

Interpretation and Its Overcoming:
Nietzsche, Active Inference, and the Organization of Experience

Flyxion
Independent Researcher

October 2026

Preface

This book compares two bodies of work that rarely meet. One is Nietzsche's treatment of interpretation as a condition of organized experience that life must nonetheless repeatedly overcome. The other is the family of theories, grouped under predictive processing and active inference, that explain perception, action and learning as the activity of a system maintaining an organized relation to its surroundings. The comparison is conducted through shared problems rather than claims of anticipation, and it keeps four kinds of claim apart: historical claims about what Nietzsche wrote, conceptual comparisons between his problems and those of contemporary modelling, mathematical reconstructions that are offered as proposals, and normative arguments about which transformations of interpretation are worth having.

The mathematical sections are marked as reconstruction and are never presented as translation. Literary analysis is treated as evidence about what Nietzsche's writing does to a reader's expectations, and not as an illustration of a theory established elsewhere.

Contents

Introduction: The Organism That Interprets	1
I Interpretation Before the Sovereign Subject	5
1 Reading <i>The Will to Power</i> as a Problematic Archive	7
2 Knowledge as Schematization	9
3 The Ego as an Interpretation of Activity	11
4 Drives, Perspectives, and Competing Interpretations	13
II Predictive Organization and Interested Action	15
5 Predictive Coding, Active Inference, and the Limits of Analogy	17
6 Action as the Transformation of Evidence	21
7 Precision, Attention, and the Authority of an Interpretation	25
8 Exploration, Resistance, and the Growth of Capacity	27
III Self-Overcoming and the Revision of Models	29
9 Learning Within a World and Changing the Terms of a World	31
10 Defensive Coherence and Interpretive Entrapment	35
11 Attachment and the Conditions of Revisability	37
12 Value Creation and the Problem of the Objective	41

IV	Literary Form as an Experiment in Interpretation	43
13	Aphorism, Fragment, and the Interrupted Reader	45
14	Zarathustra's Voice and the Hazards of Discipleship	49
15	Metaphor, Concept Formation, and the Invention of Distinctions	53
V	Beyond Individual Interpretation	55
16	Shared Models, Institutions, and Collective Self-Confirmation	57
17	Power, Truth, and the Availability of Correction	59
18	The Limits of the Synthesis	61
A	Notation and Conventions	63
B	Variational Free Energy and Expected Free Energy	65
C	Coarse-Graining and Task-Relative Sufficiency	67
D	An Example of Structural Revision	69
E	A Small Dynamical Model of Entrapment	71

Introduction: The Organism That Interprets

An organism that acts must treat some differences as mattering and others as negligible. It must carry expectations about what follows from what, and it must do so before the evidence is complete. Interpretation in this sense is a condition of organized experience, since without a selective ordering of the field there would be no stable objects, no recognizable situations and no actions that could be directed at one outcome rather than another. Yet an ordering that makes life manageable also fixes what can be noticed, and an organization adequate to one range of circumstances can become an impediment when circumstances change or when the organism's own capacities outgrow it. The central problem of this book is how interpretation can be both a condition of organized experience and something that life must repeatedly overcome.

Nietzsche approaches the problem through accounts of schematization, of perspectival knowing and of the multiplicity of drives that coordinate or contest one another beneath the appearance of a unified subject. Predictive processing and active inference approach a related problem from a different direction. They explain perception as the revision of expectations in light of discrepancy, explain action as a means of bringing about the sensory conditions that the organism's model anticipates, and explain learning as the adjustment of the model that generates those expectations (Rao and Ballard, 1999; Friston, 2010; Clark, 2016; Parr et al., 2022). Both bodies of work begin from the embodied activity of a living system and decline to treat cognition as the inspection of a world by a detached observer. That shared starting point is the ground of the comparison. It does not license the claim that either anticipated the other, and the comparison proceeds through shared problems: how simplification is possible without being mere distortion, how a self is constituted as an interpretation of activity, how competing concerns acquire a working hierarchy, and how an organization can be revised without being destroyed.

The governing thesis is stated here provisionally and defended across the chapters. Nietzsche and active inference both make cognition intelligible through the organization of an organism's engagement with its surroundings. Nietzsche additionally asks how that organization acquires authority, and when its preservation becomes an impediment to transformation. Active inference supplies exact vocabulary for the preservation of an organization, namely the maintenance of expected states under a generative model. It is less clear that it supplies an account of the difference between preserving an interpretation that makes life manageable and transforming an interpretation so that new forms of life become possible. Whether that difference can be captured within the formalism, or whether it requires an additional account of value, is the disagreement around which the book is organized.

Four kinds of claim are distinguished throughout, so that resemblance at one level never silently establishes equivalence at another. Historical claims concern what Nietzsche wrote, in which texts, and with what changes of emphasis. Conceptual comparisons concern whether a problem he poses has a counterpart in contemporary modelling and whether the two treatments are compatible, rival or merely parallel. Mathematical reconstructions are proposals, in which a Nietzschean notion is given a precise form that makes its commitments inspectable, and each is accompanied by a statement of what it leaves out. Normative arguments concern which transformations of interpretation deserve affirmation, and these are argued for on their own terms rather than read off either source.

The method has a literary dimension that is more than ornament. Nietzsche's writing is aphoristic, interrupted and addressed to a reader whose expectations it deliberately disturbs. What such writing does to a reader is evidence about the activity of interpretation, and the book attends to it as such, asking how transitions, changes of voice and unresolved ambiguities alter which features of a text a reader treats as authoritative. A generic account of prediction and error could describe some of this, but it could also flatten irony, allusion and productive ambiguity into a single quantity, and the literary chapters are written to keep that risk visible.

The book proceeds in five parts. The first concerns interpretation before the sovereign subject, taking up the textual method, the treatment of knowledge as schematization, the ego as an interpretation of activity, and the drives as competing perspectives. The second supplies the technical foundation of predictive organization and examines action, precision and exploration. The third addresses self-overcoming and the revision of models, including the problem of value creation.

The fourth studies literary form as an experiment in interpretation, and the fifth extends the argument to shared models and institutions and marks the limits of the synthesis.

Part I

Interpretation Before the Sovereign Subject

Chapter 1

Reading *The Will to Power* as a Problematic Archive

The textual basis of any account of Nietzschean interpretation must first settle what is being read. *The Will to Power* appeared as a book under Nietzsche's name only after his collapse, assembled by his associates and his sister from notebook material of the 1880s, and the numbered aphorisms of the widely used English edition (Nietzsche, 1967a) follow an arrangement that Nietzsche did not give them. The critical edition prepared by Colli and Montinari restored the notebooks to their chronological order and made it possible to see that the plan of a systematic work under this title was entertained, revised and eventually abandoned in 1888, when Nietzsche turned instead to the writings he completed (Montinari, 1982). A passage quoted from the compilation therefore comes with two kinds of uncertainty. It may record a thought that was later developed, tested against objections and published in a finished work, or it may record a trial formulation that Nietzsche dropped.

The method adopted here follows from that situation. Passages from the notebooks are used to reconstruct a constellation of problems and are never treated as articles of a finished doctrine. Where a claim about interpretation appears in the compilation, it is located in its notebook context where that can be done, and compared with its treatment in *The Gay Science*, *Beyond Good and Evil*, *On the Genealogy of Morality* and *Thus Spoke Zarathustra* (Nietzsche, 1974, 1966, 1967b, 2006). Agreement between the notebooks and a published work strengthens an attribution. Disagreement is recorded as a tension or as a change of emphasis, and is not smoothed away to produce a unified position. Systematic readings of Nietzsche exist and are argued with care (Richardson, 1996), and the present method

does not exclude them. It requires only that any systematic reading say which of its premises rest on published statements and which on material Nietzsche chose not to publish.

The comparison can be made concrete for the doctrine of the next chapter. The notebook claim that knowing is a schematizing that imposes regularity as far as practical needs require (section 515 of the compilation) has published counterparts of different weight. Section 24 of *Beyond Good and Evil* describes cognition as a will to simplification that has contrived to preserve ignorance in the service of life, and section 354 of *The Gay Science* derives conscious thought from the pressure of communication, so that what reaches awareness is already shaped by what can be shared. Both passages support the notebook's picture of cognition as practical simplification, and neither states it as a general theory of knowledge in the compilation's terms. The attribution adopted below is accordingly that Nietzsche held the schematizing thesis as a recurring problem and not that he fixed it as a doctrine, and the dating and notebook identifiers of the relevant entries in the critical edition remain to be supplied wherever a claim rests on the compilation alone.

A literary question enters at the outset and recurs throughout the book. Aphoristic and experimental writing, when reorganized into the form of a philosophical system, acquires expectations of continuity, hierarchy and doctrinal consistency that the original writing did not claim to satisfy. A reader who meets numbered sections arranged under headings on knowledge, nature, society and art will tend to assume that earlier sections ground later ones and that apparent contradictions are failures of exposition. The editorial arrangement thus supplies an interpretive frame before any sentence is read. This is itself an instance of the phenomenon under study, since an organization imposed on material determines which relations among its parts are visible, which are salient and which are treated as noise. The material form of Nietzsche's reception is accordingly part of the argument about how interpretation organizes its objects, and the reading practice adopted in later chapters aims to resist the frame without pretending to a neutral standpoint outside any frame.

Chapter 2

Knowledge as Schematization

Nietzsche's treatments of identity, regularity and the concept share a recurring claim: that the cognition which makes a stable world available does so by treating unlike things as like and by imposing regularity on a flux that does not present itself in that form. The notebooks put this in terms of schematization, an activity that, as section 515 of the compilation has it, imposes upon chaos as much regularity and form as practical needs require. The question this chapter pursues is whether such activity should be described as the discovery of an order or as the making of experience tractable through an ordering. The two descriptions differ in what they permit one to say about error. If cognition discovers order, a schema is false to the degree that it fails to match what is there. If cognition makes experience tractable, a schema may be adequate or inadequate relative to a task, a scale and a form of life, and the word error needs a different home.

The question of error is sharper than it first appears. A useful simplification does not become false merely because it suppresses differences, since every concept suppresses some. What matters is whether the suppressed differences are idle with respect to the purposes the concept serves. The same simplification can be adequate for one purpose and ruinous for another, and it can pass from the first condition to the second when circumstances change. A reading of Nietzsche on which all schematization is equally falsifying cannot make this distinction, and a reading on which it is simply true cannot explain why he treats the same schemata as both indispensable and as objects of suspicion.

Coarse-graining supplies a precise vocabulary for the point. Let X be a detailed state space and let an interpretive organization be a map

$$\phi : X \longrightarrow Z,$$

where Z contains the distinctions available to the organism. Because distinct states of X may be sent to the same state of Z , the organization necessarily discards differences. The relevant question is which discarded differences remain irrelevant to action and which later become consequential. This can be made exact relative to a task.

Definition 2.1 (Task-relative sufficiency). Let T be a task specified by a class of admissible policies and a loss ℓ defined on X . An organization $\phi : X \rightarrow Z$ is ε -sufficient for T if the best expected loss attainable by policies that depend on X only through ϕ exceeds the best expected loss attainable by policies that depend on X directly by no more than ε .

The definition gives sense to the claim that an organization is adequate without being a copy. It also makes adequacy fragile in the way the notebooks suggest. If the task changes, or if the distribution of encountered states shifts, a map that was ε -sufficient for the earlier task and distribution may cease to be so, although nothing about the map itself has altered. A simplification can thus pass from harmless to harmful through a change in what the organism does or where it lives. This gives a precise reading of the claim that errors of cognition are conditions of life: the simplification is a condition of life for the tasks it serves, and an error for those it does not.

Remark 2.1. The definition does not say that all schematization is innocent. It says that the evaluation of a schema requires a specification of task and environment, and that the specification can itself be contested. A reconstruction of Nietzsche's view that left this contestation out would explain only the half of his position on which schemata are useful.

Chapter 3

The Ego as an Interpretation of Activity

Nietzsche's criticism of a substantial subject, a simple inner proprietor who thinks, wills and owns its acts, is often taken to imply a denial of agency. The two should be kept apart. The disappearance of a simple proprietor leaves the fact of organized action in place, together with the need to explain how a plurality of processes acquires enough coordination to act, to remember, to promise and to experience itself as continuous. The explanatory task is thereby changed rather than abolished. Instead of asking what the self is, one asks how a self-describing coordination is produced and maintained, and what its description leaves out. Nietzsche's own positive proposal, in *Beyond Good and Evil*, section 12, is to understand the soul as a multiplicity, a social structure of drives and affects, and section 19 analyses willing as a relation of commanding and obeying within the one who wills. The feeling of being a unified agent can conceal the complex activity that produces it, since the one who commands identifies with the commanding side and takes credit for what the whole organization accomplishes. The ego is accordingly partly an achievement of organization and partly a misleading interpretation of that achievement.

The comparison with contemporary modelling concerns the distinction between a process and its representation of its own activity. A system may maintain a model that summarizes its bodily states, its action possibilities and the expected consequences of acting, without that model transparently disclosing the mechanisms that generate the behaviour it summarizes. Such a self-model is useful in the way schemata are useful: it compresses a large and partly inaccessible set of processes into a form that can guide planning and be communicated to others.

It can also mislead in a systematic way. A model that presents the system as a unified deliberator will explain actions by appeal to decisions and will render the contest of concerns that produced them invisible, which is the pattern that Nietzsche diagnoses in the ordinary grammar of willing.

The easy conclusion, that an interpreted self is therefore unreal, does not follow and should be resisted. An interpretation can have causal consequences. It can regulate conduct, structure what the organism attends to, shape what others expect of it, and become an object of investment, so that challenges to it are felt as challenges to the organism and not merely to a proposition. These effects obtain even when the interpretation's account of its own origins is mistaken. The self so understood is neither a substance nor an illusion in the sense of a mere appearance without effect. It is an organization that has consequences, and the questions it raises are about its authority, its revisability and the cost of revising it, which later chapters take up.

The distinction between an agent and the model an agent has of itself also bears on the relation between responsibility and metaphysics. Promising and being held to a promise require that conduct be coordinated across time, and a self-model that supports such coordination is practically necessary whatever its metaphysical standing. The Nietzschean question is therefore what kind of coordination a given self-interpretation makes possible and what kinds it forecloses, and this question can be asked without settling whether any inner proprietor exists.

Chapter 4

Drives, Perspectives, and Competing Interpretations

The term drive is used by Nietzsche in ways that resist a single analysis. Sometimes it names an impulse that presses for discharge, sometimes an evaluative orientation that gives things their significance, sometimes a competing organization that seeks to interpret the whole in its own terms. These senses are not obviously consistent, and the comparative problem is sharpened by a difference of method. Many active-inference models begin with a specified agent and a specified objective, and ask how the agent should perceive and act. Nietzsche frequently makes the organization of the agent itself a site of conflict, so that the question of whose objective governs is open and its answer can change.

A reconstruction can represent partially competing evaluative processes through objectives F_i , each of which assesses states and policies from its own standpoint, and an effective organization such as

$$F_{\text{eff}}(s) = \sum_i w_i(s) F_i(s).$$

Proposal 4.1. Treat a Nietzschean multiplicity as a family of objectives $\{F_i\}$ together with state-dependent weights $w_i(s)$ that determine which concerns govern conduct in which situations.

What matters philosophically is the weights and not the sum. The sum represents a momentary settlement, whereas the weights encode a hierarchy, and the question that Nietzsche presses is what determines which concerns dominate and how that determination can change. If the weights are fixed, the multiplicity collapses into a single objective with a complicated shape, and nothing distinctively

Nietzschean remains. If the weights depend on state, they can shift as circumstances shift, which captures the idea that a drive may dominate in one setting and be subordinate in another. If they depend on the history of the organism, a further structure is needed to describe how hierarchies are acquired, and this is where an account of formation and revision becomes necessary.

It is important to say what this proposal does not claim. It does not translate drives into computational modules, and it does not suggest that Nietzsche had a weighted sum in mind. It offers a form that makes certain commitments inspectable, and one of its uses is to reveal where the form fails. When the objectives F_i are not commensurable, so that no common scale licenses adding them, the scalar F_{eff} is an artefact of the reconstruction and not a feature of the phenomenon. In such cases the conflict is better represented as an ordering that is partial, or as a contest whose resolution is a matter of the organization taking one form rather than another. Part of the later discussion of value creation turns on whether a framework that begins from a scalar objective can represent such conflicts without silently reintroducing an evaluator that stands above them.

Part II

Predictive Organization and Interested Action

Chapter 5

Predictive Coding, Active Inference, and the Limits of Analogy

The comparison pursued in this book depends on a technical vocabulary whose terms are often run together, and the first task is to keep them apart. Predictive coding names a family of schemes in which a processing hierarchy transmits predictions downward and discrepancies upward, so that what propagates through the system is chiefly the part of the input that the current prediction fails to account for (Rao and Ballard, 1999). Bayesian inference names a normative relation between prior beliefs, a likelihood and a posterior, and says nothing about how a physical system might realize it. Variational inference is an approximation technique for cases in which the exact posterior is intractable, and it proceeds by choosing a tractable family of distributions and selecting the member closest to the posterior. Active inference extends this apparatus by treating action and policy selection as further consequences of the same optimization, so that an agent is described as selecting the courses of action under which its expected observations accord with a generative model that includes preferred outcomes (Friston et al., 2017; Parr et al., 2022). The free-energy principle is the broadest of these claims, a proposal about the conditions under which any self-organizing system that persists can be described as if it were performing such inference (Friston, 2010). Each step in this sequence adds assumptions, and a conclusion licensed at one step is not licensed at the next. These terms are kept separate throughout, and a comparison with Nietzsche that depends on the stronger claims will say so.

The central mathematical object is the variational free energy. For a generative

model $p(o, s)$ over observations o and hidden states s , and an approximate posterior $q(s)$, it is defined by

$$F[q; o] = \mathbb{E}_{q(s)}[\log q(s) - \log p(o, s)] = D_{\text{KL}}(q(s) \parallel p(s \mid o)) - \log p(o).$$

The second equality follows from the factorization $p(o, s) = p(s \mid o)p(o)$, which allows the logarithm of the joint to be split into a term that depends on s and a term that does not. Two consequences are immediate. Because the divergence is non-negative, F is an upper bound on the surprise $-\log p(o)$, with equality exactly when q equals the exact posterior. And because $\log p(o)$ does not depend on q , minimizing F over q is the same as minimizing the divergence between q and the posterior, so that perception, understood as the choice of q , becomes approximate Bayesian inference.

It is equally important to say what does not follow. Minimizing F under a given generative model yields a posterior that is best relative to that model. It carries no guarantee that the model is adequate, that its state space contains the distinctions the situation requires, or that its priors deserve the confidence they receive. The bound on surprise is a bound on surprise as assessed by the model, so that an agent whose model is poor can drive F low by attending to what the model handles well. Approximate inference under a generative model does not by itself establish a philosophical perspectivism, an ontology of power or a criterion of flourishing. Where active inference is invoked in later chapters, it is invoked for what the formalism actually delivers, and the gap between that and the Nietzschean questions about authority and transformation is itself part of the subject.

Action enters through the expected free energy of a policy π . In one common formulation,

$$G(\pi) = D_{\text{KL}}(q(o \mid \pi) \parallel p(o)) + \mathbb{E}_{q(s \mid \pi)}[\mathbb{H}[p(o \mid s)]] ,$$

where the first term measures the divergence between predicted and preferred outcomes, often called risk, and the second measures the expected ambiguity of the mapping from states to observations. Policies with low G are those expected to bring about preferred outcomes through observations that are informative about hidden states, and the decomposition of Appendix B shows that this informativeness is the expected information gain. The preferred outcomes are encoded in $p(o)$, which is part of the model and not derived from the formalism. This observation

is of first importance for what follows, since the authority of those preferences is exactly what the Nietzschean question about value presses on, and Chapter 12 returns to it.

Chapter 6

Action as the Transformation of Evidence

An organism does not receive evidence passively. What it encounters depends on where it goes, what it handles, whom it asks and what it ignores, and each of these is in turn shaped by the interpretation it already holds. Perception and action are therefore mutually implicated in a strong sense. Action determines which observations occur, which uncertainties can be resolved and which expectations meet repeated confirmation, and the resulting observations determine how the interpretation is revised. In the formal setting this appears as the dependence of the observation distribution on the policy, written $q(o \mid \pi)$, so that the evidence available for assessing a model is partly a product of the model's own use.

The Nietzschean comparison becomes especially productive at this point. An interpretation can organize practices that make its world increasingly familiar and manageable. A way of reading situations leads to a way of acting in them, the acting produces situations of the kind the reading handles well, and the reading appears to be confirmed. This is the problem of endogenous evidence: apparent confirmation may result partly from restricting encounters to situations that an existing model already handles well. A record of success is evidence about an interpretation only to the extent that the encounters composing it were not selected by the interpretation to ensure that success.

Three kinds of action should be distinguished, because they stand in different relations to the evidential standing of the interpretation they serve. Successful intervention is action that changes a situation in the way the interpretation predicts, and where the situation was not chosen to guarantee that result, it counts as a genuine test that the interpretation passed. Selective exposure is the arrangement

of one's circumstances so that only situations of a manageable kind arise, and the confirmations it produces are largely manufactured. Exploratory action seeks situations in which the interpretation's predictions are uncertain or likely to fail, and its results bear on the interpretation in proportion to that risk. The same observable pattern, a long run of confirmed predictions, can arise from any of the three, and nothing in the record of confirmations alone distinguishes them.

The distinction can be given a simple form. Let μ_π be the distribution over situations that an agent following policy π actually encounters, and let ν be a reference distribution over the situations that matter for the agent's continued viability, which may include situations it rarely meets. Write $\ell(x)$ for the model's expected prediction error in situation x .

Definition 6.1 (Insulation gap). The *insulation gap* of a policy π for a model with error profile ℓ is

$$\Delta(\pi) = \mathbb{E}_{x \sim \nu}[\ell(x)] - \mathbb{E}_{x \sim \mu_\pi}[\ell(x)].$$

A large positive gap indicates that the model performs far better on the situations its policy leads it to meet than on those that matter for its viability. This is the signature of selective exposure, but it is equally the signature of effective specialization or of successful environmental management, in which an agent legitimately transforms its surroundings so that the model's competence suffices. The gap measures a disparity in performance relative to a specified reference distribution and does not identify its cause, which requires further evidence about how the encountered situations came to be selected. A gap near zero shows only that the average error is the same under the two distributions. It does not show that the distributions coincide, since very different distributions can yield equal expected values of a single loss. The reference distribution ν cannot be read off the formalism, and its specification is a substantive and contestable choice. This is the point at which the formal reconstruction acknowledges its dependence on something outside itself, since which situations matter for the agent's continuation is a question about the agent's form of life, and Chapter 17 argues that it must be answered by an account of correction that the formalism does not by itself supply.

The notion also clarifies why defensive coherence is not obviously a failure from within. An agent that has arranged its circumstances so that its interpretation is never seriously challenged will report low error, high confidence and a sense that its view of the world is well supported. Only a comparison with situations it does not meet reveals the gap, and an agent organized to avoid such situations has little

occasion to make the comparison.

Chapter 7

Precision, Attention, and the Authority of an Interpretation

In predictive models the influence of a discrepancy depends on how it is weighted. A large error that is assigned low precision barely alters the system's estimates, whereas a small error assigned high precision can drive large revisions. Precision is thereby a mechanism of attention and of authority: it determines which discrepancies are heard. In a Nietzschean register the corresponding point is that an encounter does not arrive with an unquestionable authority independent of the organization receiving it. Whether a surprising event counts as a refutation, an anomaly, a nuisance or an opportunity depends on how the interpretation that meets it is disposed to weigh it. Some discrepancies become transformative while others are dismissed, subordinated or redescribed, and the difference lies largely in the weighting.

A local quadratic approximation makes the structure explicit. For a prediction error ε and a precision matrix Π , the weighted discrepancy is

$$\mathcal{E} = \frac{1}{2} \varepsilon^\top \Pi \varepsilon.$$

In a Gaussian model with observation noise precision Π , the negative log-likelihood of an observation has the form

$$-\log p(o \mid s) = \frac{1}{2} \varepsilon^\top \Pi \varepsilon - \frac{1}{2} \log |\Pi| + \text{const},$$

where ε is the difference between the observation and its prediction. The second term deserves attention. It shows that discounting an error by lowering precision is not free within the formalism, since the log-determinant term penalizes low pre-

cision and so charges for any claim that the observations are generally unreliable. A model that could silence every discrepancy by setting Π close to zero would be penalized for implying that its own observations carry almost no information. The formalism thus builds in a limited resistance to wholesale insulation, although it cannot prevent a model from raising precision on confirming channels and lowering it on disconfirming ones at a modest total cost.

The philosophical question therefore concerns the formation and revision of Π more than the elimination of ε . An account that treats precision as a given parameter to be estimated leaves open who or what assigns authority to some channels of evidence over others, and how that assignment itself becomes open to challenge. If precision is learned from the history of past reliability, then an agent whose history was shaped by selective exposure will have learned precisions that reflect its insulation, and the learning process reproduces rather than corrects the bias of Chapter 6. Revising precision in the face of a discrepancy that the current precision already discounts is thus a distinctive difficulty, and it is one that the literary chapters later examine as a problem of how a text can alter what a reader takes to be authoritative.

This connects the technical notion to reading. Transitions, interruptions, changes of voice and shifts in affect can be studied as ways of altering which aspects of a text a reader treats as authoritative, which is to say as interventions on the reader's precisions. A sudden irony does not supply a new proposition. It changes the weight a reader gives to what came before, and reading afterwards proceeds under a different distribution of attention. The mathematical vocabulary supplies a way of stating this effect, though it does not supply the reasons that make one such shift illuminating and another merely disorienting.

Chapter 8

Exploration, Resistance, and the Growth of Capacity

An apparent tension runs between the idea that organisms minimize free energy and the observation that they seek difficulty, novelty and resistance. If the aim were simply to reduce surprise, an organism would retreat to the most predictable environment available, and the Nietzschean praise of resistance would seem opposed to the formalism. The tension is reduced by the expected free energy introduced in Chapter 5. The decomposition of $G(\pi)$ derived in Appendix B contains the negative expected information gain, so that policies expected to resolve uncertainty about hidden states are rewarded, and exploration is not an exception to the optimization but a component of it. A policy that visits a situation of high uncertainty can have low expected free energy because of what it is expected to teach. Information seeking and the pursuit of unfamiliar situations are therefore accommodated, and the quantity that measures their value is the expected information gain

$$\text{IG}(\pi) = \mathbb{E}_{q(o|\pi)}[D_{\text{KL}}(q(s | o, \pi) \| q(s | \pi))],$$

the expected reduction in uncertainty about hidden states from the observations that policy π would produce.

Whether this suffices to reconstruct Nietzschean overcoming is a further question, and the answer depends on a distinction between two kinds of uncertainty. The first is uncertainty resolved within an existing repertoire: the agent has a model with certain variables, does not know their current values, and acts to find out. The second is the encounter that requires a new repertoire, in which the difficulty lies in the fact that the agent's variables omit something essential. Learning

a previously unknown value of a familiar variable differs in kind from learning that one's variables were the wrong ones.

The point can be stated sharply. The information gain above is defined over the hidden states s of a fixed model. It is bounded by the entropy of the agent's current beliefs about those states, and it is silent about any feature of the world that the state space does not represent. An unmodelled variable contributes nothing to IG however much it matters, because the quantity has no argument in which to register it. An agent that explores optimally in the sense of maximizing expected information gain is therefore exploring within a fixed space of distinctions, and its exploration can be exhaustive and still leave the space unchanged.

Nietzsche's interest in resistance can then be investigated as a problem of expanding or reorganizing capacities, and without identifying power with any single information-theoretic quantity. Resistance is valuable on this reading not merely because it supplies information about the familiar variables, though it may, but because it can disclose that the agent's present repertoire cannot meet it, and so create the occasion for a different repertoire. What grows in such an episode is the range of distinctions the agent can draw and the range of situations it can handle, and this is a change in the model family rather than in the estimates made within it. The next part takes up that change directly. Here it is enough to record that active inference handles seeking difficulty as a matter of reducing uncertainty about a given world, and that the Nietzschean demand extends further, to the possibility of a world whose terms differ from those currently available.

A caution against the opposite error belongs here. It would be a mistake to equate overcoming with the maximization of any measure of novelty or difficulty. An agent can pursue difficulty indiscriminately and accomplish nothing, and resistance that exceeds what an organization can absorb destroys it. The growth of capacity is therefore a matter of resistance met at a rate and in a form that the organization can convert into reorganization, and which encounters meet that condition is a question about the particular organism and its supports, which later chapters on attachment and institutions address.

Part III

Self-Overcoming and the Revision of Models

Chapter 9

Learning Within a World and Changing the Terms of a World

Three activities are easily confused under the single name of learning, and the mathematical centre of this book depends on keeping them apart. State estimation is the inference of the current hidden state of the world from observations, given a model. Parameter learning is the adjustment of the quantities that fix how states generate observations and how they evolve, again within a model whose form is held fixed. Structural revision is the alteration of the form itself, so that the variables, the dependencies among them or the kinds of cause admitted are different afterwards. The first two operate inside what may be called an admissible family of models, and the third replaces the family. An agent may improve its estimates indefinitely while retaining an inadequate family, and excellent optimization within that family cannot supply distinctions the family excludes.

Let $\mathcal{M}$ denote the admissible family. Ordinary updating selects within it, producing at each time a posterior over states and parameters that depends on the data seen so far. Structural revision is a change of the family,

$$\mathcal{M}_t \longrightarrow \mathcal{M}_{t+1},$$

and the question the chapter pursues is what could make such a transition justified. Two kinds of evidence bear on it. The first is a persistent residual. Suppose that over a long run of data the best achievable predictive error within $\mathcal{M}_t$, after all parameter learning, remains above what the noise level of the observations would explain. The excess is then not attributable to ignorance of the current state or to poorly tuned parameters, since both have been optimized away, and it indicates that the family omits some structure the data contain. The second kind of evidence

is comparative. If a candidate family $\mathcal{M}'$ is available, its marginal likelihood on the same data can be compared with that of $\mathcal{M}_t$, and the ordinary machinery of model comparison applies.

The comparative route presupposes what the question was meant to explain. A candidate family must already be specified for its evidence to be computed, so that Bayesian model comparison ranks proposals but does not generate them. The residual diagnoses a lack without saying what would fill it. This yields a precise statement of a difficulty that Nietzschean language tends to leave in metaphor, though a narrower one than a general impossibility claim would suggest. Optimization within a specified family does not, by itself, explain how an agent acquires alternatives outside that family. Formal systems can be expanded to include structural alternatives, procedures that propose them, or searches over descriptions, as work on structure learning, nonparametric Bayesian models and program induction does, and such systems do generate new families in a precise sense. The question then concerns what the expanded system presupposes, namely which space of structures is searched, by what proposal mechanism and under what prior, and how its proposals become available to the organism. Relative to a given family, structural novelty is a reorganization of the space in which inference takes place, and an account of the organism's situation, whether encounter with resistance, metaphor, other persons or the inheritance of a language, must say what makes the enlarged space and its proposals available.

Cost enters the decision. Revising a family is not free, since it disrupts the habits, expectations and coordinated action built on the old one. A simple expression of the trade-off is that revision is warranted when the expected improvement in performance over the relevant horizon exceeds the cost of transition,

$$\mathbb{E}[\Delta L] H > c,$$

where ΔL is the per-period reduction in expected loss, H the horizon and c the transition cost. This inequality is deliberately crude, and its crudeness is instructive. The terms ΔL and H are themselves estimated from within the current interpretation, so that an organization may undervalue a change whose benefits its own categories cannot register. Cost is also unevenly distributed across agents, and Chapter 11 asks what happens to c when the burden of revision can be shared.

Self-overcoming enters as a philosophical interpretation of certain such reorganizations and not as a synonym for them. The computational notion is value-neutral, since a family can change for better or worse and the criteria of residual

and evidence are purely predictive. Nietzsche's evaluative conception of becoming asks which reorganizations express an enhancement of life and which a decline, and this requires a standard that the predictive criteria do not contain. The text preserves the difference between computational revision and evaluative becoming throughout, and Chapter 12 returns to what the standard would have to be.

Chapter 10

Defensive Coherence and Interpretive Entrapment

An interpretation can maintain its coherence by reducing its exposure to correction. The mechanisms are familiar from the preceding chapters: selective attention that assigns low precision to disconfirming channels, constrained action that avoids situations where the interpretation is at risk, reinterpretation of counterexamples so that they confirm what they appeared to challenge, and dependence on social environments that reward repetition of the received view. A system organized in this way can become increasingly effective at preserving its interpretation while becoming less capable of handling changes beyond the protected domain. Its reported coherence improves as its adequacy to the wider situation declines.

Nietzsche's analysis of the ascetic ideal supplies an unusually rich case. In the third essay of *On the Genealogy of Morality*, especially in sections 7 and 8, the philosopher's withdrawal from ordinary pleasures and obligations is presented as serving the conditions of intellectual work, so that renunciation can express a powerful commitment to a task. The broader ascetic ideal, however, can turn suffering into an indictment of earthly existence and make detachment a doctrine about life as such. Outwardly similar acts of detachment can accordingly serve very different dispositions: one a discipline that enables activity, the other a defence that forecloses it. The concluding section of the essay presses the point further by observing that human beings would sooner will nothingness than not will. An apparent escape from attachment can preserve an attachment to purity, to salvation or to the identity of one who has transcended ordinary needs. The will has changed its object, and perhaps turned against itself, without ceasing. On the reading developed here, this is a precise instance of defensive coherence:

an organization that has secured its stability by making its own renunciation the object it cannot give up.

Dynamical systems supply the main mathematical vocabulary. Consider an interpretive state x evolving under a potential V with noise,

$$dx = -V'(x) dt + \sigma dW_t.$$

Minima of V are attractors, each with a basin of attraction, and the noise occasionally drives the state across the barrier separating one basin from another. For small σ the expected time to escape a basin whose barrier has height ΔV grows as $\exp(2\Delta V/\sigma^2)$, so that stability in the sense of persistence can be extremely high. The central point is that this measure of stability is silent about adequacy. A deep basin is deep because the barrier around it is high, which says nothing about whether the state at its bottom handles the wider range of situations that the organism may meet. Local stability supplies no sufficient measure of wider adequacy.

Hysteresis and metastability sharpen the distinction between persistence and flexibility. A system exhibits hysteresis when the path by which it moves between states depends on its history, so that returning conditions to an earlier value does not restore the earlier state, and an interpretation that has been reinforced may resist reversal even after the circumstances that justified it have passed. Metastability describes states that persist for long times without being the deepest available, and it captures an interpretation that is held firmly although a better organization exists elsewhere in the space. These notions allow the chapter to state its caution against the opposite error. It would be a mistake to treat instability as intrinsically emancipatory. A system that moves readily among basins may be incapable of sustaining any organization long enough to act, and the capacity to hold an interpretation is as much a condition of life as the capacity to leave it. What distinguishes a healthy organization is not low barrier height but the correspondence between its barriers and the actual costs and risks of the transitions they guard, and a barrier that has become decoupled from those costs is the formal mark of entrapment.

Chapter 11

Attachment and the Conditions of Revisability

Attachment theory enters the argument as a question about the conditions under which an organism can tolerate uncertainty and reorganization. In the tradition founded by Bowlby and elaborated empirically by Ainsworth and her colleagues, early relationships with reliable caregivers are understood to support the capacity to explore, to return to a secure base and to regulate distress, and later work extended these ideas to adult relationships (Bowlby, 1969; Ainsworth et al., 1978). The chapter uses this tradition in a restricted way. It avoids diagnosing Nietzsche, and it does not treat solitude as an attachment classification. Its subject is the possibility that reliable relationships help sustain the capacity to explore, to receive correction and to undergo changes that temporarily disrupt self-understanding.

Two questions that are easily conflated must be separated at the outset. Nietzsche's criticism of the ego concerns whether there is a simple, enduring subject behind thought and action, and his positive proposal in *Beyond Good and Evil*, sections 12 and 19, understands the soul as a social structure of drives and willing as a relation of commanding and obeying within us. The attachment-theoretical question concerns how a person develops the capacity to explore, to tolerate vulnerability and to sustain relationships. The two are logically independent. Rejecting a substantial ego does not establish emotional independence, and needing relationships does not establish a substantial ego. A self understood as an organization of drives could depend profoundly on relationships for its formation and stability, and nothing in the multiplicity thesis forbids this.

This sharpens the earlier discussion of the ego. If an identity coordinates a person's expectations and actions, as Chapter 3 argued, then challenging it

threatens more than a proposition. A criticism of a belief can feel like an injury to the person because the belief helps hold the identity together, and another person's independence can feel threatening where a sense of worth depends on being indispensable. In the terms of Chapter 9, the transition cost c of revising an identity is high exactly because so much coordinated activity depends on it. Relationships and institutions may distribute the burden. If the cost of revision is shared with those who can offer reassurance, correction and continuity during the period of disruption, the effective cost falls,

$$c_{\text{eff}} = c - s,$$

where $s \geq 0$ denotes the support available, and changes that would fail the inequality of Chapter 9 for an isolated person may pass it for a supported one. The expression is a bookkeeping device with no empirical content of its own, but it makes a substantive claim visible: autonomy can be understood partly through the availability of support for transformation, and an account that celebrates independence while omitting the supports that produced it describes only half of the conditions of revisability.

This yields the strongest question that can be put to Nietzsche in this area, namely that his account of self-overcoming may underdescribe the relational conditions that make self-overcoming possible. Someone's capacity to endure solitude, withstand disagreement or revise an identity might have been cultivated through reliable relationships, and a portrayal of the resulting independence as self-generated could overlook the support that helped produce it. The question is legitimate without first identifying solitude with avoidance, and it should not be mistaken for the claim that Nietzsche himself was insecurely attached, which no textual evidence can establish.

The paper *Beyond-Man and Relational Agency: A Reading of Zarathustra through Attachment Theory* (Ferrari, 2026) argues that Zarathustra's solitary self-overcoming resembles insecure-avoidant agency, in that independence becomes defensive self-sufficiency, relationships supply little corrective feedback and vulnerability receives insufficient acknowledgment. It presents its reading as structural and interpretive and disclaims any diagnosis of Nietzsche or his characters, a restraint that the present chapter shares, and it draws its account of avoidant attachment from Bowlby and from Mikulincer and Shaver (Mikulincer and Shaver, 2007). Its central question deserves retention: whether apparent autonomy conceals an inability to depend on others. The verdict is less secure than the question. De-

scribing Zarathustra's agency as absolute, unmodulated and largely unresponsive to others requires considerably more textual argument than the paper provides, and the next part, in Chapter 14, tests it against the evidence of the narrative. The illuminating distinction that the paper's question suggests is between independence that permits relationship and independence that must continually defend itself against relationship. Applied to the will to power, it asks whether self-mastery increases a person's capacity to encounter and be changed by others, or whether maintaining the identity of a self-sufficient person becomes an overriding constraint.

Chapter 12

Value Creation and the Problem of the Objective

The deepest difficulty in the comparison concerns what evaluates changes in what is valued. In the formalism of Chapter 5, preferences are encoded in the distribution $p(o)$ over preferred outcomes, which appears in the expected free energy of policies and which is part of the generative model and not derived from it. A model that optimizes action relative to specified preferences has not yet explained the authority, formation or transformation of those preferences. If the preferences used to evaluate policies can themselves change, one must ask what evaluates those changes.

One response is to place the preferences in a higher level of a hierarchical model, so that they are themselves subject to learning. A prior over preferences C_0 is then updated in light of experience by some operator R_0 that depends on a still higher set of commitments C_1 , and so on. One way of closing such a hierarchy ends in a level whose commitments are held fixed, and an account built in this way can represent changes in preference only by letting an unrevised evaluator carry the authority for everything below it. This is not the only possible architecture. An implemented system must have specified dynamics, but specified dynamics need not amount to an unchanged evaluator that ranks every transformation, and mutually modifying processes, in which no level is exempt from revision by the others, and path-dependent changes are not excluded by finitude. What the hierarchical case does bring out is the difference between explanation and justification. A formal account of how preferences change does not by itself justify those preferences or confer normative authority on the process that changes them, and the question of authority has to be asked of whichever architecture is adopted.

Nietzsche's value creation should be approached through this problem while resisting its reduction to adjusting a reward function. At least three kinds of change are involved and the formalism treats them unevenly. A change in desired outcomes alters which states the agent seeks and is the case that the hierarchy above represents most easily. A change in the relative authority of concerns alters which of several objectives governs conduct, which in the notation of Chapter 4 is a change in the weights w_i and not in the objectives F_i themselves. A change in what counts as an intelligible goal alters the space of possible objectives, so that outcomes that could not previously be wanted become available to want, and this is a structural revision of the kind described in Chapter 9 applied to the objective itself. The third kind is the one that the scalar picture handles worst, because a scalar objective is defined over a given space of outcomes and cannot represent the enlargement of that space as a movement within it.

The ascetic ideal again furnishes a test. When the will turns from worldly objects to willing nothingness, as the *Genealogy* describes, it is difficult to say that a preference has been reweighted, since what has altered is the way the will conceives of its own aims. The case suggests that will to power in Nietzsche's use is better understood as the organizing activity within which preferences are formed than as a preference of especially high weight. If so, it does not occupy the role that $p(o)$ occupies in active inference, and the temptation to identify them rests on a category mistake. The appropriate comparison is between the organizing activity and the process by which a model's preferences, attentional weights and structure are jointly settled, which the standard formalism leaves outside its scope.

Whether formal accounts can represent these transitions without silently installing an unchanged higher-order evaluator remains the open question of the book. The argument of the remaining parts does not claim to resolve it. It claims that the question is correctly posed here, that the formalism supplies exact tools for describing the preservation of an organization and the revision of a model within it, and that the difference between preserving an interpretation that makes life manageable and transforming it so that new forms of life become possible depends on an account of value that the formalism, as ordinarily stated, does not itself supply.

Part IV

Literary Form as an Experiment in Interpretation

Chapter 13

Aphorism, Fragment, and the Interrupted Reader

Nietzsche's literary forms belong to his philosophical method and cannot be treated as packaging for doctrines stated elsewhere. An aphorism can invite confident interpretation and then leave its grounds underdetermined. A sequence of sections can establish an expectation that a later fragment unsettles. A passage can be spoken in a voice whose relation to the author is left open. In each case the reader must decide whether to assimilate the disruption, to revise the interpretation under way, or to preserve incompatible possibilities, and what the reader does is the activity of interpretation that the earlier chapters described in the abstract. In the preface to the *Genealogy* Nietzsche insists that an aphorism has not been deciphered merely because it has been read, and that its exegesis, which demands an art of its own, only then begins. The remark presents reading as work that the text solicits and does not complete.

Reader-response theory and philosophical hermeneutics supply resources for describing these activities. Iser's account of the act of reading emphasizes the gaps and indeterminacies that a text leaves for the reader to fill, and the way expectations formed at one point are modified by what follows (Iser, 1978). Gadamer's hermeneutics treats understanding as the fusion of horizons in which the interpreter's prejudgments are brought into question by the text they bring to bear on it (Gadamer, 1989). Nehamas has argued that Nietzsche's styles of writing are bound up with his conception of a life composed as a literary character, so that the form of the work enacts the view it expresses (Nehamas, 1985). Predictive processing offers a restricted comparative vocabulary for the same phenomena, namely expectation, discrepancy and revision, and the aim of this chapter is to

say what that vocabulary captures and what it risks flattening.

A minimal model of reading treats the reader as inferring, after each fragment, a posterior over interpretive hypotheses $h \in \mathcal{H}$. Let $f_{1:t}$ be the fragments read so far, and let the posterior be

$$q_t(h) \propto p(h) \prod_{u \leq t} p(f_u \mid h, f_{1:u-1}).$$

An interruption is a fragment f_{t+1} that is improbable under the currently dominant hypothesis but not under others, so that the posterior shifts. The three responses named above then have a formal counterpart. Assimilation retains the dominant hypothesis and absorbs the fragment by lowering its precision or redescribing it as a local exception. Revision transfers weight to a different hypothesis. Plural retention keeps several hypotheses alive, so that the entropy of q_t is not allowed to collapse. The formalism does not, on its own, assign a value to this response. The entropy of q_t is a posterior uncertainty about which interpretation obtains, which is a different quantity from the ambiguity term of Chapter 5, an expectation of the entropy of observations given states, and several incompatible readings can remain plausible even when each predicts the features of the text with precision. Neither posterior entropy nor likelihood ambiguity says what distinguishes productive interpretive plurality from uncertainty that contributes nothing to understanding, and literary judgment evidently draws that distinction. This is a first place where a generic discrepancy measure falls short of the practice it describes: it has no resources for saying which plurality counts as an achievement.

Irony exposes the limitation more sharply. An ironic utterance supports a content that is the opposite of its surface, and it does so through a stance that the reader must infer from cues. A faithful model therefore needs a latent variable for stance, $k \in \{\text{sincere}, \text{ironic}\}$, so that hypotheses are pairs $h = (c, k)$ of a content and a stance, and the likelihood of the same words depends on both. The reader's inference about k is itself learned from the author's habits, from genre and from the surrounding text. Once this variable is admitted, the single quantity called prediction error splits into discrepancies at different levels, and some of the most interesting effects in Nietzsche, in which a passage is at once endorsed and mocked, correspond to postures in which the reader cannot settle k and in which the instability of k is the point. Historical allusion and the double use of a term introduce further latents. Each such addition is a modelling decision that goes beyond the basic account, and one lesson of the chapter is that the adequacy of the predictive vocabulary for literature depends on how much structure is supplied

to it by the analyst.

The material form of the text enters here, as it did in Chapter 1. A reader of the editorial arrangement of the notebooks meets sequences whose order is the editors', and may mistake the order for an argument. The consequence for interpretation is the same as that of a precision setting imposed from outside: the reader's expectations are shaped before reading begins, and the frame is mistaken for the content.

Chapter 14

Zarathustra's Voice and the Hazards of Discipleship

Thus Spoke Zarathustra, subtitled a book for all and none, multiplies the positions from which words are spoken. The chapter distinguishes Nietzsche as author, a narrator who reports Zarathustra's journeys and speeches, Zarathustra himself, the figures to whom he speaks, including his disciples, the animals who accompany him and the higher men who gather at his cave in the last part, and finally the reader, who may be addressed through any of these. It investigates how prophetic authority is constructed and then complicated by the narrative, and it declines to treat every declaration as a transparent statement of doctrine. A speech delivered by Zarathustra to a crowd that cannot understand him is a different kind of utterance from a speech delivered to his animals, and a speech whose terms the narrative subsequently undercuts has a different status again.

The central issue is the tension between teaching self-overcoming and creating followers attached to the teacher's authority. The text states the tension itself. In the prologue Zarathustra says that the creator seeks companions and not followers, and at the close of the speech on the bestowing virtue he asks his disciples to lose him and find themselves, promising to return only when all of them have denied him, and observing that a teacher is repaid poorly by a pupil who remains only a pupil. The device of departure and return is repeated across the work. Zarathustra withdraws to solitude, comes back to the company of others, finds that his teaching has been misunderstood or turned into doctrine, and leaves again. Later episodes continue the theme, among them the shadow who follows him, whom he tells to go free, and the gathering of the higher men, who in their relief at finding a master end by worshipping an ass. The comedy of the final scenes is evidence that the

narrative itself treats the worship of the teacher as a hazard to be mocked and not as an outcome to be sought.

These features bear on the attachment-theoretical reading developed in Chapter 11. That reading holds, in the terms of the paper by Ferrari (Ferrari, 2026), that Zarathustra's solitary self-overcoming resembles insecure-avoidant agency, with independence functioning as defensive self-sufficiency, little corrective feedback from relationships and insufficient acknowledgment of vulnerability. The text supplies evidence for such a reading. Zarathustra repeatedly retreats to solitude when his teaching is misunderstood, he is addressed in the work as one who longs for the solitude that he calls his home, and his friendships, where they appear, are frequently strained or marked by his impatience. A reader inclined to the attachment lens can accumulate such passages.

The text also supplies evidence that complicates it, and the interpretive test is whether the reading survives its own counterevidence. The speech on the bestowing virtue can be read as the rejection of dependent discipleship in favour of a relation between more independent participants, with the promised return indicating a relationship that continues across separation. Desire for relationship and defensive avoidance are not mutually exclusive, and the invitation to companions in the prologue or the repeated return to people therefore cannot by itself refute the avoidant reading, since an agent may long for company and still organize his conduct so as to limit what company can ask of him. The weight of the analysis falls on scenes of actual exchange, and these are mixed. In "The stillest hour" at the close of the second part, Zarathustra is addressed by a voice he cannot resist, confesses fear and says that he lacks the voice for commanding, and is told that he must still become a child, a scene of vulnerability and of being commanded, although the speaker is internal. In "The Convalescent" he collapses after confronting his most abysmal thought and lies for seven days while his animals attend to him and then speak with him, a scene of receiving care, though from animals and not from persons. His meetings with the wanderer who calls himself his shadow and with the higher men in the fourth part stage companionship and weariness together. None of this proves secure attachment, but it complicates the claim that separation expresses defensive withdrawal as a general matter. The textual basis the paper offers is also thin. Its characterization of Zarathustra as unilateral, isolated and unresponsive to relational feedback rests on two quotations, the proclamation of the overman in the prologue and the passage on burning in one's own flame, and neither is drawn from a scene in which Zarathustra and another person actually exchange anything. A further difficulty is that the paper's

description of Zarathustra's agency as absolute, unmodulated and unresponsive to others fits poorly with scenes in which he is moved, exhausted or reproached by others.

One textual correction is needed. The paper locates the passage about burning in one's own flame in the speech titled "On the Sublime Ones" in the second part, whereas the passage belongs to "On the Way of the Creating One" in the first part. The correction is more than a matter of reference, because the image concerns transformation through self-overcoming, which is a statement about a process of becoming, and by itself supplies little evidence about whether a person can receive care or respond to relational feedback. Evidence of the second kind has to be found in scenes of actual exchange.

The conclusion of the chapter retains the question and weakens the verdict. The distinction that organizes the evidence is between independence that permits relationship and independence that must defend itself against relationship, and the narrative stages both without simply choosing between them. A reader who finds only avoidance has not accounted for the invitations and returns, and a reader who finds only openness has not accounted for the retreats and the contempt. The text's own refusal to settle the matter is consistent with the reading of Chapter 13, on which the form of the work asks the reader to hold competing interpretations of autonomy and relationship and to decide what to make of the instability.

Chapter 15

Metaphor, Concept Formation, and the Invention of Distinctions

Metaphor is a possible means of reorganizing salience and conceptual structure, and Nietzsche gives it a foundational role. In the early essay “On Truth and Lies in a Nonmoral Sense,” which he did not publish (Nietzsche, 1979), concepts are described as the residue of metaphors whose figurative origin has been forgotten, so that what passes for literal truth is the accumulated and worn-out product of transfers between spheres. Whatever one makes of the historical claim about language, the thought has a systematic consequence for the framework of this book. Schematization, in the sense of Chapter 2, requires a stock of distinctions, and the origin of new distinctions is the question that Chapter 9 left open. Metaphor is one candidate for where some of them come from.

Two functions can be distinguished. Some metaphors help a reader assimilate an unfamiliar subject to an existing category, so that the unfamiliar is handled by the repertoire already in place. Others make previously unnoticed relations available for thought, so that something is seen as having a structure it was not seen to have. Only the second function bears on structural revision. A way of stating the difference is that assimilating metaphors act within the current representation, whereas generative metaphors alter it.

The mathematical discussion returns to representation maps. Let $\phi : X \rightarrow Z$ be the existing organization and let a candidate representation be $\phi' : X \rightarrow Z'$. Three relations between them correspond to three effects. If $\phi = g \circ \phi'$ for some map g , the new representation refines the old one and introduces distinctions that the old one lacked. If $\phi' = g \circ \phi$, it coarsens it and renders earlier distinctions dispensable. If neither holds, the two cut the state space along incomparable lines,

and the new representation both preserves some distinctions and loses others. A metaphor can be modelled as a partial structure-preserving map h from a source domain with relations R_S to a target domain, proposing the relations $h(R_S)$ as new relations on the target. Whether the proposal is productive is then assessed by the criterion of Chapter 2: whether the representation it induces is sufficient, to within an acceptable margin, for the tasks that matter, and whether it opens tasks that the previous one could not formulate.

Model selection and minimum description length supply a second comparison. A description length criterion prefers the representation minimizing the sum of the length of the model and the length of the data given the model, in the manner of Rissanen's principle (Rissanen, 1978). Such criteria capture a real virtue, the compression of regularities, and they might be expected to favour metaphors that unify disparate phenomena. Neither shorter descriptions nor greater predictive accuracy is assumed here to exhaust literary or philosophical value. A metaphor may lengthen the description of the data in the short term, since it introduces a structure whose payoff arrives only later or in another domain, and a metaphor's worth can lie in the questions it makes askable and not in the accuracy of its answers. The criteria also assume that the data to be described are given, whereas a generative metaphor can alter what counts as data.

The chapter therefore asks how a text can expand interpretive possibilities without immediately settling them. A metaphor that settles its relation to the target at once becomes a definition and loses its generative force, whereas one that leaves the transferred structure partly open invites the reader to continue the mapping. This returns the argument to the reader. The transformation of the reader's repertoire is accomplished through the reader's own work with the figure, and to that extent the text enacts the claim of Chapter 9 that new distinctions are not obtained by inference within a fixed family but made available to it. Whether such availability can be assessed by any standard other than the reader's subsequent capacities is a question that the final part carries into the discussion of shared models and institutions.

Part V

Beyond Individual Interpretation

Chapter 16

Shared Models, Institutions, and Collective Self-Confirmation

The argument so far has treated interpretation as the achievement of an individual organism, but the organization of experience is largely inherited and socially maintained. Language supplies the distinctions that a person can draw before any inquiry begins. Education trains attention and fixes which discrepancies a learner is taught to notice. Scientific instruments, record-keeping practices, habits and institutions determine what can be encountered at all and what can be recognized as evidence when it is. A model that an individual holds is therefore frequently a local instance of a model that a community maintains, and the questions of the earlier chapters, about insulation, precision and revisability, recur at the collective scale with their own mechanisms.

The formal treatment introduces coupled agents $i = 1, \dots, n$ with distinct observations, models and preferences, who exchange reports and influence one another's updates. The simplest tractable case is linear averaging. Let $x_t \in \mathbb{R}^n$ collect the agents' current estimates of a quantity and let W be a row-stochastic matrix recording how much weight each agent places on each other's estimate, so that

$$x_{t+1} = Wx_t.$$

When the network is connected and aperiodic, the estimates converge to a consensus, and the consensus value is $\pi^\top x_0$, where π is the left eigenvector of W for eigenvalue one, normalized to sum to one. The weights π_i measure the long-run influence of each agent's initial estimate. Agreement is thus guaranteed by the structure of the coupling, but its content depends on who holds influence. Golub and Jackson showed that if initial estimates are independent noisy signals of a

truth, the consensus converges to the truth as the group grows only if no agent retains non-vanishing influence (Golub and Jackson, 2010). Where one authority is heavily weighted, the group converges just as firmly, and converges on that authority's initial error.

Agreement can arise in several ways that the observation of agreement alone does not distinguish. It can arise from shared evidence, in which case the convergence is justified to the extent that the evidence is good. It can arise from dependence on a common authority, in which case a single source is counted many times. It can arise from coordinated inquiry in which agents test one another's claims against independent encounters. And it can arise from the suppression of dissent, in which disagreement is removed from the data of the system and not resolved. The double-counting case admits a simple statement. Suppose that n agents each form a log-odds estimate ℓ for a hypothesis from one and the same signal, and that a naive pooling procedure sums the reports as though they were independent. The pooled value is $n\ell$, although the evidence supports only ℓ , so that apparent confidence grows with the number of agents while the information stays fixed. A community in which agents cite one another, each citing in turn the same source, displays exactly this structure.

The notion of insulation introduced in Chapter 6 carries over. Let μ denote the distribution over situations that a community's institutional practices lead it to encounter, and ν the distribution over situations that matter for its continued viability. A large gap between the error the community's model makes on ν and on μ is compatible with collective selective exposure, a self-confirming arrangement in which instruments, categories and incentives combine to supply only the evidence the model already handles. Like its individual counterpart, the gap may instead reflect legitimate specialization, and which reading is correct requires evidence about how the encountered situations were selected. Where the arrangement is one of self-confirmation, because it is distributed, no individual need intend it, and no individual may be in a position to see it. The chapter accordingly examines the conditions under which collective convergence improves knowledge and those under which it merely stabilizes a shared error. Convergence improves knowledge when the sources feeding it are independent, when influence is not concentrated, and when disagreement can enter the record. It stabilizes error when these conditions fail, and from inside the community the two cases may look alike.

Chapter 17

Power, Truth, and the Availability of Correction

A perspectival account of knowledge faces an obvious objection. If every interpretation is a perspective shaped by interests and capacities, there seems to be no ground for distinguishing stronger inquiry from effective self-insulation, since both are ways of organizing experience in the service of an organization's persistence. This chapter proposes a criterion and argues for it. The criterion concerns an interpretation's ability to sustain consequential encounters with what it does not already accommodate. Such a criterion requires argument, and it is not presented as an automatic consequence of either Nietzsche or active inference.

Nietzsche supplies a resource in the third essay of the *Genealogy*, where he rejects a view from nowhere and claims that the more affects are allowed to speak about a thing, and the more different eyes can be brought to observe it, the more complete the concept of that thing, and the objectivity attainable, will be. Perspectivism in this form is a claim about the conditions of better knowing: knowledge improves with the number and difference of the perspectives brought to bear and not through escape from perspective. The remark already points toward the institutional dimension of the present chapter, since a plurality of perspectives has to be housed somewhere and its challenges have to be able to reach the organization they challenge.

The discussion connects epistemic authority to the practical availability of correction, and the following definition states the three conditions that the criterion involves.

Definition 17.1 (Correctability). An organization is *correctable* with respect to a class of challenges $\mathcal{C}$ if, for each $c \in \mathcal{C}$, (i) the challenge can be registered by the

organization at a precision that is not negligible, (ii) the cost borne by the party raising it is bounded and does not exceed what that party can reasonably sustain, and (iii) a revision in response to the challenge does not destroy the conditions of the organization's own activity.

Each condition names a way correction fails. The first fails when precision on the relevant channel has been driven to zero by the mechanisms of Chapter 7, so that the challenge is received and not heard. The second fails when raising a challenge exposes the challenger to penalties severe enough that challenges are not raised, which makes the record of dissent a poor guide to the existence of grounds for it. The third fails when the organization cannot revise without collapse, so that even a challenge heard and welcomed cannot be acted on, and it is the condition through which the problem of transition cost from Chapter 9 enters. Who can challenge a model, what costs do they bear, and can the organization receiving the challenge change without destroying the conditions of its own activity are, in this formulation, three questions with three distinct answers.

This gives the monograph an institutional dimension and prevents self-overcoming from remaining an exclusively heroic individual achievement. The supports discussed in Chapter 11 for a person reappear as the arrangements through which a community can hear and absorb correction: procedures for adversarial testing, protected channels for dissent, redundancy of independent sources, and structures that spread the cost of revision. Where such arrangements exist, a community can transform itself without the intervention of an exceptional individual. Where they are absent, a community depends on the rare person willing to bear the full cost of challenge, and the conditions of the third kind make it likely that even then nothing changes.

The criterion should not be mistaken for a theory of truth, and it carries a limit that the final chapter records. It identifies the availability of correction as a mark of an inquiry that is capable of becoming better, and it is silent on whether the direction of change is an enhancement. An organization can be highly correctable and be corrected toward a worse condition by challenges that are themselves illusory. The criterion thus discriminates between self-insulating and self-revising organizations, which is what the perspectival account needed, without deciding which revisions deserve affirmation.

Chapter 18

The Limits of the Synthesis

A comparison should state where it ceases to hold. The preceding chapters identified several points of real contact between Nietzsche's treatment of interpretation and the active-inference family of models. Schematization corresponds to coarse-graining relative to a task, with the difference that the formal notion is exact about what is lost. The ego as an interpretation of activity corresponds to a self-model that summarizes an organism's processes without disclosing them. The dependence of an error's influence on its weighting corresponds to precision, and the dependence of evidence on what the organism does corresponds to the policy-dependence of observations. Each of these correspondences is a resemblance of problem, supported by a formal construction, and each was advanced with a statement of what it leaves out.

Three limits deserve final statement. The first is whether will to power can be represented through a scalar objective. The argument of Chapter 4 showed that a weighted sum of objectives is a coherent form only where the objectives are commensurable, and Chapter 12 argued that the will to power functions in Nietzsche's use as an organizing activity within which preferences are formed, and not as a preference of high weight. If so, the attempt to identify it with the quantity minimized in active inference, or with any single information-theoretic measure, mistakes the category of the notion. The second limit concerns evaluative conflicts, which may not admit a common measure at all. Where the conflict is between ways of life and not between values on a shared scale, the formalism's tools of weighting and tradeoff presuppose what is in dispute. The third limit is whether literary transformation can be captured by changes in a generative model. Chapters 13 to 15 found that the predictive vocabulary describes some of what a text does to a reader, namely the shifting of expectation and the reallocation of attention, but that it needs added structure to represent irony and allusion, and

that it offers no way of separating productive plurality from idle uncertainty or of saying why one disruption illuminates while another merely disorients.

A distinction runs through all three limits and should be kept explicit in the conclusion. Empirical questions about cognitive mechanisms, such as whether perception is implemented by hierarchical prediction, are answered by evidence about organisms, and the formal tools may prove accurate or inaccurate as models of them. Philosophical judgments about which forms of life deserve affirmation are questions of a different kind, and no result about mechanism settles them. The syntheses most likely to mislead are those that present a finding of the first kind as though it carried a verdict of the second, or that borrow the authority of a mathematical result for a valuation that the mathematics does not contain.

The conclusion that the book defends preserves a productive asymmetry. Active inference offers formal tools for understanding perception, action, learning and exploration, and it supplies exact vocabulary for what it is to preserve an organization and to revise a model within a family. Nietzsche presses questions about the authority and transformation of the organizations those tools describe, in particular how a family of models or a hierarchy of preferences acquires its standing and what it would be to replace it. The conjunction is most useful when each exposes questions left open by the other. From the formal side, the open questions are the origin of structural novelty and the authority of the objective. From Nietzsche's side, they are the conditions under which self-overcoming can be sustained and the standards by which its results are to be judged, and the chapters on attachment, institutions and correctability suggest that part of the answer lies in the relations and arrangements that make revision bearable.

The question posed at the outset was how an organism can remain sufficiently organized to act while retaining the capacity to revise the interpretations through which action becomes possible. The answer developed here is conditional and in part negative. The capacity to revise depends on exposure that is not engineered to confirm, on weights that can themselves be challenged, on supports that distribute the cost of transformation, and on access to alternatives that lie outside the family currently in use, which optimization within that family does not itself explain. Undergoing such revision may therefore require more than accepting a better proposition. It may require changes in attention, affect, identification and the reader's relation to authority, which is why a book about interpretation had to attend to how its principal author writes, and not only to what he says.

Appendix A

Notation and Conventions

Throughout, X denotes a detailed state space and Z a space of distinctions available to an organism. Maps $\phi : X \rightarrow Z$ are called interpretive organizations. Objectives are written F or F_i , and the symbol $\mathcal{M}$ is reserved for an admissible model family. Quotations and paraphrases of Nietzsche follow the English editions listed in the bibliography and are cited by section number where the work is divided into numbered sections, which allows readers to consult any edition. Passages of the compilation *The Will to Power* are cited by its section numbers, and the critical edition of Colli and Montinari is the authority for their notebook dating.

Appendix B

Variational Free Energy and Expected Free Energy

Variational free energy

Let $p(o, s)$ be a generative model and $q(s)$ an approximate posterior. The variational free energy is

$$F[q; o] = \mathbb{E}_{q(s)}[\log q(s) - \log p(o, s)] .$$

Since $p(o, s) = p(s | o) p(o)$, the logarithm of the joint splits as $\log p(s | o) + \log p(o)$. Substituting, and noting that $\log p(o)$ does not depend on s ,

$$F[q; o] = \mathbb{E}_{q(s)}[\log q(s) - \log p(s | o)] - \log p(o) = D_{\text{KL}}(q(s) \| p(s | o)) - \log p(o) .$$

Because $D_{\text{KL}} \geq 0$ with equality if and only if $q = p(\cdot | o)$, it follows that $F \geq -\log p(o)$, and minimizing F over q yields the exact posterior whenever the family of q contains it. When the family is restricted, the minimizer is the member of the family closest to the posterior in the sense of D_{KL} , and F at the minimum is a tighter or looser bound on the surprise accordingly. The bound is always relative to the generative model p .

Expected free energy and its two decompositions

For a policy π write $q(o, s | \pi) = q(s | \pi) p(o | s)$ for the predicted joint, let $p(o)$ encode preferred outcomes, and let $q(s | o, \pi)$ denote the posterior that would

follow from observing o under π . Define

$$G(\pi) = \mathbb{E}_{q(o,s|\pi)}[\log q(s | \pi) - \log p(o) - \log q(s | o, \pi)].$$

Grouping the first and third terms gives the negative expected information gain, since

$$\begin{aligned} \mathbb{E}_{q(o,s|\pi)}[\log q(s | \pi) - \log q(s | o, \pi)] &= -\mathbb{E}_{q(o|\pi)}[D_{\text{KL}}(q(s | o, \pi) \| q(s | \pi))] \\ &= -\text{IG}(\pi). \end{aligned}$$

Hence

$$G(\pi) = -\text{IG}(\pi) - \mathbb{E}_{q(o|\pi)}[\log p(o)],$$

the sum of an expected cost, measured by the negative log preference of predicted outcomes, and an epistemic term that rewards expected resolution of uncertainty. The information gain equals the mutual information between states and observations under the predicted joint, so that

$$\text{IG}(\pi) = H[q(o | \pi)] - \mathbb{E}_{q(s|\pi)}[H[p(o | s)]] .$$

Substituting and using $D_{\text{KL}}(q(o | \pi) \| p(o)) = -H[q(o | \pi)] - \mathbb{E}_{q(o|\pi)}[\log p(o)]$ yields the second form used in Chapter 5,

$$G(\pi) = D_{\text{KL}}(q(o | \pi) \| p(o)) + \mathbb{E}_{q(s|\pi)}[H[p(o | s)]] ,$$

a risk term plus an ambiguity term. Both forms contain $p(o)$, which is stipulated and not derived, and this is the formal location of the difficulty examined in Chapter 12.

Appendix C

Coarse-Graining and Task-Relative Sufficiency

Recall the definition of Chapter 2. For a task T with loss $\ell(a, x)$ over actions a and states $x \in X$, and a distribution over X , let R_X denote the least expected loss over policies that may depend on x , and let R_ϕ denote the least expected loss over policies that depend on x only through $\phi(x)$. The organization ϕ is ε -sufficient for T if $R_\phi - R_X \leq \varepsilon$. Since policies depending on $\phi(x)$ are a subset of those depending on x , $R_\phi \geq R_X$ always.

Proposal C.1. If the loss depends on x only through $\phi(x)$, that is $\ell(a, x) = \tilde{\ell}(a, \phi(x))$ for some $\tilde{\ell}$, then ϕ is 0-sufficient.

Proof. Take any policy π_X depending on x . For each value z in the range of ϕ , the loss is constant across the fibre $\phi^{-1}(z)$ for any fixed action, so the expected loss of π_X is a mixture of fibre-wise losses and is minimized by choosing in each fibre the action that minimizes $\tilde{\ell}(\cdot, z)$ averaged over the within-fibre distribution of π_X 's choices. That action depends on x only through $z = \phi(x)$, so a policy depending only on $\phi(x)$ attains the same minimum. $\square$

An example shows the fragility discussed in Chapter 2. Let $X = \mathbb{R}^2$ with coordinates (a, b) and let $\phi(a, b) = a$. For a task whose loss is $(u - a)^2$, where u is the action, ϕ is 0-sufficient, because the best policy is $u = a$ and requires nothing about b . If the task changes to the loss $(u - a - b)^2$, the best policy using ϕ alone is $u = a + \mathbb{E}[b \mid a]$, whose excess loss over the best policy $u = a + b$ is, conditionally on a , $\text{Var}(b \mid a)$. The overall excess expected loss is therefore $R_\phi - R_X = \mathbb{E}[\text{Var}(b \mid a)]$, assuming the second moments exist. The map ϕ is unchanged, and its sufficiency has passed from zero to the expected conditional variance of the discarded coordinate.

Appendix D

An Example of Structural Revision

Let the observations be $y_t = A \sin(\omega t) + \eta_t$, with independent noise η_t of variance σ^2 , and let the admissible family $\mathcal{M}_1$ consist of linear trends $y_t = \alpha + \beta t$. Over a long record spanning many periods, the least-squares fit within $\mathcal{M}_1$ converges to a nearly constant line at the mean, and the residual variance converges to $\sigma^2 + A^2/2$, because the oscillation contributes mean square $A^2/2$ that no member of the family can represent. No amount of parameter learning within $\mathcal{M}_1$ lowers the residual below this value, and the persistent excess over the noise floor σ^2 is the signal of Chapter 9 that the family omits structure.

The extended family $\mathcal{M}_2$ adds a sinusoidal component with free amplitude, frequency and phase. Within $\mathcal{M}_2$ the residual variance approaches σ^2 . Under squared-error loss the per-period reduction in expected loss from moving to $\mathcal{M}_2$ is $\Delta L = A^2/2$, so the criterion $\mathbb{E}[\Delta L] H > c$ of Chapter 9 reads $A^2 H/2 > c$. The example exhibits the pattern in a transparent case. The residual diagnoses the lack, comparison of marginal likelihoods ranks $\mathcal{M}_2$ above $\mathcal{M}_1$ once $\mathcal{M}_2$ is proposed, and the criterion weighs the gain against the cost. What the example does not supply is the proposal of $\mathcal{M}_2$, which here is made by the analyst who knows to look for periodicity. In the setting of the main text that is the part of the process that optimization within the family does not itself explain. A system that searched over candidate structures could make such proposals, at the price of presupposing the space it searches.

Appendix E

A Small Dynamical Model of Entrapment

Consider the overdamped dynamics $dx = -V'(x) dt + \sigma dW_t$ with the symmetric double-well potential $V(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$. The minima lie at $x = \pm 1$, where $V = -\frac{1}{4}$, and the barrier between them is at $x = 0$, where $V = 0$, so that $\Delta V = \frac{1}{4}$. For small σ the mean time to cross from one well to the other is of order $\exp(2\Delta V/\sigma^2) = \exp(1/(2\sigma^2))$, by Kramers' estimate, with a prefactor depending on the curvatures at the minimum and the barrier.

Suppose now that one well, say $x = -1$, corresponds to an interpretation that is adequate only on a protected domain, and the other to an interpretation adequate on the wider domain that matters for viability. The potential V records the stability of each state and not its adequacy, and the two are logically independent, which is the point of Chapter 10. To model reduced exposure, let the effective noise decline with the residence time τ spent in the protected well, measured from entry and reset on crossing, taking $\sigma_\tau = \sigma_0/(1 + \kappa\tau)$ for a constant $\kappa > 0$. This is a schedule that depends on history through residence, and it does not yet make exposure the outcome of what the organism does, which a fully coupled model would require. For a system that starts in the protected well and has not yet crossed, $\tau = t$, so the schedule coincides with $\sigma_t = \sigma_0/(1 + \kappa t)$. Assume, as an explicit simplifying condition, that the instantaneous crossing rate is $\Gamma_t = C \exp(-2\Delta V/\sigma_t^2)$ for a constant prefactor C . Under this stipulated hazard model the probability of at least one crossing is exactly $1 - \exp(-\int_0^\infty \Gamma_t dt)$, and the integrated hazard is bounded by

$$\int_0^\infty \Gamma_t dt = C \int_0^\infty \exp\left(-\frac{2\Delta V(1 + \kappa t)^2}{\sigma_0^2}\right) dt \leq \frac{C\sigma_0^2}{4\Delta V\kappa} \exp\left(-\frac{2\Delta V}{\sigma_0^2}\right).$$

The bound follows from substituting $u = 1 + \kappa t$ and using $u \geq 1$ on the range of integration. A finite integrated hazard establishes a positive probability of never crossing, and the bound is small when σ_0^2 is small relative to $2\Delta V$, so that under these assumptions a system that begins in the protected well has a small probability of ever leaving it, however long it runs. A comparison system with constant noise σ_0 has crossing rate $C \exp(-2\Delta V/\sigma_0^2)$ at every time and therefore crosses with probability one in the long run. The first system is entrapped and the second is flexible, and the difference arises from the dependence of exposure on residence and not from the potential.

The model is an illustration of the stated assumptions. It shows that persistence can be produced by a decline in exposure with residence and that the resulting stability carries no information about adequacy. It does not show that any organism behaves in this way, and it makes no claim about which interpretation is correct.

Bibliography

- Ainsworth, M. D. S., Blehar, M. C., Waters, E., and Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Lawrence Erlbaum.
- Bowlby, J. (1969). *Attachment and Loss*, Vol. 1: Attachment. Basic Books.
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, 29, 1–49.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.
- Ferrari, E. F. (2026). Beyond-Man and Relational Agency: A Reading of Zarathustra through Attachment Theory. PhilArchive preprint, archived 16 February 2026, <https://philarchive.org/archive/FERBAR-3>.
- Gadamer, H.-G. (1989). *Truth and Method*. 2nd rev. ed., trans. J. Weinsheimer and D. G. Marshall. Continuum.
- Iser, W. (1978). *The Act of Reading: A Theory of Aesthetic Response*. Johns Hopkins University Press.
- Mikulincer, M., and Shaver, P. R. (2007). *Attachment in Adulthood: Structure, Dynamics, and Change*. Guilford Press.
- Nehamas, A. (1985). *Nietzsche: Life as Literature*. Harvard University Press.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465–471.
- Golub, B., and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1), 112–149.

- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Nietzsche, F. (1967). *The Will to Power*. Trans. W. Kaufmann and R. J. Hollingdale. Random House.
- Montinari, M. (1982). *Nietzsche Lesen*. De Gruyter.
- Nietzsche, F. (1966). *Beyond Good and Evil*. Trans. W. Kaufmann. Vintage.
- Nietzsche, F. (1967). *On the Genealogy of Morals* and *Ecce Homo*. Trans. W. Kaufmann and R. J. Hollingdale. Vintage.
- Nietzsche, F. (1974). *The Gay Science*. Trans. W. Kaufmann. Vintage.
- Nietzsche, F. (1979). On truth and lies in a nonmoral sense. In *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870's*, ed. and trans. D. Breazeale. Humanities Press.
- Nietzsche, F. (2006). *Thus Spoke Zarathustra*. Ed. A. Del Caro and R. Pippin, trans. A. Del Caro. Cambridge University Press.
- Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87.
- Richardson, J. (1996). *Nietzsche's System*. Oxford University Press.