fits neatly with Ismael's broad conception of the self-governing agent. In Ismael's account, freedom – our capacity to decide and act in ways that transcend the limits of automated behaviour and environmental control – is a consequence of the way the brain processes information. The predictive processing paradigm enables us to add detail to this picture, describing action, cognition, and perception in unified, information theoretic terms. In this framework, the brain is understood to be a predictive mechanism, which constantly anticipates its encounters with the world, predicts the flow of sensory data, and constructs and updates its probabilistic model of the causes of that data. The framework has extreme explanatory power, presenting a unified conception of brain function in which perception, cognition, and action are understood to be complementary ways of fulfilling the brain's main task: that of minimising prediction error between sensory input and the internal model that accounts for that input (Clark, 2016; Hohwy, 2013). In its most far-reaching formulation, the predictive processing framework is underpinned by the free energy minimisation account developed by theoretical neurobiologist Karl Friston and his collaborators. Friston's free energy principle, outlined in the following section, serves as the unifying foundation of this computational theory of brain function, and can be regarded as the organising principle of life (Friston, 2009, 2010).

3.1. Action from First Principles

The unified account of brain function developed by Karl Friston and others begins by answering a fundamental question: how do living systems maintain their organisation and structure (for a finite time) in an environment characterised by thermodynamic entropy? A simplified version of Friston's argument (Friston, 2009, 2010; Parr et al., 2022, see also Wiese & Metzinger, 2017) is outlined below.

For any adaptive system to exist it must, by definition, maintain the boundary that distinguishes it from everything else. In statistical terms, this boundary is delineated by the system's Markov blanket – a statistical construct that marks the partition between an open system and its environment[9] and mediates exchanges between the two. The Markov blanket effectively screens off internal states from external ones, rendering internal and external variables conditionally independent of each other. The blanket itself consists of two further states: *sensory states* and *active states*. *Sensory states* are variables which influence internal states, and are influenced by but cannot directly influence external states (though they can influence external states indirectly by modifying active states). *Active state*s are variables which influence external states, and are influenced by but cannot directly influence internal states (though they can influence internal states indirectly by modifying sensory states). For a living system to maintain its structure and boundaries is just for its Markov blanket to persist (Kirchhoff et al., 2018; Parr et al., 2022). The persistence of the Markov blanket requires that the system maintain itself within high probability states.

For an adaptive organism to sustain itself in the face of dissipating environmental forces, it is necessary that the organism occupy a narrow range of high probability states. Relative to the set of all possible states, only a very restricted set of states is compatible with the organism's viability, while most states represent an existential threat. It follows that if we were to randomly sample a living organism over its lifetime, by definition most samples would fall within the restricted range of states compatible with its viability, with few samples outside of that range. A multidimensional scatter plot of the sampled states would have some small, densely populated regions, while most regions of the state space would be sparsely populated; the probability distribution of the sampled states would be said to have low entropy[10]. In order for a living organism to continue to exist, the states the system visits over its lifetime must necessarily be characterised by this low-entropy probability distribution. The long-term viability of a biological system, then, depends on the system occupying high probability states in the short term (Friston, 2009, 2010; Parr et al., 2022; Wiese & Metzinger, 2017).

---

[9] In the living world, Markov blankets delineate the (statistical) boundaries of whole organisms, and can be nested within each other, defining, for example, the boundaries of organelles, cells, brains, organisms, and communities (Kirchhoff et al., 2018).

[10] Entropy, in this context, is an information theoretic quantity, analogous to but distinct from thermodynamic entropy. By contrast, a high entropy probability distribution is one in which sampled states are distributed more evenly across the entire state-space.

To put this another way, in statistical terms, the continued viability of the organism requires that the organism minimise *surprise*[11]. Surprise, in this context, is an information theoretic measure that quantifies the unlikeliness (formally, the negative log probability) of an event under a given model. In statistical terms, an improbable event is highly surprising, and as the probability of an outcome increases, surprise decreases. Surprise is relative to the system, or the model – for a fish, being out of water would be a surprising state; for a human being, being adrift in an open ocean would be a surprising state. The entropy of a probability distribution is a measure of the average surprise of outcomes sampled from that distribution. So, maintaining a low entropy distribution of states depends on minimising surprise, or in Bayesian terms, maximising evidence for the model (self-evidencing).

So far, Friston has shown that in order for an adaptive system to maintain its boundaries and structure in the long term, it is necessary for it to maintain a low-entropy distribution of states, which requires minimising surprising states in the short term. But although minimising surprise is an existential imperative, from the perspective of the organism, evaluating which states are surprising is an intractable computational problem. Systems bound by a Markov blanket have access to information about the world only via sensory states, and for living systems, that information is partial and noisy. Consequently, mapping the worldly causes of sensory data is difficult, since any unit of sensory input may in principle have a number of causes. Moreover, calculating which states are surprising would require complete knowledge of all possible states.

Although living systems cannot minimise surprise directly, they can, using the information available to them via sensory states and brain states, do so indirectly by minimising free energy. Free energy, in an information theoretic context, is a measure of the difference between two probability distributions, or probability densities. Applied to biological systems, free energy is a function of sensory data and the recognition density – the internally coded, probabilistic model of the causes of that data. Or, to put it in terms of the Markov blanket, free energy is a function of sensory states and internal states. The free energy principle states that any quantity that can change, changes to minimise free energy. Minimising free energy,

---

[11] The terms surprise and surprisal are often used interchangeably in discussions about the free energy principle, but surprisal is sometimes preferred to emphasise the distinction between surprise as an information theoretic measure of unexpectedness, and surprise as a personal-level response to something unexpected. Surprise is used here only in the former sense.

which under some simplifying assumptions reduces to minimising the prediction error between the sensory data and the probabilistic model that accounts for the data, implicitly minimises surprise, because free energy is always greater than surprise. [12] (Friston, 2009, 2010; Parr et al., 2022). Minimising free energy then, transforms the intractable computational problem of surprise minimisation into a manageable optimisation problem.

To recap, adaptive systems resist the tendency to disorder by minimising free energy. Biological organisms must minimise surprising states in the short term in order to maintain their structure and boundaries over the long term. Surprising states are those outcomes which are unexpected given the model (the model here includes information about preferred states, or states which are compatible with the organism's window of viability). The existential necessity of minimising surprise is resolved indirectly, by optimising the alignment between sensory input (that carries information about hidden states of the world) and the recognition density – the organism's internally coded model of the causes of sensory data.

The free energy principle is a theoretical framework that enables a deeply coherent account of action to be built from first principles. The principle says that any exchange between an adaptive system and its environment must minimise (information theoretic) free energy. In practice, free energy minimisation is achieved by prediction error minimisation, which is implemented by the dual strategies of action and perception.

3.2. Perception and Action as Error Minimisation Strategies

According to the predictive processing framework, the primary task for the brain is prediction error minimisation, that is, to minimise the discrepancy between sensory data and the internally coded model of the causes of that data. This biological imperative requires the organism to (a) orient towards plausible hypotheses to explain the sensory flux, thereby aligning internal states to sensory states, and (b) sample the evidence in a way that matches its internally coded model, thereby aligning sensory states with internal states (Hohwy, 2013). The ongoing, dynamic process of minimising the discrepancy between brain states and

---

[12] Free energy is greater than surprise because it can be mathematically demonstrated that free energy = surprise + the divergence between the recognition density (or internal model) and the conditional density (the true causes of the sensory data) which is always a positive number.

sensory states is implemented by the deeply intertwined strategies of perception and action, or perceptual and active inference.

The understanding of the brain as an inference machine was originally developed in the domain of perception, beginning with Helmholtz's insight that perception is necessarily a process of causal inference[13] (Clark, 2013), an insight which may itself be rooted in the work of Kant (Swanson, 2016; Wiese & Metzinger, 2017). The key idea is that the brain, with no unmediated access to the external world, is tasked with mapping the hidden structure of the external world from the evolving body of sensory data, which it does by generating probabilistic hypotheses. Reverse mapping worldly causes from their sensory effects is not straightforward. Bare sensory data is insufficient to map the source of that data; there is no one-to-one mapping of external causes to sensory states, or sensory states to external causes. Any given unit of sensory data may in principle be linked to a number of different external causes, just as any given external state may have a number of different sensory effects. With no direct access to external states, the brain infers the sources of sensory input using internally available information. In doing so, it realises a process akin to Bayesian updating, generating hypotheses based on prior knowledge, weighing them against the evidence, and selecting the hypothesis that best fits the evidence (Hohwy, 2013).

The brain achieves a probabilistic reverse mapping of the worldly causes of sensory effects using a computational process of perceptual inference (Clark, 2013, 2016; Friston, 2009; Hohwy, 2013). Restating Hohwy's (2013) simplified Bayesian description, the process is as follows: The fundamental perceptual problem for the brain is to predict the hidden states of the world w based on sensory data s. To resolve the problem, the brain maximises the value of all available information – sensory evidence and prior knowledge – simultaneously computing probability distributions for alternate hypotheses and selecting the most probable. Formally, the computational process can be represented by Bayes theorem:

$$p(w|s) = p(s|w)\ p(w) / p(s)$$

The left side of the formula, $p(w|s)$ represents the perceptual task – computing the probability of world state w given sensory data s (the posterior distribution, or updated belief). This is equal to the likelihood of the data (the probability of sensory data s given world state w) multiplied by the probability of world state w (the prior distribution, based on prior

---

[13] First presented by Helmholtz in Handbuch der physiologischen optik (1860)

knowledge of the world) divided by the marginal likelihood (the probability of sensory data *s* whose presence in the formula ensures that the posterior probability falls between 0 and 1, meeting the norms of probability). Applied to a simple example, suppose the visual system encounters a tall patch of green in the park. Calculating the posterior probability p(w|s) of there being a leafy tree given the visual data, depends on estimations of likelihood p(s|w) – the probability of the visual data if there was a leafy tree – and prior beliefs p(w) – the probability of there being trees here at all. Thus, the structure of perceptual experience is a dynamic balance of prior knowledge and sensory information. This sub-personal computational process of Bayesian inference is what makes the seemingly immediate and transparent structured experience of perception possible.

In the predictive processing framework, perception is understood to be a constructive, top-down process in which the brain predicts sensory information rather than passively receiving it. The brain constantly anticipates current and future sensory input, checks its hypotheses against incoming sensory evidence, and registers and corrects prediction errors. As the system registers mismatches between the predictions and evidence, updated probabilities (posterior beliefs) are dynamically incorporated back into the model, becoming new priors, and bringing the predictive model progressively into alignment with the causes of sensory input. Perceptual experience, in all its nuance and complexity, emerges out of the interplay between hierarchically ordered sensory expectations and error signals, which is realised at every level by the process of Bayesian inference.

The hierarchical ordering of the environment is mirrored in the hierarchical ordering of perceptual processing in which the contours of noise versus signal, and evaluations of accuracy and error at each level are shaped by expectations encoded at the level above. Regularities in the world are arranged in temporal and spatial hierarchies. In the temporal domain, processes that play out over the long term, which we conceive of in coarse-grained terms, set the frame of reference for, and have implications for shorter-term processes which we conceptualise in increasingly granular detail. Similarly, in the spatial domain, each level of the hierarchy sets the frame of reference for the levels below. Urban environments, for example, tend to contain office blocks, that contain offices and hallways, that contain furniture, that contain drawers, that contain paperclips, and so on. In the perceptual domain, prior beliefs[14] at each level of the hierarchy anticipate and constrain perceptions at the level below; high level probability estimates about the most likely patterns of information at lower (or adjacent[15]) levels, shape the processing of information at those levels. Visually, for example, lower-level processes such as edge and colour detection are informed by features predicted at the level above, such as eyes and noses, which are in turn constrained by predictions of faces at the level above, which are themselves influenced by social and environmental contextual predictions. Likewise with the other sensory modalities: encoded expectations at each level of the hierarchy anticipate and inform perceptions at the level below. As encoded expectations at each level of the predictive hierarchy constrain and contextualise percepts at the level below, prediction errors – reports of the mismatch between the predictive model and the incoming flow of sensory data – are fed up the cortical hierarchy, recoding expectations and restructuring the model to accommodate sensory input. The perceptual inference account has been expanded by Friston and others (Clark, 2013, 2016; Friston, 2009, 2010; Hohwy, 2013; Parr et al., 2022) into a unified theory of the brain in which both action and perception are brought together under the prediction error minimisation scheme.

While perception minimises prediction error by enabling the predictive model to be brought progressively closer to sensory input, action minimises prediction error by enabling sensory data to be actively sampled in a way that aligns with the predictive model. According to the active inference framework, action is not initiated by command but is a consequence of the sensory expectations encoded in the predictive model (Clark, 2016, 2020; Hohwy, 2013; Nave et al., 2022). Action is produced when counterfactual exteroceptive, proprioceptive, and interoceptive predictions – the predicted consequences of an action – initiate a stream of prediction errors, which are in turn quashed when the organism moves to align the sensory flow with the prediction (thereby providing evidence for the model).

---

[14] Beliefs, in the predictive processing framework, are sub-personal, probabilistic models rather than propositional formulations. They and are not necessarily linguistically mediated, or even consciously accessible to the agent.

[15] As Clark (2016) notes, although the structure is hierarchical, there is a degree of nuance that a strict hierarchical description does not capture. The structure of neuronal populations creating context, generating predictions, and registering errors for other neuronal populations, might be better conceived of as concentric circles of influence, or hierarchical plus horizontal.

In the predictive processing framework, perception and action are not fundamentally separate processes. Rather, action and perception are deeply intertwined strategies which work synergistically to minimise prediction error. Together they operate in a continuous reciprocal cycle where each constrains, structures, and contextualises the other. As we forage for information, for example, action enables us to test perceptual inference by moving the body to sample the evidence in different ways. When we reach out to touch, lean in to listen or look more closely, or view an object from different angles, we are able to test our hypotheses, reduce uncertainty, and track external states more accurately (Hohwy, 2013). Likewise, perception is built into action. Perception does not deliver an 'action-neutral' perspective on objective reality but rather functions to put creatures in touch with the world of affordances – the possibilities for action and intervention that their environments make available (Clark, 2016; Nave et al., 2022). The continuous oscillation between action and perception, and the hierarchical processes of action and perception themselves are modulated at every level by precision weighting – context-sensitive estimations of certainty.

3.3. Precision Mediated Error Minimisation

The dynamic balance of top-down sensory expectations and bottom-up error signals is mediated by ongoing estimations of confidence – or *precision* expectations – of those signals (Clark, 2016). Built into the constant process of evaluating and re-evaluating the accuracy of predictions by comparing them to incoming sensory signals, the brain continually evaluates the reliability of those signals and adjusts the depth of processing on either the predictions or sensory input accordingly. *Precision* estimates – dynamically shifting evaluations of the expected reliability, or confidence in sensory signals – modulate the relative neural gain on either the predictions or the error signals, determining whether information processing emphasises the predictive model or the sensory signals more heavily. In high certainty environments, precision is weighted more heavily in favour of the predictive model, which facilitates ease and economy of processing. In low-certainty (unfamiliar, noisy, or unexpected) environments, higher precision is assigned to prediction error signals, increasing the neural gain on those signals. Error signals are then processed in higher definition (with increased precision) enabling the predictive model to be updated with increased accuracy. The continuous, dynamic shifting of precision weighting is built into the predictive machinery at every level of the cognitive hierarchy and plays a key role in action production.

Precision is a critical part of the action account. Recall that action suppresses prediction error by actively sampling sensory input in a way that aligns with the currently deployed predictive model. Action production results from generating counterfactual exteroceptive, proprioceptive, and interoceptive predictions, which then initiates a stream of prediction errors, which are in turn quashed when the organism moves to align the sensory flow with the prediction. The mechanism of action production is precision. Reflex arcs initiate action automatically when the brain assigns high precision weighting to the predicted sensory consequences of an action; that is, when the brain actively disattends to current sensory data while increasing the gain on the counterfactual interoceptive and proprioceptive simulations that correspond to the movement. The resulting cascade of error signals is quashed when the movement that corresponds to the predicted sensations is executed (Clark, 2016, 2017; Hohwy, 2013; Nave et al., 2022).

Precision is implemented by both top-down and bottom-up (or endogenous and exogenous) attention (Clark, 2017; Hohwy, 2013). For example, when implementing endogenous attention in the sensory domain – perhaps I am looking for a friend's face in a crowd – higher precision is allocated to the sensory data in my visual field, increasing the neural gain on error signals as I visually scan the space, and making it more likely that I will locate my friend. Likewise with exogenous attention. If some unexpected sound such as a siren captures our attention, high precision is allocated to the sensory signal, increasing the neural gain on the error signals as the brain attempts to locate the source and incorporate the information into its model. In the same way, attention optimises precision in the domain of action too. Assigning high precision to interoceptive predictions associated with an action, brings about the predicted action. We achieve control of motor movements as we learn the sensory patterns associated with those movements, and the brain learns to systematically disattend to – lower precision on – information about current states. Self-generated action is enabled by lowering the precision on prediction error signals (which relay the mismatch between the deployed model and current sensory input) thereby decreasing their influence and allowing volitional movement (Clark, 2017). This mechanism of precision-mediated error minimisation is responsible for both simple motor actions and for complex action sequences that play out over extended timescales.

3.4. The Temporally Deep model – From Action to Action Policies

While all adaptive systems are obliged to minimise free energy by minimising the discrepancy between their sensory states and internal states – that is, their sensory data and the predictive model that accounts for that data – systems with generative models that incorporate inferences over multiple timescales execute complex action sequences (or action policies) that minimise the free energy expected over time. Corcoran et al (2020) describe three levels of generative model in the natural world. At the most basic level, such as single-celled organisms, the free energy principle operates in the context of simple generative models, enabling reflexive, self-sustaining behaviour to counter environmental perturbation. More complex organisms have evolved hierarchical generative models with parametric depth, where predictions at one level contextualise predictions at the level below, enabling behaviours which anticipate and adapt to environmental fluctuations. In still more complex creatures like us, generative models are afforded not only parametric but temporal depth, enabling us to track patterns of regularity and action-dependent contingencies over extended timescales. Temporally deep generative models enable us to simulate future states, make counterfactual inferences over multiple timescales, and give us the ability to evaluate not just (variational) free energy, but *expected* free energy – the free energy expected to accrue over the course of an entire action policy, or sequence of actions (Corcoran et al., 2020; Nave et al., 2022). While (variational) free energy is a function of past and present observations, expected free energy includes a prospective element, and has both pragmatic and epistemic implications. The expected free energy of an action policy depends on the extent to which the action policy is expected to resolve uncertainty and the extent to which it fulfils the prior preferences built into the model (Nave et al., 2022).

Action selection over multiple spatiotemporal scales is enabled by hierarchical, temporally deep generative models, which enable counterfactual simulations and inferences about evolving causal relations and future possibilities. Motor action is initiated when high precision counterfactual proprioceptive and interoceptive predictions set off a cascade of prediction error signals, which are in turn quashed when we move the body to align the sensory flow with the predicted sensations. Action policies – action sequences that are spread out over extended periods of time – unfold in the same way: high level priors, incorporating long term goals and preferred states, contextualise, predict, and optimise lower-level inferences over shorter time periods through the same process of prediction error minimisation, which are then unpacked in moment-to-moment action-perception loops (Clark, 2020; Nave et al., 2022). Thus, action policies are accommodated within the predictive processing framework as high level predictions that filter opportunities and guide behaviours at shorter timescales without positing additional mechanisms of motivation, desire, or cost functions (Clark, 2020). In selecting action policies to minimise expected free energy, an agent must model itself, as an endogenous controller of its own current and future sensory flow.

3.5. The Self-Model

The self-model, in the predictive processing paradigm, is the part of the overall generative model which models the generative model itself. In the same way that we use internally available information – inference and observation – to model the environment, we model ourselves as part of the causal structure of the world (Allen & Tsakiri, 2019; Hohwy, 2013; Hohwy & Michael, 2017). The self-model is a necessary part of the world model, enabling us to intervene and interact with other components of the environment. Having a model of itself enables the system to attenuate sensory input when the system itself is the cause of those sensations, lowering precision on sensations such as saccadic eye movement, blinking, or the optic flow of our own movement through an environment, as well as to predict and bring about preferred sensory states (Clark, 2016; Hohwy, 2013). The self-model emerges out of the same inferential processes we employ to model the environment.

Our internalised model of our body is constructed out of interactions with the environment; we model ourselves as objects within a spatially ordered world of objects by making inferences about the causes of visual, audial, and tactile sensory input, and bringing the model progressively into line with sensory observations. These exteroceptive predictions are enriched with interoceptive predictions and observations – feelings of hunger and satiation, fluctuating glucose levels, breath rhythm, and heart rate (Barrett & Simmons, 2015) – and contextualised with social and environmental information. These streams of interoceptive and exteroceptive predictions are amalgamated into a unified model of self. In a system that tracks patterns over multiple spatiotemporal scales, the self-model is constructed as a hierarchical, temporally extended locus of experience and agency.

We model ourselves, both implicitly and explicitly, not only as physical subjects but as temporally extended, psychological and autobiographical subjects. In systems with hierarchical and temporal depth, the self-model incorporates layers of complexity, temporal structure, and context, into a multi-layered construction which in turn contextualises all kinds of experience. Moment-to-moment exchanges with the environment are contextualised by interoceptive predictions; events are contextualised by personal narratives and timelines; interpersonal exchanges are contextualised by relationships which are contextualised by social roles; generalised, long-term predictions about our life trajectory contextualise the actions we take year-to-year and day-to-day. We model ourselves and others through the ongoing, iterative, and hierarchical processes of predictions, sensory observations, and precision-mediated error minimisation.

In summary, according to the predictive processing framework, perception, cognition, and behaviour are motivated by the biological imperative for organisms to minimise the mismatch between their sensory states and the probabilistic model that accounts for those states. In order for living systems to resist the tendency to disorder, any adaptive change, whether in rapidly changing moment-to-moment environmental exchanges, extended projects and life trajectories, or evolutionary timescales, must minimise free energy (Friston, 2009). The free energy principle involves a computational strategy with three components: (1) Recognition density – top-down predictions that encode beliefs about the hidden states of the world that cause the sensory data (2) Incoming sensory signals, which carry information about hidden states of the world and the system itself, and (3) the precision-weighted difference between the two – the prediction error. Perception and action work synergistically to minimise prediction error, as the brain continuously oscillates between updating the predictive model to accommodate sensory information (enabled by perception) and modifying the sensory flow to accord with the predictive model (enabled by action). In systems with hierarchical, temporally deep, generative models, predictions at lower levels, or over shorter timescales, are parameterised by predictions at lower levels and over longer timescales. Such generative models necessarily model themselves, enabling counterfactual simulations and evaluations of action-based contingencies over multiple timescales. The predictive processing account provides an elegant, coherent framework of the cognitive structures that support the phenomenology of agency and choice, enabling us to replace the ill-formed question of whether we have free will with a naturalistic enquiry into agency, choice, and autonomy.

4. Autonomy – Freedom to Create Ourselves

In this chapter, I argue that the agent is constituted by the system as a whole – the entire embodied intelligence (the brain and the nervous system) rather than only the top-level reflective viewpoint, or self-model. Mind and body, conscious and unconscious processes are in a continuous state of reciprocal causation. As such, the division between conscious and unconscious processes is arbitrary. The agent is in an ongoing process of self-creation through action, exploration, and reflection, continually setting up causes and constraints both in the external environment and in its own cognitive architecture.

4.1. Self and Agent

Ismael grounds her account of free agency in the capacity for self-governing systems to experience, think, and act as a unified self[16]. The self, in Ismael's account, is an emergent, internal point of view; a structure that unifies streams of information and enables coherent, all-things-considered judgements. The unity of self is unpacked in terms of synthesis – the integration of streams of sensory information into a common frame of reference; univocity – the capacity of the system to attribute its own thoughts, beliefs, attitudes to itself via linguistically mediated thought, thereby enabling multiple streams of internal information to be resolved into a unified, coherent collective voice; and dynamical unity – the capacity of the system to act on its own behalf, and exert top-down volitional control over its own behaviour according to the interests and priorities of the system as a whole. It is this unified self, as a high-level, reflective standpoint, that separates the things that I do, from the things that happen to me. (Ismael, 2014, 2016). And it is because I possess this standpoint as a reflective subject that I am capable of acting freely, of exerting top-down control over my beliefs, choices, and actions. Ismael's three functions of the self are fluently accommodated by the predictive processing account, though without the need to position the emergent self as the agent.

In the predictive processing paradigm, the capacity for unified perceptual experience, decision making, and directed action over multiple timescales is fulfilled by the hierarchical, temporally deep, predictive model. Synthesis is inherent in structured perceptual experience. The subject-object division, the stabilised conception of a spatiotemporally ordered world,

---

[16] See Section 2.3.

and the modelling of oneself as a locus of feeling and action are all deep structures of the unified generative model – the result of prediction, sensory data, and precision-mediated prediction error. Univocity – the capacity for higher level integration of the component parts of the system, enabling a unified, interoceptive viewpoint – is built into the hierarchical model which necessarily leans towards coherence via prediction-error minimisation processes. Dynamical unity – the capacity for a deliberative standpoint and top-down control of our own behaviour – rests on the capacity to model multiple future possibilities, infer causal pathways over multiple timescales, and select actions which align with the interests of the system as a whole. The hierarchical, predictive mind integrates layers of personal and sub-personal processing into a unified model which experiences, thinks, and acts as a unified whole.

In the predictive processing framework, the self is part of the mind's overall generative model. The subject-object divide, our conception of ourselves as physical entities in a spatiotemporally ordered world, and our sense of being psychologically continuous subjects embedded in a network of social relations are all products of inference – components of the probabilistic model, constructed out of experience-based hypotheses about the causal structure of sensory input. The self-model – our conception of ourselves as physical, feeling, psychological beings – is part of the infrastructure of thought and action, enabling us to model ourselves as agents embedded in the causal structure of the world, and indeed, to act as causes in the world (Hohwy, 2013). Similarly, we model others as agents with sensory expectations that mirror our own. Andy Clark suggests (2016) that we predict the behaviour of others using the same encoded sensory expectations that drive our own actions, to infer the intentions of other agents. Just as we understand and gain control of our own motor actions by learning the proprioceptive and interoceptive patterns associated with specific movement trajectories and incorporating those patterns into our generative model, we develop mirror systems that enable us to transpose those sensory expectations onto others, thereby learning to understand and predict the actions of others. By simulating our own motor routine in response to visual input of the movements of others, combined with contextual information, we are able to understand others' intentions (pp.139-140). By modelling ourselves and others as agents, we are able to predict complex contingencies and infer the downstream effects of our own and others' interventions.

Because the self-model is temporally extended, interoceptively available, and central to all sorts of top-level processes, it is easy to conceive of the self as the source of agency, but the self is merely an aspect of the predictive model – a construction within the unified point of view, rather than the whole unified package, or driver of behaviour. By default, the sense of being or having a self is built into the structure of experience and our interactions with the world. The self-model is constructed both implicitly and explicitly; we explicitly cultivate, refine and affirm the self-model through social roles, narrative, and identity construction, but we also have an implicit sense of self that operates as a kind of gravitational centre of psychological experience. However, although the sense of self as subject and agent is incorporated effortlessly into moment-to-moment processing, the self is more fragile than we might ordinarily take it to be. The self-narrative is constructed and reconstructed, our autobiographical past is constantly under revision, the sense of identity is context-dependent and requires ongoing maintenance, and the feeling of being the source of centralised control is constantly in flux. Although the metacognitive recognition of the constructed self is not ordinarily part of everyday awareness, it is easy to grasp. If we turn awareness on itself and look for the self introspectively – that is, if we increase precision on the internal locus of experience – it becomes immediately clear that the self is merely a contraction within a wider unified field of conscious awareness. The self which is assumed to be the subject of experience is merely an object of, or construction within, a still more expansive experience.

The self is not the agent; the self is a construction within the agent. The self is a high-level construction within the structured field of conscious awareness, which is grounded in the brain and nervous system, which itself constitutes the agent. The agent is the whole embodied intelligence, the entirety of the brain and the nervous system. It is the whole system that constructs the world and the self-model within it, the subject-object division (the stabilised conception of oneself as a sensorimotor presence distinct from the environment) and the collection of narratives and identifications that comprise what we think of as *I*. It is not the self-model that acts, but the whole system. Modelling ourselves and others as temporally extended entities enables us to predict the sensory and other consequences of actions, weigh long-term and short-term interests, and balance priorities. While modelling the self as an enduring entity or source of agency across time, is essential in order to deliberate, act, and implement action sequences, the self is not the source of deliberation.

Moreover, conscious and unconscious levels of processing are so deeply interconnected that the distinction between deliberated and automated, or reflective and reflexive behaviour is not as clear-cut as it appears. Conscious, personal-level processes cannot be cleanly separated from the rest of the system. Unconscious and conscious processes, mind and body are in continuous reciprocal exchange, with conditions at each level of processing setting constraints and possibilities for the other. For example, top-level, linguistically mediated deliberation takes place within the bounds of sub-personally computed action space. The possibilities for action that arise for consideration, along with the frameworks that are used to select from among those possibilities, all arise in consciousness automatically; the content and structure of conscious experience is rooted in unconscious processes. Conversely, unconscious processes – internal processes which are not available to the deliberative standpoint – are influenced by top-level, conscious direction. What we consciously reflect on, deliberate over, and direct our attention to feeds back into the system, creating context for sub-personally computed processes.

Likewise, there is continuous reciprocal exchange between mind and body. Consciously directed practices alter the neural substrate itself, which in turn affects both conscious and unconscious levels of processing, and alters the quality of subjective experience. Effortfully learning a second language increases hippocampus volume (Mårtensson et al., 2012); regular meditation practice decreases the volume of the right amygdala and increases subjectively evaluated ability to cope with stress (Gotink et al., 2018); intense aerobic exercise increases the volume of the anterior mid-cingulate cortex, which in turn enhances the capacity for effortful practice and the ability to override the impulse to disengage in challenging tasks in other domains (Touroutoglou et al., 2020). Similarly, physiological fluctuations such as variations in hydration, oxygen, and glucose levels, together with the brain's sub-personal evaluations of anticipated demands on these resources impact subjective emotional and psychological experience (Barrett, 2017a).

The agent is the brain and the nervous system, from which all structured experience – including the implicit and explicit construction of self – arises. While self and selves are built into the predictive model, and enable us to model ourselves and others as agents, we can talk about agency as a capacity or function of the whole system, without identifying the self as doer, decider, or believer. The predictive mind structures perceptual experience, models itself as an integrated locus of value, belief, and agency, deliberates and acts as a unified system over multiple timescales. It is within this integrated system that the world is perceived and understood, opportunities for action are grasped, choices are made, and plans are formulated.

4.2. The Mechanics of Agency and Choice

Commonplace belief in free will is grounded in the experience of agency. By default, we experience ourselves as intentional agents, with ownership and control over our choices and actions, and this is built-in to our understanding of ourselves, the world, and other agents. Phenomenologically, agency is the sense of that an action is caused by one's intention - a sense of ownership or control that is associated with volitional actions. Hohwy (2007) suggests that this sense of ‘mineness’ is due to prediction confirmation, since the sense of self and ownership must have a frame of reference. Agency is a precision-mediated prediction error minimisation strategy.

Action, in the Predictive Processing framework, is a precision-mediated prediction error minimisation strategy – the result of the sensorimotor expectations encoded in the generative model. Over multiple timescales actions and action sequences are initiated by counterfactual predictions which are cashed out by simple reflex arcs (Clark, 2016). Endogenous control of motor movement is enabled when the brain assigns high precision to counterfactual proprioceptive predictions and the system moves to align its sensory input with the predicted sensations, thereby cancelling out the error signals generated by the counterfactual prediction. Layers of nuance and complexity are dynamically incorporated into the bidirectional, hierarchical flow of information processing. Predicting the sensory consequences of reaching for a cup of coffee, for example, initiates a stream of proprioceptive predictions about the unfolding states of muscle spindles, tendons, and joints involved in that movement trajectory. This in turn generates a cascade of prediction errors which are systematically cancelled out as reflex arcs bring about the predicted movement, thereby fulfilling the proprioceptive predictions (Clark, 2020). This same mechanism of prediction error minimisation is responsible for endogenous control of behaviour over longer timescales. Precision weighted long-term predictions contextualise the flow of information and entrain action policies (sequences) over shorter timescales. These policies in turn contextualise the processing of real time action-perception loops, determining the salient features and perception of affordances in the environment, and entraining in-the-moment motor action.

The predictive processing account eliminates the classical distinction between belief states and motivational states and integrates both belief and motivation into a unified computational architecture, in which behaviour is driven by the dynamic hierarchy of precision-weighted probabilistic predictions and error minimisation processes (Clark, 2020). In the common-sense view, beliefs and desires contribute to action selection in different ways. My belief that there is drinking water in the kitchen contributes to the action of walking to the kitchen only if I desire to drink a glass of water. Conversely, my belief that I cannot fly prevents me from climbing onto the roof, despite my desire to fly from the top of the building. Under the predictive processing paradigm though, belief states and motivation states are not fundamentally different cognitive kinds but are accounted for by a rich continuum of precision-mediated predictions – predictive hierarchies that include predictions about our actions and the interoceptive and proprioceptive consequences of those actions, and extend from the specific to the general, immediate to long-term, and concrete to abstract. High-level predictions function as controllers and motivators of behaviour, both predicting future states and preparing the agent to bring those states about (Clark, 2020). In this elegant, unified, and powerfully explanatory account of action selection, high level predictions stand in for desires and intentions, proprioceptive and interoceptive predictions stand in for motor commands, and action selection, at every level and timescale, is mediated by precision.

The brain continually generates predictions that guide action and perception, which play out in both conscious and unconscious levels of processing. At every level of the cognitive hierarchy, the evolving flow of information – the precision-mediated flow of prediction and prediction error, the separation of signal to noise, and the opportunities for action that are perceived in the environment – is modulated by information by higher or adjacent levels, according to task and context. The generative model, in which actions and action policies are contextualised, entrained, and controlled, is an intricate, dynamic network of predictions which incorporates both sub-personally computed expectations based on past experience, and personal-level predictions that are employed in the process of deliberation and choice.

The process of deliberation and choice involves the simulation of multiple counterfactual predictions, optimisation of precision weightings on those predictions, and the selection of predictions that cohere with the hierarchical model. The ability to deliberate, plan, and formulate distal goals is underpinned by our capacity for top-down cognitive control of our own thought processes, which evolved out of the ability to simulate the consequences of action, without performing those actions. (Pezzulo, 2012). High-precision proprioceptive and interoceptive predictions, over short and long-term timescales, are realised through action and action policies. Action selection is biased towards predicted states that fit coherently into the overall model. If a predicted state cannot plausibly be attained from the current state (as is the case for predictions that involve flying off the top of buildings, for example) it is assigned low precision, and will not entrain action or action policies. In that case, error minimisation will be achieved via perception, rather than action (Friston et al., 2014). Together, the capacity for counterfactual cognition and cognitive control, and the ability to simulate more than one actionable scenario at a given point, afford the agent multiple possibilities that can be realised through action; we predict experience and enact those predictions.

4.3. Freedom is Relative to Context

Action selection takes place within a nested, contextual hierarchy of predictions, in which action and perception oscillate in continuous, reciprocal exchange to minimise prediction error. The context for every level of processing is set by prior expectations flowing down from the level above, while error signals fed up the hierarchy continually update those expectations, creating a continuous, circular transfer of information between levels (one which resembles Ismael's description of the structure of self-governing systems). The hierarchical system is built for coherence; as perception is optimised, the agent constructs a model of its world; as action is optimised, the agent learns to navigate and intervene in its world. Action and perception work in synthesis to deliver a world in which opportunities for action are built into the structure of perceptual experience according to the expectations built into our model (Clark, 2016; Nave et al., 2022). Perceptual experience is contextualised by our tasks and goals. The entire domain of structured experience, and therefore the entire scope of possibility for action, is defined by the structures that comprise the predictive model.

Deliberation and choice unfold within layers of context embedded in the probabilistic model, which determine the contours of the choice space and define the parameters within which the agent can exert endogenous control over its behaviour. The scope within which the agent can make rational, reflective choices, and the frame of reference for resolving conflicting predictions is a function of the patterns of information in the overall model – the overarching context which incorporates the behavioural repertoire, conceptual frameworks, the mapping of causal pathways and contingencies, and conceptions of value. Within this space, the constructed self-model operates as a high-level context that plays a pivotal role in action selection – a contextual model which incorporates in which layers of sub-personal and personal-level expectations about the kinds of states I occupy, the kinds of experiences I have, and the kinds of things that I do. Representations of anticipated proprioceptive and interoceptive states are built into every plan, choice, and action. We act in ways that cohere with our probabilistic expectations about ourselves and the kinds of things we do.

Freedom – the scope of behavioural flexibility and capacity for top-down, non-mechanical action selection – is relativised to the generative model; that is, to an experiential rather than objective frame of reference. For any given configuration of agent and event, there is no objective set of possibilities for action. The architecture of the choice space is defined by the set of possibilities the agent perceives is available in a given context. Epistemic limitations are built into the structure of possibilities; we can act only within our understanding of ourselves and the world. The limits of motor control, volitional action, and action policies are defined by the nature of the simulations available in the generative model. Conceived possibilities for action, the structure of deliberation and choice, and the resolution of those processes are nothing but the bidirectional processing of information, and the systematic, layer-by-layer minimisation of prediction error. The degree to which we can exert rational, reflective control over – or at least influence – our actions (within the range of possibility afforded by circumstance) depends on our ability to introduce, alter, or select context within the model.

4.4. Autonomy: the Freedom to Create Context

One worry about this picture is that even deliberated choice and action selection may be cast as automated processes of prioritisation, in which competing hypotheses, goal states, or action policies are sorted, weighted, and selected according to sub-personal contextual parameters. If this is all that the brain is doing, the account leaves no room for meaningful top-level choice and genuinely free action selection. But crucially, the predictive framework that underpins judgement and action is open to top-down intervention. Our capacity for cognitive control enables us to introduce and alter context, which allows us to actively shape the very structures that define the parameters of choice and deliberation. This ability to alter the way information is structured and processed enables us to shape the frame of reference from which possibilities, choices, and decisions emerge and transform the structures that define the scope of behavioural possibility. The deep interdependence of personal and sub-personal processes, and the unified processes of action, perception, and cognition, allow the agent to actively shape the deep structure of the mind. Continuous reciprocal exchange between linguistically mediated, personal-level thought and sub-personal processes enable us to set high-level context for sub-personal processes via conscious-level thought and practice, which affords us far-reaching, top-down influence over our own behaviour.

We continually fold context into the model from the top level. Our deliberated actions and judgements, reflections, enquiries and interpretations are constantly being fed back into the model and incorporated into the frame of reference that structures future experience. We do this in both effortful, self-directed ways, and in absent-minded, other-directed ways. When we engage in rational enquiry, we introduce a high-level contextual framework by which to interpret and evaluate inferences. When we consciously formulate goals and plans, or clarify values, we introduce high-level predictions that play out over long timescales. When we actively reframe an experience, we recontextualise that experience, and that context will be brought to bear on comparable future experiences. When we deliberately expand our behavioural repertoire, expose ourselves to new environments, or explore a body of knowledge, new information is incorporated into our overall generative model, within which all experience is structured and contextualised. Shallow-processed, frictionless consumption of information also shapes the underpinning contextual framework, biasing processing towards experiences and interpretations that cohere with that information. The agent is engaged in a continual process of self-construction, as it feeds top-level context into the predictive model, which in turn alters the underpinning contours of the choice space, and sets the conditions for automated processes of prioritisation.

There are many ways to individuate ourselves within the constraints of our genetics and environment, and many ways we can use the mind we have right now to alter the course of future events, and change the scope of future choices. While we cannot integrate predictions that are radically incongruent with our overall generative model, initiate actions for which we have no frame of reference, or perceive opportunities for action that prior experience has not prepared us for, we can push out the edges of our current paradigm, ‘try on’ alternatives, expose ourselves to new experiences and information, and transform our overall frame of reference through incremental changes over time. We are free to deconstruct and reconstruct the frameworks that inform action and choice over multiple timescales, and we do that actively when we set goals, for example, commit to a programme of training or behaviour change, learn philosophy, practice meditation, or engage in self-enquiry. As we thread context into the model through top-level processes, we alter the deep structures of the mind and its neurobiological underpinning. Conscious-level interventions modify neural structures and connections. The brain and the nervous system are being continually re-sculpted through experience (that is, the agent is continually re-sculpting itself). As neural representations and patterns of influence between neural populations change, the possibilities for action change.

4.5. Attention and Information

The scope of freedom is defined by information and attention. Judgement, choice, and action are information processes, describable in the information theoretic terms of predictive processing and the free energy principle. The entire scope of belief and action is contained in information structures and processes of the mind. The limits of behavioural flexibility and control are defined by informational structures embedded in the generative model – the network of probabilistic beliefs that structures experience, and within which choices are made, priorities are set, and actions are initiated. As we select, process, and create information, we shape the frame of reference for deliberation and choice. We constantly exert meaningful influence over own behaviour, through top-down selection of information, and endogenous control of our attention.

Our capacity for attentional control affords us top-down influence over the selection and precision weighing of information, and the ability to alter the frame of reference that contextualises our decisions about what to believe and how to act. Recall that the mechanism of agency is precision – the relative neural gain accorded to either the prior beliefs or error signals (Hohwy, 2013). If higher precision is assigned to counterfactual priors, action follows; if higher precision is assigned to the error signals, the predictive model is updated. Importantly, precision is implemented by attention. Attentional control gives the agent the capacity to direct the process of selecting action and action policies within the limits of the model, and to extend those limits by deliberately attending to and integrating new information into the model.

Were it not for our capacity for top-down attentional control, the predictive processing scheme could be used to argue that the phenomenology of deliberation, choice, and volitional action is merely an epiphenomenal consequence of a highly complex but mechanistic information processing system. However, the distinction between endogenous and exogenous attention suggests that top-down, directed attention cannot be reduced to a mechanistic process. The distinction between top-down (endogenous) attention and bottom-up (exogenous) attention can be illustrated most clearly using the example of visual attention. When we attend to the visual environment, movement in the periphery of the visual field automatically triggers eye movement or body movement that bypasses conscious evaluations of salience to focus the moving object on the fovea, allowing us to attend to that part of the scene in more detail – an involuntary, bottom-up process of attention allocation. But we may also attend to an area of the visual field in a targeted way. When we scan the environment to look for our keys, for example, or scrutinise a face to check whether we recognise it, we employ endogenous attention. The capacity for attentional control, along with the pivotal role of attention in action production has implications for freedom. The ability to direct our attention enables us to modulate precision (neural gain) on any aspect of perceptual or cognitive experience, which affords us enormous variability in how we understand and navigate the world, in ways that surpass mechanistic processes. Attention is a pivotal link in endogenous control over action and action policies.

Attentional control is an important strategic intervention for exercising and expanding the scope of personal freedom. We can deliberately cultivate our attentional capacity in order to bring more depth and detail to aspects of our embodied or cognitive experience (as is the case with mindfulness, meditation, or concentration practices) or to suppress the impulse to attend to aspects of our mental life (Mamat & Anderson, 2023) thereby reducing its sway over our behaviour. Conversely, neglecting to actively direct our attention in the ways that align with our high-level action policies diminishes our capacity to act on those policies. High-level predictions in the form of long-term goals are useful when they are implemented as short-term action policies, and in-the-moment motor control. Failure to implement those high-level predictions is at least partly a failure to attend to them, such that they do not sufficiently entrain actions at shorter or more immediate timescales[17]. Furthermore, neglecting to allocate attention in ways that align with consciously selected priorities leaves us susceptible to

---

[17] I am thinking here not of predictions that are deliberately superseded in favour of some other prediction or goal state – a change of plan – but of predictions that are retained but not actioned, such as the 80% of new year’s resolutions that fail only to be reinstated the following year, or partly finished projects that are stalled but not resolutely abandoned.

distraction traps which hijack the attention, and external manipulation of our informational environment which further undermines endogenous control and narrows or modifies the scope within which we can think and act. Depending on how we implement top-down control of our attention, behavioural patterns become more or less engrained, worldviews are expanded or narrowed, we become more or less functional, and we construct ourselves as agents with more or less top-level control.

4.6. The Autonomous Agent Modifies its Own Cognitive Environment

The highest level of freedom is the ability to construct, or at least influence the internal structures that underpin choice and action, such that automated actions, conceived choice points, and opportunities for action, are aligned with one's consciously chosen beliefs and values. Autonomy surpasses limited conceptions of freedom in which a top-level agent arbitrates conflicting desires, suppressing some and authorising others, or in which freedom merely consists of being unobstructed in acting on one’s desires. Autonomy is the deep freedom created by the agent itself, established as the agent transforms the deep structures of the mind, enabling the agent to think and act as a unified whole.

Our capacity for autonomy lies in our ability to intervene in our own cognitive environment, which sets the very conditions of our freedom. We not only have sensorimotor agency – the freedom and obligation to act and engage in reciprocal exchanges with the world – but cognitive agency too – the freedom to shape and direct our own minds. As we allocate attention, curate our information environment, think, move, and act in the world, we are in a continual process of constructing ourselves as agents and setting the conditions for our own future processing. With every experience, we modify our own neural structures; the brain and the nervous system are being continuously re-sculpted by everything we do. We can exploit the process of experience-dependent neuroplasticity to expand the possibilities for freedom, using deliberate, sustained practice, attentional control, and directed learning to shape the brain in ways that support optimal flourishing of the whole system. This progressive process of internal alignment expands the agent’s freedom to act on its own behalf.

Autonomous control over our mental life is not the default, but something we can gain by practice. The deepest structures of thought are rooted in our biology, early experiences, and cultural conditioning. With clarity and effort, we shape our own cognitive infrastructure, and go beyond the limitations of these inherited structures. We introduce new information, expose ourselves to new experiences, and generate high-level context, with far-reaching, downstream effects in our own thinking processes. By going to the edge of what our current paradigm, abilities, or experience allows, we gradually transform our minds and expand the parameters of our freedom. There is a lot of psychological and practical work involved in constructing highly functional models that afford the agent maximum freedom to act on what matters to them; this kind of freedom is a cognitive skill acquired by action, thought, enquiry, and practice.

We actively participate in our own self-creation and have far-reaching influence over the long-term parameters of our own freedom. Current actions and choices are incorporated into the network of probabilistic beliefs that set the frame for future experience, choices, and actions. We can act deliberately to modify our internal and external conditions, and to attend to, integrate, and interpret information in ways that support and enhance the freedom of our future selves. Conversely, our future selves may become prisoners of our own unexamined processes, habitual patterns, and constrained or manipulated informational environments. We can expand our capacity for freedom by engaging in the work of psychological growth and integration that aligns our deep processes with our top-level frameworks, values, and priorities. To the extent that we establish the conditions for own freedom, we are capable of autonomous action.

5. The Parameters of Freedom

Building on Ismael's enquiry into the physics of human freedom, we have explored the cognitive structures that support agency, choice, and autonomy, and drawn on the insights of predictive processing to develop an enriched, layered account of freedom that remains firmly grounded in a naturalistic worldview. The account does not deliver the kind of free will that aligns entirely with our pre-philosophical intuitions; the default experience of choice and action being driven by intention, or by some kind of self, is merely an artefact of the way the embodied intelligence of the brain and nervous system models itself and other agents. However, this nuanced, nested picture points to an expansive vision of freedom that surpasses our pretheoretical intuitions, and provides clues about how we can develop our capacity to think and act freely, and widen the parameters of our own freedom. We have seen that freedom exists on a gradient, supported by structures that reach deep into physical reality, cognitive processes that allow for constrained but meaningful deliberation and choice, and causal pathways that allow agents to remodel themselves from the top down, enabling us to transform the very conditions of our own freedom. This chapter revisits these three levels of freedom, moving from Ismael's account of the physics that allow for free action, to the cognitive architecture that supports volitional choice and action, to the possibility of constructing ourselves as fully-developed, autonomous agents. I discuss the limits and possibility of freedom, and explore how we might expand the parameters of personal freedom by intervening in the underpinnings of our own psychology and neurobiology.

5.1. Freedom as Possibility

The possibility of freedom is built into the fabric of reality. At the deepest level of physical reality, the universe is a network of logically interconnected variables. As discussed in Section 2.1, determinism at the microscopic level does not entangle us in a predetermined sequence of causes that reaches back to the origin of the universe; as Russell observed, causation disappears from the microscopic level of physical description altogether. When the notions of causality and determinism are disentangled which, as Ismael and Hoefer argue, is precisely what is required when physics is understood on its own terms, we are left with determinism as a network of logical interconnection. Instead of asymmetrical compulsive relationships at the micro-level, we find a matrix of interdependent variables that lack any intrinsic direction – deterministic laws, invariant under time-reversal, that describe logical, rather than causal relationships. This acausal, deterministic underpinning of macroscopic reality means that our choices and actions are fundamentally unbound by prior conditions at the microscopic level. Determinism, far from being a threat to freedom, is the basis of a physical reality in which reliable intervention – the ability of one part of the whole to act on another part of the whole towards some strategic ends – is possible. Were it not for the functional interdependencies inherent at every level of physical reality, systems could not sustain themselves or interact with each other in reliable ways. Determinism is the basis of a universe which is open to intervention.

Likewise, at the macro level, we are not bound by global laws. As Ismael argues, our actions and choices contribute to global, emergent patterns, but there are no pre-existing laws of global evolution that determine our fate, or that condemn us to a future that is set independently of our reflected choices and decisions. Our choices and decisions are part of what structures the future. From the point of view of the embedded participant in History, openness is a property of events before they happen and not after; the future is genuinely open, and depends on our choices and actions (Ismael, 2016, chap.6). The openness that characterises future events as our lives unfold in A-Series time is completely compatible with the transcendent picture of the block universe, in which nothing is unfolding, the categories of past, future, and *is* – foundational concepts in our experiential and linguistic framework – break down, and the nature of being becomes inexpressible. From the embedded perspective, our decisions are made in the context of a universe in which there is more than one lawful possibility for how the course of events unfolds. From the transcendent perspective, our decisions are part of the integrated, logically structured whole.

Within this integrated whole, living systems, with senses coupled to the macroscopic level of reality, infer and utilise the causal information available at the macro-level to sustain their existence. Causal inference is built into the structure of perception, cognition, and action. Because the thermodynamic gradient means that actions taken in the present propagate asymmetrically into the future[18] the macroscopic level of reality is rich with information about the past, and we are able to shape and imprint the future by reordering the present. Adaptive systems, bound by a Markov blanket, necessarily model their environment and act upon it in order to counter the dissipating effects of entropy and resist, for a finite amount of time, the second law of thermodynamics (Friston, 2013). It is an essential condition of life

---

[18] See Section 2.5 for a discussion of Ismael's explanation of emergent macro-level asymmetries.

itself, that systems minimise the discrepancy between their model and the sensory flow; we must act to structure sensory input to align with the probabilistic expectations embodied by our model (which includes information about preferred states, or states which support the continued viability of the organism (Friston, 2010, 2013)[19]. Moreover, agents that embody generative models which simultaneously model the world (and themselves) over multiple timescales, must act to minimise error accordingly; creatures like us must act to align sensory inputs with expectations over the timescales built into the structure of our models. Thus, the freedom and obligation to act is built into the deepest structure of living systems.

The upshot here is that the possibility of free action is fundamentally built into our existence as adaptive agents situated in space and time. At a very basic level, we are free to act on, or intervene in our environment. The material universe allows for it, and our own nature as biological systems requires it. Moreover, complex, adaptive systems like us, that have evolved the capacity for cognitive control and counterfactual cognition, have the ability to simulate multiple possibilities, make choices, and bring behaviour under endogenous control. We can – and we must – make decisions and act.

5.2. Freedom as Volitional Choice and Action

In a minimal sense, freedom is just the ability of a system to influence its internal information processes such that its behaviour is under its own control, rather than under the control of the environment. At least in a rudimentary form, this capacity for voluntary behaviour – the ability to initiate action based on internal processes – is built into non-human animals, just as it is in us (Heisenberg, 2009). Behavioural flexibility and endogenous control of action is essential for the survival of adaptive organisms (Brembs, 2011; Heisenberg, 2009). The ability for a system to initiate action on the basis of internal variables or processes, such that the causal link between stimulus and response is overridden, has been selected for because it provides a key advantage in outmanoeuvring predators and competitors, and enables organisms to try out varying responses to novel challenges and unfamiliar environments (Brembs, 2011). Even simple creatures, such as unicellular organisms and insects vary their behaviour independently of environmental stimuli and history, and produce novel responses to identical conditions even in cases where it would be more efficient to replicate previous responses (Brembs, 2011). Such actions, which are initiated on the basis of internal states,

---

[19] See Section 3.1.

independent of environmental stimuli, surpass the category of *behavioural response* (Heisenberg, 2009). According to Brembs and Heisenberg this capacity for behavioural flexibility and endogenous control of behaviour – that is, the capacity for volitional action which is decoupled from environmental causes – supports a naturalistic (and in their view, consciousness-independent) notion of free will as a non-metaphysical, biological property.

The picture of behavioural variability and endogenous control of action as basic functions of adaptive organisms fits well under the active inference (or free energy minimisation) paradigm. The behaviour of adaptive organisms is not merely a matter of random variability combined with reinforcement mechanisms. Adaptive systems are actively engaged in hypothesising the causes of sensory stimuli, and they do this by interacting with their environment in different ways to optimise perceptual inference. The degree of flexibility and endogenous control available is a reflection of the degree of complexity embodied by the model and scope of the behavioural repertoire. In unicellular bacteria, endogenous action selection is limited to swimming and tumbling behaviours. Flies can narrow attention to a restricted area of their visual field, and vary their flight trajectories independently of sensory input (Heisenberg, 2009). In humans, these simple processes of behavioural flexibility and endogenous action selection have evolved into the capacity for high-level deliberation, abstract reasoning, and systematic sequencing of actions and action policies. The ability to integrate increasingly abstract concepts and frameworks into our hierarchical models and make inferences over multiple timeframes brings the ability to engage in complex deliberation, infer intricate causal networks, and strategize over multiple timescales.

Freedom, as volitional choice, is a dynamic property of the network of interconnected predictions embodied in our models – the constantly changing interplay of neuronal populations influencing other neural populations; the contours of behavioural flexibility, choice, and control continually shift as context is configured and reconfigured. Choice is enabled by the brain’s capacity for cognitive control and counterfactual cognition which is implemented at varying levels of complexity, from simple, binary, in-the-moment decisions to temporally extended action sequences that navigate complex contingencies and changing social and environmental dynamics. Possibilities for action, in the form of counterfactual interoceptive and proprioceptive predictions, are generated within the constraints of the behavioural repertoire, inferred environmental conditions, habitual patterns, internalised norms of behaviour, and whatever other factors are incorporated into the frame of reference.

We select from among these predictions by implicitly or explicitly invoking a higher level context within that frame of reference, and fulfil the selected prediction by reconfiguring the sensory flow to align with it[20]. We encounter a continuous stream of choice points that emerge from and are resolved within the network of predictions and information processes that comprise the generative model.

Our capacity for volitional choice enables us to resolve competing hypotheses about what to believe and what to do from a high-level deliberative standpoint. To facilitate economical processing, some volitional choices become automated, habitual, context-dependent action sequences, requiring minimal reflection. Others invite (or demand) conscious reflection, deliberation and varying degrees of effortful processing; we weigh competing priorities (as predictive hypotheses) set up proximal and distal causes, devise long-term strategies, initiate and navigate complex chains of events. However, even conscious, top-level deliberation takes place within a largely automatically generated action space. Much of the machinery of choice is driven by unconscious automated processes, mediated by conceptual and value frameworks, reality paradigms, priorities, and influences that were put in place by processes outside of our control and that bypass conscious reflection.

Volitional choice and action are straightforward (though constrained to a greater or lesser extent) implementations of freedom, but volition is not necessarily the highest level of freedom. Volitional choices that emerge out of imposed, distorted, or manipulated frameworks or fragmented, disunified agents cannot be regarded as fully-fledged expressions of freedom. Just as freedom can be undermined by external influences such as coercion, deception, and manipulation – influences that corrupt the model within which choices are made – freedom can also be undermined by internal influences. When volitional choices and actions are driven by imposed or internalised maladaptive frameworks, unconscious, non-rational influences and processing errors, psychological conflicts, self-deception, biases, or judgement errors, our ability to think and act freely is threatened. Nevertheless, both our in-the-moment choices, and the predictive frameworks we deploy to contextualise and inform those choices are susceptible to top-level intervention. As Ismael (2016) points out, we are

---

20 As discussed in Chapters Three and Four, under the predictive processing paradigm, action selection and choice are precision mediated error minimization strategies, explainable without positing additional mechanisms of desire, intention, or will.

not only free to make decisions, we are free to choose the basis on which we make our decisions (p.197). To do so is to move towards a deeper level of freedom. The potential for truly autonomous action lies in our ability to transform our own frameworks and information processes, and establish an internal relation of deep unity and alignment.

5.3. Freedom as Autonomy / Self-mastery

There is an important distinction between volitional action and autonomy. Volitional freedom is exercised when agents simulate alternate behaviours and select from among them. Autonomy emerges when we actively reorganise the internal frameworks that underpin choice and experience. This is a higher-level implementation of volitional choice, in which the agent actively creates context and develops a coherent overall model to guide decision-making, thereby redefining the choice space itself. The capacity for autonomous action is established as the agent restructures its own processes and transforms its model, such that volitional choices are contextualised by a predictive framework that aligns with the agent's freedom and flourishing, enhancing and expanding the agent's ability to act on its own behalf[21].

If we do not have sufficient authority over the frameworks that contextualise our choices – that is, if we have internalised frameworks that conflict with our reflective, considered values, undermine our own agency, or serve the agenda of an external agent – then we fail to act autonomously. The frameworks – the predictive models, or hierarchical networks of probabilistic beliefs – we use to make decisions determine the nature of the possibilities for action under consideration and set the conditions for choice and resolution. We need not be entangled in addiction or subject to coercion for our authority to be compromised or abdicated; we may merely, through negligence, misperception, or enculturation, for instance, be entangled in a corrupted internal information processing environment. For example, people very often fail to act in the ways they consider optimal. Nevertheless, such suboptimal action selections are necessarily coherent within the predictive framework they are conceived and selected in. The failure to act in ways that align with the flourishing of the overall system

---

[21] The ability to make decisions on one's own behalf is one of the loose definitions of freedom offered by Ismael (2016), "To be free is to be able to make decisions on your own behalf. Nothing more and nothing less. So carry on as you were, making choices with confidence that not just your fate, but the fate of the universe hangs in the balance as you do" (p.211). My point here is that the extent of our ability to decide and act on our own behalf is variable, and that this capacity is itself open to self-intervention.

is not a failure of self-control or weakness of will in the way that might typically be thought – action selection is a prediction error minimisation strategy, rather than an implementation of will or intention – but rather a failure to construct and attend to an overarching model that guides decisions in a way that preserves or expands the parameters of one's freedom. Alternatively, such sub-optimal outputs may result from a failure to deconstruct disempowering or faulty aspects of the internal model that undermine the agent's ability to act in ways that support its own autonomy and integration. Autonomous self-intervention enables the agent to make the appropriate repairs to the model, and intervene in the space within which choices are conceived and actions are initiated. Thus, the development of autonomy is a progression towards self-mastery – the internally directed transformation of the agent itself.

The movement towards autonomy is achieved by progressively unifying and integrating divergent parts of the model under a coherent overall framework that supports and expands the agent's freedom. Autonomy surpasses conceptions of freedom in which lower-order desires, appetites, and preferences are arbitrated or controlled by a high-level, constructed inner authority. In its fullest expression, freedom is not an internal battle for supremacy, but an internal alignment and integration of an agent's own information processes. The notion of freedom being an expression of inner authority is a picture of fragmented agency in which some part of the mind is free, while other parts are under its control. The fragmented agent whose attention is divided among conflicting parts of its own model cannot act as a unified, autonomous whole, and has not yet aligned its cognitive processes towards the integration and flourishing of the entire system. When the system is unified within the overarching progression towards flourishing – the fullest expression of the agent's nature – the conflict is genuinely reconciled. In the psychologically accomplished free agent, apparently conflicting motivations are transcended and unified, rather than arbitrated.

Autonomy is an enhanced capacity to act on our own behalf – a self-directed realignment brought about by enquiry, reflection, and deliberate practice. By bringing action, values, and rational thought into coherent alignment, we move from indoctrinated constraints and preprogramed limitations of thinking and action, towards a state of deep internal alignment and authentic empowerment. As the agent progressively resolves disunity in their model, deconstructs freedom-undermining values and frameworks, and transforms the deep structures of their mind, they restructure the internal foundation that contextualises their volitional choices. Autonomy then, is both a cognitive skill and a psychological state.

5.4. Expanding the Parameters of Freedom

The parameters within which we can think and act freely are defined (within the broader set of limitations defined by genetics and environment) by the structures of our predictive models. The structure of present experience and the possibilities for action conceived within it are products of the dynamic network of probabilistic beliefs that encode our action repertoire, interpretive frameworks, and experience-based expectations of self and world. Present experience is largely a reconstruction of the remembered past. Prior beliefs – sub-personally computed probability distributions – set the frame for direct perceptual experience, and are built into our encounters with the world. We actively anticipate a world which looks much the same as our past experience, tend to replicate experiences, themes, and actions, and encounter familiar patterns of limitation and opportunity as we reconfigure past experiences into present ones. At the same time, we constantly learn from experience and update our models, and these updated beliefs become new priors that lay the ground for future experience. This process can be strategically exploited; we can, through deliberate thought and action, systematically reorder our minds and environments in ways that support the development of maximum agency and autonomy.

Our ability to expand the parameters of our freedom lies in our capacity to intervene in our own cognitive processes, restructure our generative models with deliberate practice, and actively support the brain to generate increasingly functional models of ourselves and the world. In doing so, we transform the structure of experience in ways that support autonomy – personal responsibility for one’s own experience and an enhanced capacity to act on one’s own behalf – and freedom – the ability to select models of thought and action that support optimum functioning and exchanges with the environment. We can, through learning, enquiry, and action, diminish the influence of beliefs that have been incorporated into our models (via culture, conditioning, trauma, training, or error, for example) that undermine agency, or limit optimally functional exchanges with the environment. As we establish unified, coherent, and functional models of reality, resolve fragmentation in our models, eliminate freedom-undermining frameworks of interpretation and action, or expand our behavioural or interpretive repertoires, we increase our capacity to think and act freely.

There are many ways we can intervene in our behaviours and top-level judgements to create conditions for increased agency and autonomy. We might, at minimum, refrain from engaging in behaviours and environments that compromise or undermine our freedom in the long term. Obvious examples include addictive behaviours or technologies, communities whose price of admission is adherence to dogma or submission to an epistemic authority, and excessive exposure to misinformation – a miscalibrated orientation to reality is a disunifying influence on the minds and undermines rationally coherent decision making processes. Given that we embody a model of our environment, we might treat with caution information environments and technologies that surreptitiously erode our wellbeing, distort our worldview, or lull us into unthinking consumption of information, and instead curate our information environment to align with the mind we want to have. We might engage in philosophical enquiry, cultivate our capacity for rational thinking, and actively develop coherent overarching structures to guide judgement and choice. When we reset prior beliefs that limit freedom, acquire more functional models for thinking or interpreting experience, engage in the work of psychological integration, or expand our behavioural repertoire, we restructure the choice-space of future experience and, in doing so, expand the scope of our freedom.

In addition, we might reconsider how we process emotions; the way we process and regulate emotions has far-reaching implications for our ability to think and act freely. Emotions can influence judgements and decisions even when the emotional response is unrelated to the target of evaluation, and in ways that bypass rational, top-level direction or awareness[22] (Lerner et al., 2015). Experience is imbued with fluctuating feelings and emotions that can not only influence our judgements, they can compel us to act in ways that are counter to our best judgement; we may speak or act unwisely when our minds are clouded by anger or impatience, we may find ourselves avoiding tasks when they evoke feelings of unease, even if those tasks are self-selected; and we are subject to a myriad of social and environmental nudges that exploit our vulnerability to influence by evoking emotional rather than rational responses. Freedom partly depends on our ability to make judgements and choices that are not biased by faulty, emotionally driven evaluations. The scope within which we can think

---

[22] This is not to say that emotions cannot also influence judgment and action in useful and informative ways.

and act freely is expanded when our judgements and actions are unbound by internal compulsion or hidden influences which diminish our autonomy.

Resetting the emotional ground of experience – being deliberate about how we conceptualise emotional experience out of interoceptive information – can expand our capacity to think and act freely. According to Lisa Feldman Barrett (2017, 2017a), emotional experience is inference about interoceptive states, based on the brain's predictions about the body's anticipated energy requirements. The brain, along with anticipating sensory input and computing the best responses, constantly predicts the body's energy requirements, and proactively prepares the body's systems to meet energy demands by regulating physiological resources (such as oxygen, glucose, and blood flow). We experience this flow of interoceptive information in consciousness awareness as *affect* – the ever-present feeling tone of experience, that precedes the construction of emotion; that is, affect precedes the conceptualisation of feeling as an instance of an emotional category. Interoceptive information, as affect, is built into our concepts, infuses our judgements, decisions, and actions, and makes us susceptible to what Barrett calls *affective realism* – the tendency to evaluate objects and people as inherently positive or negative, and imbue concepts and objects of perception with what we take to be objectively positive or negative qualities (Barrett et al., 2016). As such, the way we process and interpret the affective component of experience has a crucial role in our ability to make rational judgements, and make decisions and action selections accordingly. Skilful emotional regulation enables us to interrupt automated responses, and counter the influence of emotional content on judgement and action selection.

Managing affect skilfully expands our ability to relate to the world from a deliberate, rational standpoint, widens our window of available, optimally functional responses, and enhances our capacity for autonomous action. Barrett suggests that we can effectively alter the way we process the affective component of experience by expanding our repertoire of emotional concepts, mindfulness meditation practice, or cognitive reappraisal (Barrett, 2017, Chapter 9). We might, for example, recontextualise the feeling aspect of experience by reframing affect as primarily delivering interoceptive information about body systems and anticipated resource demands, thereby interrupting the seemingly automated construction of emotion, and enabling us to put some space between our feelings, emotions, and judgements. Towards the same ends, we might practice mindfulness techniques that include the observation of feeling without conceptualisation. The practice of mindfulness (and other types of meditation) has the additional benefit of training one's capacity for attentional control. Given that our capacity for top-down control of our attention is critically important for the quality of our experience, the nature of conceived possibilities for action, and the process of choice and action selection,[23] mindfulness practice appears to offer a profoundly useful tool for expanding the parameters of our freedom.

When past experience impinges on present experience in ways that undermine an agent's ability to respond, engage, and interact with the world in the ways they choose, there are also more radical ways to reset the priors that diminish freedom. Researchers propose that the therapeutic efficacy of psychedelics such as psilocybin and LSD (in treating post-traumatic stress disorder, for example) lies in their ability of those substances to relax the precision weighting of high-level priors, which enables the brain to reset beliefs that it is not ordinarily able to update. Psychedelic compounds act to briefly destabilise and 'flatten' the free-energy landscape of the brain which creates a heightened state of plasticity, and receptivity to bottom-up information. As it restabilises, the brain is able to restructure its model, and reset the high-level beliefs that operate as gravitational centres of experience and contextualise the flow of information at lower levels of the hierarchy (Carhart-Harris & Friston, 2019). These interventions have immense potential to support the brain to go beyond conditioned limitations, outdated models, and entrenched viewpoints that limit or undermine freedom.

Freedom is always constrained, but the project of maximising our freedom partly depends on selecting the constraints within which we will act, and on making efforts to constrain our actions and action policies to those that preserve and expand freedom. Paradoxically, expanding the parameters of freedom does not depend on an unconstrained increase in possibilities, but on the deliberate shaping, and even reduction of the internal variable subspace – we can expand our freedom by narrowing the possibilities for action. Expanding our freedom is not necessarily achieved by generating a broader range of action possibilities (interoceptive predictions) or removing barriers to enacting those predictions, but by our efforts towards deeper integration, clarity and coherent expression. Moreover, some behavioural possibilities do not become available, either cognitively, or environmentally, until we narrow the scope of the action space. In order to be available for the kinds of

---

[23] As discussed in Section 4.5. Attention and Information.

experiences that most fully embody what I choose and value, I must voluntarily narrow my options, and reduce the possibilities in one area, in order to advance my freedom and autonomy in another. When I commit myself to any experience, or allocate my time and attention to a long-term project, I necessarily narrow the scope of my behavioural choices, and forgo many alternative possibilities. We are free to the extent that we are able to deliberately select values and decision-making frameworks that support freedom, and act on those frameworks. This is the result of long-term efforts to construct a coherent, overarching framework to guide action, and consistent, deliberate practice to align our actions with rational thought.

5.5. Architects of Our Own Experience

We are engaged in a continuous process of constructing and directing experience through top-level processes of attention, deliberation, judgement and action. Conscious processes and practices shape immediate, non-reflective perceptual experience, create context for automated, default behaviours, and reconfigure the deep structures of our minds. At the same time, these conscious-level processes alter the neural environment which underpins conscious experience itself; the brain modifies its own structures and connections in response to deliberate practice, new experiences, and repeated patterns of behaviour. The cycle of reciprocal influence between mind and brain is mirrored in the relationship between agent and environment. The agent encodes a model of its environment which sets the frame for experience and choice, but the agent also acts on and remodel that environment; as we interact in the social, material, and technological domains, we actively shape the world that shapes us. Deliberate, top-level processes have a profound influence on the quality of our experience, both in the most immediate sense, and as our choices play out over time. As we select what we attend to and make decisions about how to understand and engage with the world, we transform ourselves as agents, and restructure the context of future experience and choice.

We continually create information that is carried forward into future experiences, choices, and actions. Thought patterns tend to repeat, we replicate our past responses, and weave past experience into the structure of current experience. We are free to exploit this process by creating conditions in the present that enhance future experience, and preserve or expand our capacity for freedom in the future. We can broaden the scope of behavioural flexibility by building our knowledge, exposing ourselves to new experiences, developing our mental skills or expanding our behavioural repertoire. We can constrain our action selections to those actions that align with our highest values and overall flourishing, and in doing so, increase the likelihood that the choice space we encounter in the future is contextualised by these chosen higher-order predictions. We can cultivate attentional control and engage in the systematic transformation of our minds in ways that enhance our agency and autonomy. Alternatively, we can passively subject our minds to agency-eroding information, environments, and technologies, and allow our window of freedom to be diminished by distraction, distortion, and manipulation. Each deliberated choice and action contributes to the ongoing process of narrowing or expanding the parameters of our own freedom, and takes us incrementally towards deeper entanglement, or towards self-mastery and genuine autonomy.

References

Allen, M., & Tsakiri, M. (2019). The body as first prior: Interoceptive predictive processing and the primacy of self-models. In M. Tsakiris & H. De Preester (Eds.), *The interoceptive mind: From homeostasis to awareness* (pp. 27–45). Oxford University Press.

Albert, D. Z. (2000). *Time and chance*. Harvard University Press. https://doi.org/10.4159/9780674020139

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017a). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160011. https://doi.org/10.1098/rstb.2016.0011

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

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: Spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2010.2325

Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics. *Pharmacological Reviews*, *71*(3), 316–344. https://doi.org/10.1124/pr.118.017160

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A. (2017). Predictions, precision, and agentive attention. *Consciousness and Cognition*, *56*, 115–119. https://doi.org/10.1016/j.concog.2017.06.013

Clark, A. (2020). *Beyond desire? Agency, choice, and the predictive mind*. https://doi.org/10.1080/00048402.2019.1602661

Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, *35*(3), 32. https://doi.org/10.1007/s10539-020-09746-2

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. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface*, *10*(86), 20130475. https://doi.org/10.1098/rsif.2013.0475

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130481. https://doi.org/10.1098/rstb.2013.0481

Gotink, R. A., Vernooij, M. W., Ikram, M. A., Niessen, W. J., Krestin, G. P., Hofman, A., Tiemeier, H., & Hunink, M. G. M. (2018). Meditation and yoga practice are associated with smaller right amygdala volume: The Rotterdam study. *Brain Imaging and Behavior*, *12*(6), 1631–1639. https://doi.org/10.1007/s11682-018-9826-z

Heisenberg, M. (2009). Is free will an illusion? *Nature*, *459*(7244), Article 7244. https://doi.org/10.1038/459164a

Hoefer, C. (2002). Freedom from the inside out. *Royal Institute of Philosophy Supplement*, *50*, 201–222. https://doi.org/10.1017/S1358246100010572

Hoefer, C. (2023). Causal determinism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/determinism-causal/

Hohwy, J. (2007). The Sense of Self in the Phenomenology of Agency and Perception. *Psyche*, *13*(1).

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

Hohwy, J., & Michael, J. (2017). Why should any body have a self. In F. de Vignemont & A. J. T. Alsmith (Eds.), *The Subject's Matter: Self-Consciousness and the Body*. MIT Press. https://doi-org.ezproxy.waikato.ac.nz/10.7551/mitpress/10462.003.0020

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209–223. https://doi.org/10.1111/mila.12281

Hossenfelder, S. (2022). *Existential physics: A scientist's guide to life's biggest questions*. Atlantic Books.

Ismael, J. T. (2014). On Being Someone. In A. R. Mele (Ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience* (pp. 274–297). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199333950.003.0014

Ismael, J. T. (2016). *How Physics Makes Us Free*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190269449.001.0001

Ismael, J. T. (2021, October 1). *Causation from the Point of View of Physics*. The Franke Program in Science and the Humanities Understanding the Nature of Inference Colloquium Series, Yale University. https://www.youtube.com/watch?v=FgwiXUB5MDQ

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, *15*(138), 20170792. https://doi.org/10.1098/rsif.2017.0792

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Mamat, Z., & Anderson, M. C. (2023). Improving mental health by training the suppression of unwanted thoughts. *Science Advances*, *9*(38). https://doi.org/10.1126/sciadv.adh5292

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Mele, A. (2013, February 1). *Free Will*. Wiley Online Library; John Wiley & Sons, Ltd. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444367072.wbiee159

Nave, K., Deane, G., Miller, M., & Clark, A. (2022). Expecting some action: Predictive processing and the construction of conscious experience. *Review of Philosophy and Psychology*, *13*(4), 1019–1037. https://doi.org/10.1007/s13164-022-00644-y

O'Connor, T., & Franklin, C. (2022). Free will. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/freewill/

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. The MIT Press.

Pezzulo, G. (2012). An active inference view of cognitive control. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00478

Russell, B. (1912). On the notion of cause. *Proceedings of the Aristotelian Society*, *13*, 1–26.

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, *75*(1), 3–13. https://doi.org/10.1111/pcn.13138

Strawson, G. (2010, July 22). Your move: The maze of free will. *New York Times*. https://archive.nytimes.com/opinionator.blogs.nytimes.com/2010/07/22/your-move-the-maze-of-free-will/

Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnsys.2016.00079

Touroutoglou, A., Andreano, J., Dickerson, B. C., & Barrett, L. F. (2020). The tenacious brain: How the anterior mid-cingulate contributes to achieving goals. *Cortex*, *123*, 12–29. https://doi.org/10.1016/j.cortex.2019.09.011

Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Wiese, W., & Metzinger, T. K. (2017). Vanilla PP for philosophers: A primer on predictive processing. In T. K. Metzinger & W. Wiese (Eds). Philosophy and Predictive Processing. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573024

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.