

COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS



Prediction and Perturbation: Towards A Theoretical
Integration of Predictive Processing and Systemic Therapy

MASTER'S THESIS

Study program: Cognitive Science

Field of study: Computer Science

Supervising department: Department of Applied Informatics

Supervisor: Priv.-Doz. Mag. Dr. Isabella Sarto-Jackson

Consultant: doc. RNDr. Martin Takáč, PhD.

2026

Silvana Adamyova



THESIS ASSIGNMENT

Name and Surname: Silvána Adámyová
Study programme: Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study: Computer Science
Type of Thesis: Diploma Thesis
Language of Thesis: English
Secondary language: Slovak

Title: Prediction and Perturbation: Towards A Theoretical Integration of Predictive Processing and Systemic Therapy

Annotation: This thesis examines the theoretical connections between Predictive Processing framework and therapeutic disruption techniques employed in Milan systemic psychotherapy. Both frameworks converge on a shared insight: that cognitive and relational systems change through encounters with the unexpected. The thesis identifies three core theoretical parallels concerning top-down constraint on perception, mechanisms of change, and active inference, and tentatively proposes that systemic interventions may derive their efficacy from inducing prediction errors that disrupt generative models.

Aim:

1. Review and synthesise the current literature on predictive processing, with particular focus on prediction error, precision weighting, and active inference.
2. Review the core principles of Milan systemic therapy, including circular questioning, perturbation, and Maturana's structural determinism.
3. Identify and articulate potential theoretical parallels between the two frameworks across three levels: top-down constraint on perception, mechanisms of change, and self-reinforcing patterns.

Literature: Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
Boscolo, L., Cecchin, G., Hoffman, L., & Penn, P. (1987). Milan systemic family therapy: Conversations in theory and practice. Basic Books.

Supervisor: Priv.-Doz. Dr. Isabella Sarto-Jackson, PhD.
Consultant: doc. RNDr. Martin Takáč, PhD.
Department: FMFI.KAI - Department of Applied Informatics
Head of department: doc. RNDr. Tatiana Jajcayová, PhD.

Assigned: 20.03.2026

Approved: 24.03.2026
prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

Student

Supervisor

Declaration

I hereby declare that the work presented in this thesis is original and the result of my own investigations. Formulations and ideas taken from other sources are cited as such.

Acknowledgements

I am deeply grateful to my supervisors, Isabella Sarto-Jackson and Martin Takáč, for their guidance throughout this project. Their willingness to engage with a thesis that crosses the boundaries between cognitive neuroscience, computer science, and clinical practice made the work possible, and their thoughtful feedback shaped both its argument and its scope.

I also wish to thank my boyfriend, Lenny, whose steady support sustained me through the long stretches of reading, writing, and revising that this thesis required. His cooking — consistently wonderful, and reliably timed for the moments when I most needed to step away from the screen — was an incredible show of love, and the thesis is infinitely better for it.

Abstrakt

Psychoterapia sa historicky rozvíjala v predstihu pred svojimi teoretickými základmi. Mnohé z jej najetablovanejších techník fungujú v praxi bez uspokojivého vysvetlenia, prečo fungujú, a systemická rodinná terapia túto medzeru ilustruje obzvlášť výrazne: jej klinická metodológia dlho predbiehala teoretický aparát, ktorý stále vychádza prevažne z biologických a kybernetických myšlienok spred niekoľkých desaťročí. Nedávny vzostup prediktívneho spracovania ako zjednocujúceho výpočtového rámca pre kogníciu, vnímanie a konanie tento obraz začal meniť a prístupy založené na prediktívnom spracovaní už boli navrhnuté pre viaceré terapeutické modalities vrátane EMDR, koherenčnej terapie a telesne orientovanej intervencie pri traume.

Táto diplomová práca rozširuje rovnaký projekt na milánsku systemickú terapiu. Práca má teoretický charakter a postupuje prostredníctvom literárneho prehľadu v oblastiach kognitívnej neurovedy a systemickej rodinnej terapie a prostredníctvom konceptualizácie teoretického rámca, ilustrovanej dôkladnou analýzou dvoch kanonických milánskych prípadových štúdií. Rozvíja tri hlavné príspevky: diagram zosúlad'ujúci základné dynamiky prediktívneho spracovania s dynamikami opísanými milánskymi systemickými terapeutmi; rámec PTPN — Perturbácia (*Perturbation*), Dôvera (*Trust*), Preferencia (*Preference*) a Nekonfrontačnosť (*Non-confrontation*) — špecifikujúci štyri podmienky, za ktorých možno očakávať, že nová informácia poskytnutá počas sedenia aktualizuje generatívny model rodiny namiesto toho, aby bola odmietnutá; a multiagentné rozšírenie, v ktorom sa argumentuje, že simultánna a iteratívna aktualizácia generatívnych modelov viacerých členov rodiny vedie k posunu v interakčnom vzorci, ktorý milánski terapeuti opisujú ako rozpustenie problému.

Kľúčové slová: prediktívne spracovanie; aktívna inferencia; milánska systemická terapia; predikčná chyba; PTPN; multiagentná inferencia

Abstract

Psychotherapy has historically developed in advance of its theoretical foundations. Many of its most established techniques work in practice without a satisfactory account of why, and systemic family therapy in particular illustrates the gap: its clinical methodology has long outpaced its theoretical apparatus, which still draws largely on biological and cybernetic ideas from several decades ago. The recent emergence of predictive processing as a unifying computational framework for cognition, perception, and action has begun to change this picture, and predictive-processing accounts have already been proposed for several therapeutic modalities, including EMDR, coherence therapy, and body-oriented trauma intervention.

This thesis extends the same project to Milan systemic therapy. The work is theoretical, proceeding through literature review across cognitive neuroscience and systemic family therapy and through framework conceptualisation, illustrated by close reading of two canonical Milan case studies. Three contributions are developed: a diagram aligning the core dynamics of predictive processing with those described by Milan systemic therapists; the PTPN framework — Perturbation, Trust, Preference, and Non-confrontation — specifying the four conditions under which new information delivered in a session can be expected to update the family's generative model rather than be rejected; and a multi-agent extension in which the simultaneous and iterative updating of several family members' models is argued to produce the shift in interactional pattern that Milan therapists describe as the dissolution of the problem.

Keywords: predictive processing; active inference; Milan systemic therapy; prediction error; PTPN; multi-agent inference

Table of contents

Introduction	1
1. Rationale and Scope	2
1.1 The Object of Analysis: Milan Systemic Therapy	2
1.2 The Candidate Framework: Predictive Processing	2
2. Foundations of Predictive Processing	3
2.1 A Note on Terminology	3
2.2 The Free Energy Principle	5
2.3 The Bayesian Brain Hypothesis	7
2.4 Predictive Processing	7
2.5 Action-Oriented Predictive Processing	9
2.5.1 Epistemic Value and Pragmatic Value Active Inference	10
2.5.2 Counterfactual Processing	12
2.6 Mechanisms of Prediction Error Minimisation	14
2.6.1 Perceptual Inference and Active Inference	14
2.6.2 Precision Weighting	16
2.6.3 Learning	17
2.6.4 False Inference	19
2.7 Multi-Agent Predictive Processing	23
2.7.1 Shared Generative Models	24
2.7.2 Counterfactual Processing in Multi-Agent Contexts	25
2.8 Predictive-Processing Accounts of Other Psychotherapies	27
3. Foundations of Milan Systemic Therapy	29
3.1 Case Studies	29
3.1.1 John, “The Crying Boy” (late 1970s)	30
3.1.2 Antonella, “The Sweater Girl” (early 1990s)	32
3.2 Theoretical Foundations of Systemic Therapy	33
3.2.1 General Systems Theory	34
What is General Systems Theory?	34
Implications of GST in MST	35
Examples in the Case Studies	35
3.2.2 Cybernetics	36
What is Second-Order Cybernetics?	36
Second-Order Cybernetics in Systemic Therapy	38
Examples in the Case Studies	39
3.2.3 Information as a Difference	40

What is Information as a Difference?	40
Information as a difference in systemic therapy	41
Examples in the Case Studies	41
3.2.4 Structural Determinism and Perturbation	42
What is Structural Determinism and Perturbation?	42
Structural Determinism and Perturbation in Systemic Therapy	44
Examples in the Case Studies	45
3.2.5 Constructivism	45
What is Constructivism?	45
Constructivism in Systemic Therapy	46
Examples from the case studies	47
3.2.6 Social Constructionism	48
What is Social Constructionism?	48
Social Constructionism in Systemic Therapy	49
Examples in the Case Studies	50
3.3 Key Concepts in Systemic Therapy	51
3.3.1 Resistance	51
Examples in the Case Studies	53
3.3.2 Neutrality	54
Examples in the Case Studies	55
3.3.3 Non-Normativity	56
Examples in the Case Studies	57
3.4 Key Methods in Milan Systemic Therapy	59
3.4.1 Circular Questioning	59
Examples in the Case Studies	64
3.4.2 Positive Connotation	65
Examples in the Case Studies	67
3.4.3 Prescription	68
Examples in the Case Studies	69
4. The Problem System	72
5. Milan Systemic Therapy from a Predictive Processing Perspective	74
5.1 How Systemic Therapy Delivers New Information	76
5.1.1 Questions Themselves Create Prediction Error Through New Information	76
5.1.2 Circular Questioning Reframes the Problem as Relational	77
5.1.3 Circular Questioning Introduces Multiple Perspectives on Reality	79
5.1.4 Answers as a Further Source of Prediction Error	80
5.1.5 Circular Questioning Widens the Space of Possible Outcomes	80
5.1.6 Positive Connotation Delivers a Non-Threatening Reframe	82

5.1.7 The Intervention as a Whole Functions as a Surprise	83
5.2 How Milan Systemic Therapy Creates Prediction Error That Updates Priors	84
PTPN: The Four-Factor Therapeutic Prior Update Framework	84
5.2.1 Perturbation	85
Examples in MST	86
5.2.2 Trust	87
Examples in MST	89
5.2.3 Preference	91
Examples in MST	93
5.2.4 Non-Confrontation	94
Examples in MST	95
5.3 How the Problem Dissolves in a Multi-Agent System	96
6. Discussion	99
6.1 Implications	99
6.1.1 Implications and Next Steps for Psychotherapy as a Whole	99
6.1.2 Implications and Next Steps for Predictive Processing	101
6.1.3 Implications and Next Steps for Milan Systemic Therapy	101
6.2. Limitations	102
6.2.1 Predictive Processing's Poor Falsifiability	102
6.2.2 Mismatch in Time Scales	104
6.2.3 Reinforcement Learning as an Alternative Framework	105
6.2.4 Interventions Outside the Scope of This Analysis	108
Conclusion	108
References	109
Appendix A: Background and evolution of Milan systemic therapy	133
A.1.1 The 1940s-1975: Early Influences and the Strategic-Systemic Approach	133
A.1.2 1975-1980: The Emergence of New Ideas and the Milan Systemic Approach	136
A.1.3 Early 1980s: The Split and the Rise of the Observer	137
A.1.4 Early 1980s: Applications to Individual Therapy	138
A.1.5 Late 1980s, 1990s: Critique of Systemic Thinking and Postmodern Ideas	139
A.1.6 Early 2000s: the Rise of Disorder	141
A.1.7 The Newer Models of Milan Systemic Therapy	141

List of Acronyms

CBT — cognitive-behavioural therapy

CMTF — Centro Milanese di Terapia della Famiglia (Milan Centre for Family Therapy)

EMDR — eye movement desensitisation and reprocessing

FEP — free energy principle

GST — general systems theory

HGF — hierarchical Gaussian filter

MI — motivational interviewing

MST — Milan systemic therapy

PP — predictive processing

PTPN — Perturbation, Trust, Preference, Non-confrontation

RL — reinforcement learning

List of Figures

Figure 1. Map of concepts and terminology related to predictive processing, as used in this thesis	4
Figure 2. Simplified scheme of the hierarchical predictive coding framework	9
Figure 3. A single-agent model of counterfactual processing.....	13
Figure 4. A comparison of the two ways prediction error is minimised	14
Figure 5. The prism glasses adaptation, showing learning through prediction error	19
Figure 6. Illustration of false inference under poor lighting conditions	20
Figure 7. The dress photograph and the effect of illumination context	21
Figure 8. The hollow mask illusion	22
Figure 9. An illustration of John's family	31
Figure 10. An illustration of Antonella's family	32
Figure 11. An illustration of Maturana's structural determinism and comparison to instructive interaction	43
Figure 12. An illustration of the homeostatic problem	73
Figure 13. A big-picture view of Milan systemic therapy from a predictive-processing perspective	75
Figure 14. How the problem changes from a multi-agent PP perspective	97

Introduction

In an era of growing public discussion of mental health and a proliferation of therapeutic models that address it, it is imperative to understand how therapeutic interventions produce their effects in the brain, so that the most effective forms of help can be identified. Yet many therapies, including systemic therapy, lack a mechanistic account of why their interventions succeed. Although a complete understanding of the brain remains a distant goal, this thesis offers a tentative framework intended to begin closing that gap and to guide future research.

This thesis proposes predictive processing (PP) as a candidate neural mechanism underlying the efficacy of Milan systemic therapy (MST). Drawing on a review of the relevant literature in both fields and a close reading of two published case transcripts, it identifies several points at which Milan interventions can be given a predictive-processing interpretation.

Chapter 1 sets out the reasoning behind the two central choices of this thesis: why MST is examined in particular, and why PP is proposed as its underlying cognitive framework. Because the argument draws on neuroscience, cognitive science, and psychotherapy, the thesis is written for an interdisciplinary readership; Chapters 2 and 3 therefore provide background on PP and on MST, respectively, so that readers from either side can orient themselves before the integrative chapters that follow.

Chapter 4 characterizes the problem system, and Chapter 5 then offers a PP reading of how the therapist intervenes in this problem system to produce change. Specifically, this chapter examines how prediction error is introduced into the clients' generative models, what causes this prediction error to lead to an update in priors, and how the analysis extends from the individual to the multi-agent level. Finally, Chapter 6 discusses the implications of the analysis for both PP and MST and identifies its limitations as well as suggestions for future work.

1. Rationale and Scope

1.1 The Object of Analysis: Milan Systemic Therapy

The analysis offered here focuses on Milan systemic therapy (MST) as developed and articulated by the original Milan team — Selvini Palazzoli, Boscolo, Cecchin, and Prata — between approximately 1975 and 1990. Several considerations support the choice to take this therapy as the object of analysis, particularly in its early formulation.

First, no neurocognitive account of MST currently exists. Other therapeutic modalities — including EMDR (Chamberlin, 2019), psychoanalysis (Carhart-Harris & Friston, 2010), and body-oriented trauma intervention (Fujimoto, 2026) — have been analyzed through PP or related frameworks, but no equivalent analysis has been offered for MST. This represents a gap in the literature. Second, although one of the original MST techniques examined in this thesis (i.e., positive connotation) is no longer in common clinical use, it had remarkable success in its time in producing therapeutic change. The mechanism of an intervention whose clinical effects are well documented but whose theoretical basis remains underspecified is of particular interest. Third, the original Milan team produced several detailed case studies that include near-verbatim transcripts of full sessions, which permits close analysis of the techniques as they were actually delivered.

1.2 The Candidate Framework: Predictive Processing

Several considerations support the choice of predictive processing (PP) as the candidate framework for this analysis.

Firstly, PP is one of the most actively developed theoretical programs in contemporary cognitive science, with a growing body of computational, neural, and behavioural evidence accumulating across perception, action, learning, and psychopathology (Badcock & Davey, 2024; Clark, 2013; Friston, 2010; Hohwy, 2013). The framework remains contested, but it has matured to the point of being applied to domains well beyond its original focus on perception. Second, PP has already been proposed as a candidate framework for several established

therapeutic modalities, including EMDR, psychoanalysis, and body-oriented trauma intervention, as mentioned in §1.1. Third, and perhaps most importantly, PP has particular potential for close alignment with MST because the two traditions share an intellectual ancestry. MST is grounded in Gregory Bateson’s cybernetic theory of mind, and the cybernetic tradition forms part of the conceptual genealogy of PP itself (Pezzulo et al., 2024). Both traditions take seriously the notion that minds, whether individual or collective, are nonlinear systems that update through feedback, and both reject linear stimulus-response accounts of how change occurs. This shared lineage is visible in a convergence on the same underlying claim: that movement toward a more accurate or more workable model of the world depends on encountering information that does not fit current expectations. The bridge attempted in this thesis is therefore not the forced combination of two unrelated frameworks but a translation between two lineages whose conceptual roots already overlap.

2. Foundations of Predictive Processing

2.1 A Note on Terminology

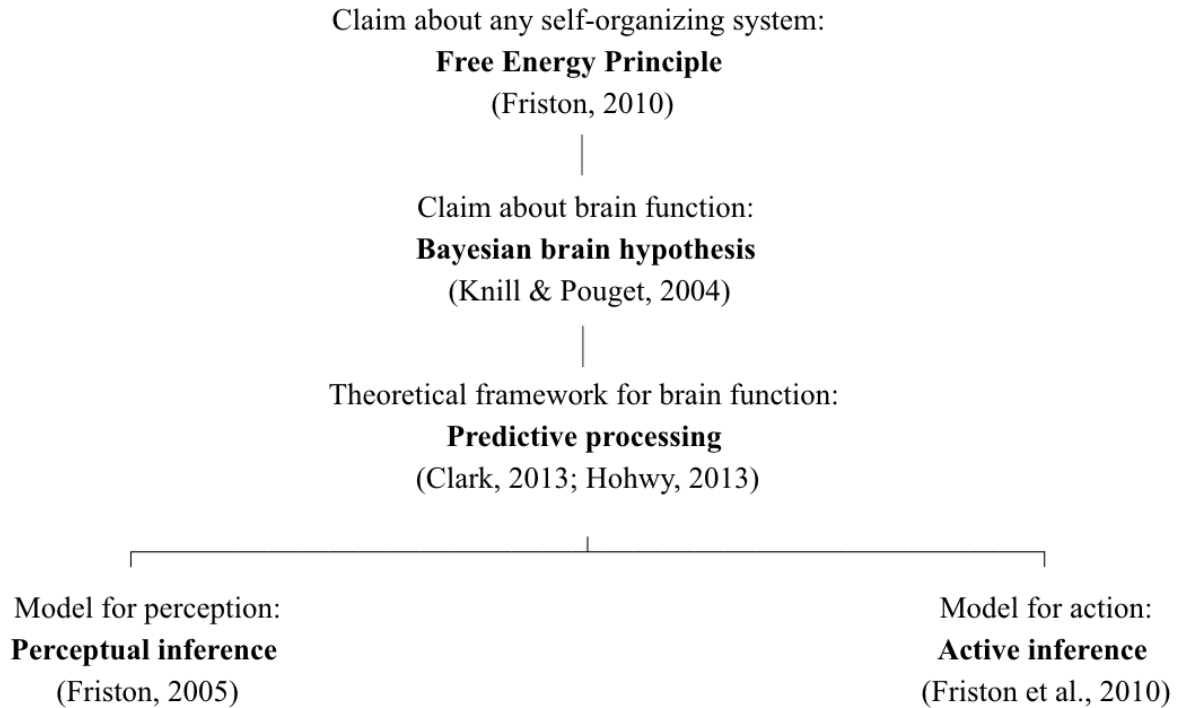
Predictive processing (PP) and its associated frameworks form a rapidly developing field, and the resulting terminology is not always used consistently. This can be disorienting for any reader new to the literature — as it was for me when I first encountered it — so this section briefly clarifies how the central terms are used in this thesis.

At the most general level sits the *free energy principle* (FEP), a formal claim about any self-organizing system. Below the FEP sits the *Bayesian brain hypothesis*, the claim that the brain performs Bayesian inference (Knill & Pouget, 2004) — although whether the Bayesian brain hypothesis follows formally from the FEP is contested (Andrews, 2021). Below that, PP is a specific kind of Bayesian brain theory in which the brain performs inference through top-down predictions compared against bottom-up sensory signals (Clark, 2013; Hohwy, 2013). PP originated in the perceptual domain, where it is implemented as predictive coding (Rao & Ballard, 1999). Active inference subsequently extended the framework to action, treating motor behaviour as a means by which the organism brings sensory input into line with

its predictions (Friston et al., 2010). Figure 1 maps these terms, consistent with Clark (2013, 2016) and Hohwy (2013).

Figure 1

Map of concepts and terminology related to predictive processing, as used in this thesis



A potential source of confusion is that *active inference* has gradually been broadened in the literature to serve as a cover term for the whole framework, effectively as a substitute for *PP*. Parr et al. (2022), for example, titled their textbook *Active Inference* and use the term in this broader sense, covering both perception and action. Meanwhile, other authors continue to use *PP* as the umbrella term, treating active inference as situated within *PP* (Clark, 2023; Hohwy, 2020; Khorrami Banaraki et al., 2024; Seth, 2021; Sprevak & Smith, 2023)

This thesis follows the usage shown in Figure 1: *PP* is the umbrella term, and *active inference* refers specifically to the action component of the framework. Where this thesis quotes or paraphrases authors who use *active inference* in the broader sense, the phrase *active inference framework* will be used to signal that broader meaning.

A final terminological note. The active inference framework literature has several names for the quantity whose minimisation drives updates to priors and generative models — most commonly *variational free energy* and *expected free energy*, with *prediction error* appearing as a specific limiting case of these quantities under particular assumptions about the generative model. The reader will encounter all of these terms over the course of this chapter. Authors who use *PP* as the umbrella term, however, tend to retain *prediction error* as the lingua franca for what drives updates to priors and models (Clark, 2013, 2016; Hohwy, 2013, 2020; Sprevak & Smith, 2023). For simplicity, and to keep the present account aligned with the earlier theoretical work in which the general principle was first articulated, this thesis also uses *prediction error* as the general term for whatever quantity drives updates to priors and generative models, except where a more specific term is required by the source being cited or paraphrased.

2.2 The Free Energy Principle

We live in a universe governed by entropy — things tend toward equilibrium, which in physical terms means disorganization or uniform dispersion (Clausius, 1865/1867). A familiar example is hot water poured into cold water: the two volumes do not stay neatly separated but mix into a single, uniformly warm body as the molecules disperse to a common temperature. The one exception to entropy is life, in which case equilibrium would mean death: in order to keep existing, a living system must resist entropy and maintain order against dissipative forces (cells must stay organized, body temperature must stay within a range, and so on). This capacity to maintain stable states against entropic forces has been termed *negentropy* (Schrödinger, 1944), and the negative-feedback system that maintains such states is called *homeostasis* (Cannon, 1929; see §3.2.2 for further discussion).

To maintain homeostasis, a system must have boundaries that separate it from the rest of the world while still allowing interaction with it. In the active-inference-framework literature, these boundaries are formalized as *Markov blankets* (Friston, 2013; Kirchhoff et al., 2018). Familiar examples include cell membranes for cells, the skin for the body, or at a higher level, the boundary of a family or an organization (Friston, 2013; Kirchhoff et al., 2018; Palacios et al., 2020).

The free energy principle (FEP) is the negentropic-homeostatic principle applied to any self-organizing system, including the brain (Friston, 2010, 2013). In this sense, *entropy* is used in Shannon's (1948) information-theoretic sense, to quantify the uncertainty of a probability distribution over states. High entropy means the distribution is spread out and outcomes are difficult to predict (e.g., it is equally likely for a body's temperature to be 20°C or 37°C); low entropy means the distribution is concentrated and outcomes are predictable (e.g., it is highly likely for a body's temperature to be 37°C, and highly unlikely for it to be 20°C). To persist, an organism must maintain low entropy and keep its sensory states confined to a narrow range of phenotype-defining values rather than being dispersed across all possible states (Friston, 2010, 2013). To do this, any self-organizing system must be able to *anticipate* which sensory states it is likely to encounter and act to bring about the ones consistent with its phenotype.

Yet, a system can never predict the world perfectly; there is always some residual mismatch between the sensory input it predicts and what is actually observed. This discrepancy is called *free energy*, or more specifically *variational free energy* to signify that it is an information-theoretic quantity, distinct from the thermodynamic free energy used in physics (Friston, 2010).

The FEP proposes that any system bounded by a Markov blanket must, to persist over time, minimise its variational free energy (and therefore its prediction error). If it does not, it will drift away from its characteristic states and dissolve (Friston, 2010, 2013). This is where the formulation appeared to raise a problem. If minimizing free energy is what an organism must do, then why does it not simply retreat into an isolated dark room, where sensory input is minimal and highly predictable? The response in the literature to this *dark-room problem* has two parts (Friston et al., 2012). First, a dark room is not in fact a low-free-energy state for an organism. A living system's generative model already contains strong probability over its characteristic states — such as being fed, being warm, and being in social contact (Friston, 2013; Friston et al., 2012). A dark room does not satisfy any of these. Its sensory consequences therefore violate the organism's expectations and generate substantial free energy. The free energy an organism minimises is not the absence of incoming signals but rather deviation from the kind of sensory inputs an organism of this kind expects to encounter (Friston et al., 2012). Second, when free energy is extended to action over time, the quantity

minimised includes an epistemic component: uncertainty about hidden states of the world (see §2.5.1). Exploring the environment, even at some immediate cost, reduces this uncertainty and therefore reduces expected free energy in the longer run (Friston et al., 2015; Schwartenbeck et al., 2019). On both grounds, the FEP predicts that organisms will avoid dark rooms rather than seek them out.

In summary, the FEP, if true, offers a unifying first-principles account of how any self-organizing biological system — from cells to brains to (potentially) families and other multi-agent systems — must minimise long-run free energy to persist over time (Friston, 2013).

2.3 The Bayesian Brain Hypothesis

Below the FEP, the Bayesian brain hypothesis is a claim specific to brain function (Knill & Pouget, 2004; Hohwy, 2013). The brain has no direct access to the world: it is locked inside the skull, and receives only neural signals from sensory transduction (Helmholtz, 1867/1962; Hohwy, 2013). Therefore, to infer the state of the external world, it must estimate the hidden causes of these signals. It does so by combining two sources of information through Bayes' rule. The first is the *prior*: the brain's belief, shaped by past experience and accumulated learning, about the likely state of the world before any new sensory evidence is taken into account. The second is the *likelihood*: how probable the current neural signals would be under each candidate state of the world. The combination of prior and likelihood yields the *posterior*: an updated belief about the most likely state of the world, given both what was expected and what was observed (Doya et al., 2007; Knill & Pouget, 2004). Each new observation updates the posterior, which in turn becomes part of the prior for the next observation. In this way, the brain continuously refines its beliefs about the world.

2.4 Predictive Processing

Predictive processing (PP) extends the Bayesian brain hypothesis with a specific framework for *how* inference is performed (Sprevak & Smith, 2023). In particular, it holds that the brain has a *generative model* of the world, which is what produces the prior and the likelihood (Clark, 2013; Friston, 2005). The generative model is a computational structure encoded in

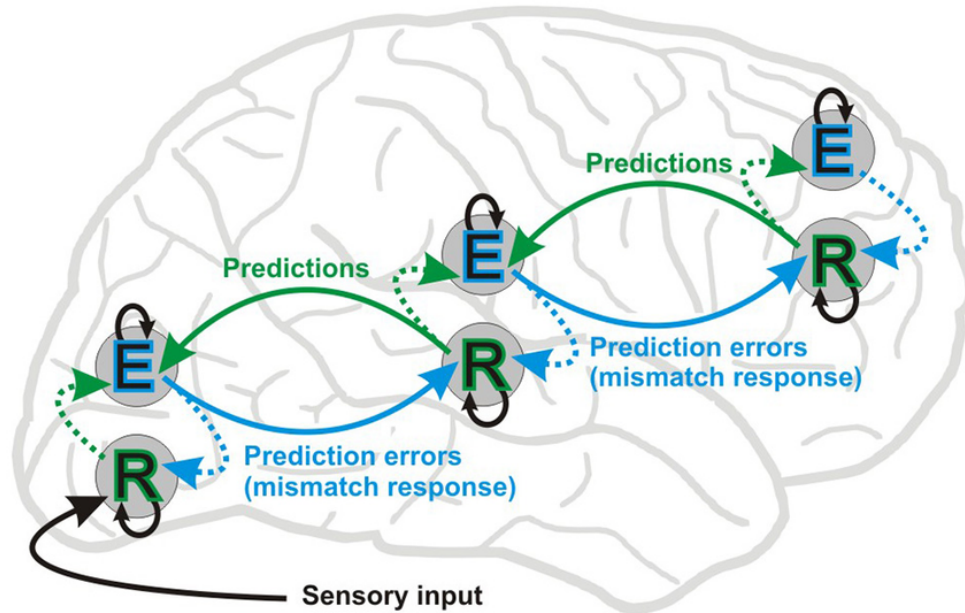
neural activity and synaptic weights across the cortical hierarchy (Sprevak & Smith, 2023). Predictions are produced at every level of the hierarchy and flow downward toward the sensory surfaces (Clark, 2013). At each level, they are compared to sensory signals arriving from the bottom up. Where a prediction matches the incoming signal, it is *explained away* by the model and further processing of the sensory evidence is suppressed (Clark, 2013; Hohwy, 2013). If, however, there is a discrepancy (i.e., variational free energy), this information propagates back upward through the hierarchy, where it drives new, revised predictions to be passed back down (Clark, 2013; Parr et al., 2022). As such, the information flow is bidirectional and recursive. The algorithmic theory most closely associated with PP is *predictive coding* (Rao & Ballard, 1999; Friston, 2005), although other algorithmic instantiations are possible within the broader framework (Sprevak & Smith, 2023). See Figure 2 for an illustration.

Note that in biological and neuroscientific contexts, where the focus is on explaining brain signals, the term *prediction error* is typically used instead of *variational free energy* (Parr et al., 2022, p. 25). The two terms are not strictly synonymous, but under specific conditions variational free energy mathematically reduces to precision-weighted prediction error, so the terms are often used interchangeably in applied PP work (Bogacz, 2017; Friston, 2005; Parr et al., 2022; Rao & Ballard, 1999).

PP and predictive coding find support in the biology of the brain itself. In the visual cortex, the number of feedback (top-down) connections substantially exceeds the number of feedforward (bottom-up) ones (Muckli & Petro, 2013). It should be noted, however, that the empirical status of PP remains unsettled. The overall evidence for the framework's key claims is mixed, and many findings interpreted as supporting it could also be accommodated by alternative models (Sprevak & Smith, 2023). Critics have argued that the theory's flexibility makes it difficult to falsify (Walsh et al., 2020). That said, clear-cut counterevidence has yet to emerge (Walsh et al., 2020), and the framework's core logic offers productive conceptual tools for understanding how the brain may operate, regardless of whether its precise neural implementation is fully confirmed.

Figure 2

Simplified scheme of the hierarchical predictive coding framework.



Note. The figure shows message passing between two neuronal populations in the hierarchical predictive-coding architecture: error units (E) and representation units (R). Representation units encode the brain’s current beliefs about the hidden states of the world and generate the predictions that are passed down (top-down or backward connections). Error units carry information about the discrepancy between those predictions and the signals arriving from below, and convey those prediction errors to higher levels of the hierarchy (bottom-up or forward connections). Reprinted from “Visual mismatch negativity: A predictive coding view,” by G. Stefanics, J. Kremláček, and I. Czigler, 2014, *Frontiers in Human Neuroscience*, 8, 666, p. 4 (<https://doi.org/10.3389/fnhum.2014.00666>). CC BY 4.0.

2.5 Action-Oriented Predictive Processing

Within the PP framework, action is treated as a further form of inference — what the literature calls *active inference*, or *action-oriented predictive processing* (Clark, 2013; Hohwy, 2013; Parr et al., 2022). Just as perception confronts the problem that the same sensory signals can be produced by many possible hidden causes, action confronts the problem that the same desired outcome can be brought about by many possible sequences of motor behaviour. The two problems are structurally similar, and both call for an inferential solution (Hohwy, 2020). In active inference, the system selects a *policy* — a sequence of sensorimotor states — by

inferring which policy is most likely to bring about the outcomes its generative model expects (Parr et al., 2022).

In the active-inference-framework literature, the quantity used to select between candidate policies is termed *expected free energy*. It is “expected” rather than “variational” because it is computed over anticipated future observations under a policy, before those observations have actually occurred — in contrast to variational free energy, which is computed over current observations (Friston et al., 2015; Parr et al., 2022).

Once a policy is selected, the system treats the sensory outcomes expected under that policy as if they were already occurring (Clark, 2013; Parr et al., 2022). The discrepancy between these expected outcomes and the actual sensory state then generates variational free energy, which is minimised by engaging the body to bring sensory input into alignment with the predictions. For example, suppose I form the prediction that I will be holding a cup of coffee. The prediction is, at present, contrary to fact: the cup is on the kitchen counter, and proprioception tells me my hand is empty. The mismatch generates prediction error, which is resolved by reaching toward the cup and grasping it — that is, by acting on the world to produce the sensory pattern that the prediction specifies (Clark, 2013; Hohwy, 2020). Perceptual inference and active inference are therefore tightly coupled: the latter depends on the former. If my model of where I left the cup is inaccurate, my action policy for retrieving it will be similarly imprecise (Hohwy, 2020).

2.5.1 Epistemic Value and Pragmatic Value Active Inference

Action under active inference can serve two distinct purposes (Friston et al., 2015; Parr et al., 2022). The first is *epistemic value*: action that reduces uncertainty about hidden states of the world, such as looking up to check whether the cup is still on the counter (Friston et al., 2015; Schwartenbeck et al., 2019). The second is *pragmatic value*: action that fulfills preferred outcomes, such as drinking the coffee to relieve thirst (Friston et al., 2015; Parr et al., 2022). Both types of action are formally unified within active inference, because both minimise expected free energy (Friston et al., 2015). From the framework’s point of view, then, there is no fundamental difference between exploration and goal-pursuit: the system simply selects

whichever action best minimises prediction error in the longer run. This framing also clarifies the dark-room objection introduced in §2.2.

As such, literature on the active inference framework distinguishes two functionally different kinds of priors: *priors over states* and *prior preferences*.

Priors over states are beliefs about what the world is — its structure, causal regularities, and the entities and relationships that compose it. They can be hardwired or acquired. Some appear to be present from birth: foundational expectations about object persistence, causality, spatial extension, and temporal continuity (Spelke & Kinzler, 2007). More sophisticated priors over states — such as beliefs about specific people, social roles, and how relationships tend to unfold — are learned through experience and become encoded at higher levels of the cortical hierarchy (McGovern & Otten, 2024). Higher-level priors generally operate over longer timescales and update more slowly than lower-level ones (Kiebel et al., 2008; Friston, 2008), which makes deeply held high-level beliefs particularly resistant to single contradicting observations.

On the other hand, *prior preferences* encode expectations about what sensory states an organism of this kind is built to experience — for example, the body temperature it expects to maintain, the oxygen saturation it expects to find, and the food it expects to consume (Friston, 2013; Friston et al., 2015). For this reason they are sometimes called *phenotype-defining priors*: they encode what kind of organism this is. In the active-inference extension of PP, prior preferences drive action. An organism does not merely passively update its beliefs about the world, it also acts to bring its sensory states into line with its prior preferences (Friston et al., 2015; Parr et al., 2022).

Prior preferences are not restricted to homeostatic variables. In humans, they include a wide range of expected sensory states with significant psychological content, including positive self-evaluation, social acceptance, secure attachment, predictability, and agency (Friston, 2013; Veissière et al., 2020; Fisher & Hohwy, 2024). Even if an interaction unfolded exactly the way an individual expected — meaning there is no variational free energy — discrepancy between the situation the individual experienced and its preferred states is itself a form of

surprise. Formally, it is the pragmatic-value component of expected free energy (Friston et al., 2015). An agent's actions are selected over time to reduce that discrepancy (Friston et al., 2015).

2.5.2 Counterfactual Processing

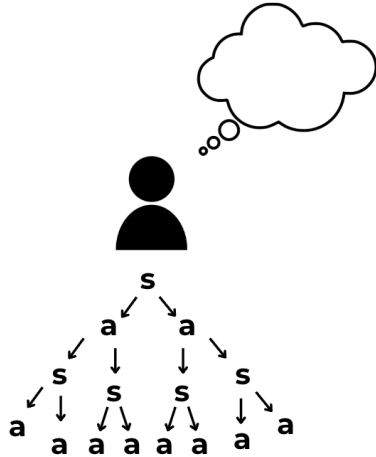
A central problem for active inference is determining which action the agent should take (Parr et al., 2022). Action selection requires comparing alternatives whose future consequences have not yet been observed, and an action that reduces expected free energy in the immediate moment may not do so over a longer horizon (Lehmann et al., 2024). In active inference, this problem is solved through *counterfactual processing*: imagining what sensory input would arise if a particular action were executed, before any action is actually taken (Corcoran et al., 2020; Friston, 2018; Lehmann et al., 2024; Palmer et al., 2015).

Counterfactual processing is not limited to a single step ahead: the generative model can also project the actions that would become available after a given sensory consequence, the sensory input that would in turn follow from those actions, and so on (Friston, 2018). Action selection therefore proceeds through what is called a *deep tree search*, in which each branch represents a candidate action (or policy) and the expected sensory consequences it would generate (Friston et al., 2021). The agent compares the expected free energy of each branching trajectory and selects the action whose downstream consequences best match its prior preferences (Lehmann et al., 2024; Parr et al., 2022). For example, a glance at my watch yields sensory data about the current time; depending on that data, different subsequent actions become available — continuing the current activity, preparing to leave the house, or sending a message that I will be late. Figure 3 illustrates this counterfactual structure.

The depth and richness of an agent's counterfactual processing — how far ahead the generative model projects, and how many alternative trajectories it can entertain — has been argued to underlie the perceptual sense that the world (and other minds within it) is genuinely present rather than merely arriving as data (Seth, 2014; Palmer et al., 2015), and more broadly to be the defining feature of cognition itself, distinguishing cognitive agents from agents that only regulate their immediate physiological state (Corcoran et al., 2020).

Figure 3

A single-agent model of counterfactual processing



Note. The figure depicts the deep-tree structure of counterfactual processing in a single-agent setting. The symbol *a* represents candidate actions, and *s* represents the hidden states inferred under each candidate action. Adapted from “An active-inference approach to second-person neuroscience” (Figure 2a), by K. Lehmann, D. Bolis, K. J. Friston, L. Schilbach, M. J. D. Ramstead, and P. Kanske, 2024, *Perspectives on Psychological Science*, 19(6), p. 935 (<https://doi.org/10.1177/17456916231188000>). CC BY 4.0.

In its more advanced form, counterfactual processing extends to sophisticated inference: action selection in which the agent considers not only the expected sensory consequences of an action but also how that action will modify its own and others’ future beliefs (Friston et al., 2021; Hesp et al., 2021). Sophisticated inference becomes especially important in social contexts, where one agent’s action is simultaneously a sensory input for another agent, who is in turn modeling the first (Friston & Frith, 2015; Lehmann et al., 2024). The recursive structure of counterfactual processing therefore becomes nested across agents, with policy selection on each side informed by anticipated beliefs on the other (Lehmann et al., 2024).

2.6 Mechanisms of Prediction Error Minimisation

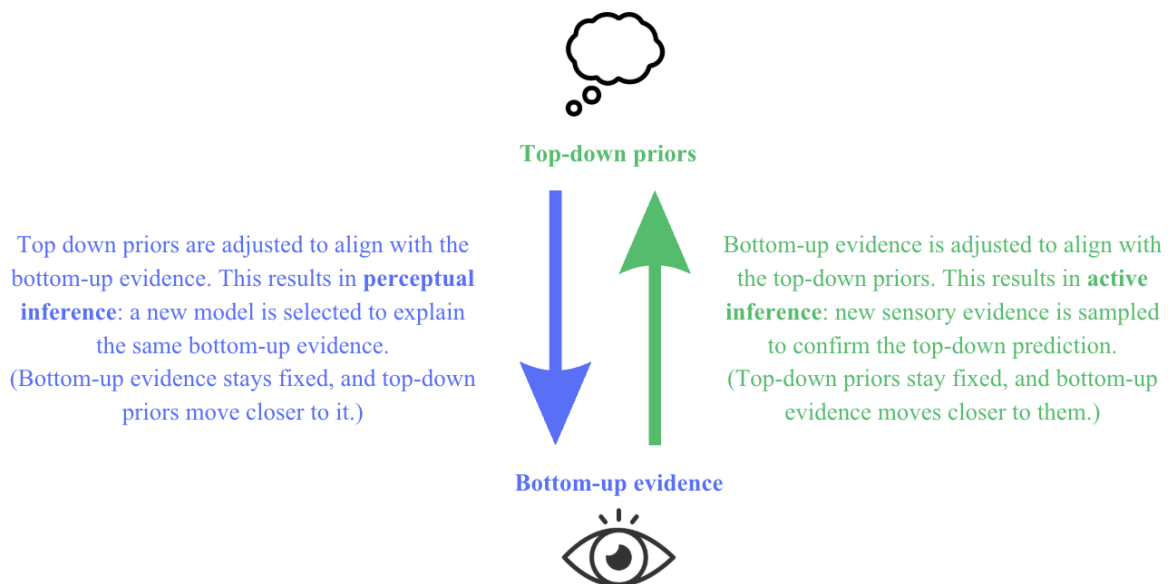
2.6.1 Perceptual Inference and Active Inference

When there is a discrepancy between a generative model prediction and sensory data, it can be resolved in two ways: the model predictions can be revised, or the observed data can be changed.

The former would mean revising the agent's priors so that the model produces the correct prediction — called *perceptual inference*. The latter would mean acting on the world to sample data that better matches the priors and is less surprising under the model — called active inference. As Pezzulo et al. (2024) explain, “‘action’ is a suitcase word and can include any means to exert control over external stimuli; ranging from autonomic reflexes (e.g., thermoregulation) to sophisticated plans (e.g., visiting one’s favorite ice cream shop)” (p.2) Under this framework, perception and action do not have two distinct objectives, but rather optimize the same function: minimizing free energy. Figure 4 offers an illustration of this.

Figure 4

A comparison of the two ways prediction error is minimised



Let us see an example for each kind of inference.

Perceptual inference: Suppose I am walking along a forest trail and notice something long, thin, and coiled in the leaves ahead. My brain's generative model entertains several possible hidden causes for the visual input — snake, rope, branch — and at first glance the prior over hidden states favours snake, because the surrounding context and the ambiguous shape together raise the prior probability of that hypothesis. The posterior tips toward snake, and my conscious percept is of a snake on the path. However, as I look more closely, the texture appears fibrous rather than scaled, the object does not move, and one end is visibly frayed. Each of these observations is unexpected under the snake hypothesis and so generates prediction error. Because the new evidence is precise (well-lit, sustained, foveal) and the snake-prior is not unrevisable, the prediction errors update the posterior: the rope hypothesis now best explains the data, and the percept flips to rope (Clark, 2013; Parr et al., 2022). Crucially, nothing in the world has changed; only my brain's inference about the hidden cause has.

Active inference: Consider what happens when my body temperature drops below its expected range. My brain maintains a prior preference (in the sense established in §2.5) that body temperature should be approximately 37°C — a preference held with extremely high precision, because deviations are detrimental to survival (Seth & Friston, 2016). If my interoceptive receptors signal that body temperature has dropped to 35°C , my brain does not simply update its generative model to expect this lower temperature. Instead, it triggers action: at the most basic level, this involves autonomic reflexes such as shivering and vasoconstriction, but it can also extend to deliberate actions such as putting on warmer clothing, seeking a heat source, or moving around. These actions all serve to bring actual sensory input into alignment with the prior preference (Seth & Friston, 2016; Tschantz et al., 2022).

As can be seen from the examples, each type of inference is more suitable for some scenarios than others. In the body temperature example, it would be lethal to use perceptual inference and update the model to expect body temperature to be 35°C . The case for the forest walking example — turning my head and looking around until I find sensory evidence of a snake — is

slightly more logical from the point of view of threat detection, but generally not a good strategy for accurate perception of the environment around me.

In reality, these two types of inference are not deployed in isolation, but in a continuous perception-action loop in which the organism and the environment exchange information reciprocally. Parr et al. (2022) illustrate this:

“Essentially, for each action-perception cycle, the environment sends an observation to the organism. The organism uses (an approximation to) Bayesian inference to infer its most likely hidden states. It then generates an action and sends it to the environment in an attempt to make the environment less surprising. The environment executes the action, generates a new observation, and sends it to the organism. Then, a new cycle starts. The sequential description here is written for didactic purposes; it is important to realize that these are not really discrete steps but are continuous dynamical processes” (p. 24).

2.6.2 Precision Weighting

What determines whether the brain updates its model to fit the sensory data (perceptual inference) or acts on the world to fit the model (active inference)? The answer lies in *precision*: the relative weight assigned to top-down predictions versus bottom-up sensory evidence (Clark, 2013; Feldman & Friston, 2010). When sensory data carries greater weight than the prior, the posterior shifts toward the data and perceptual inference or learning follows; when the prior carries greater weight, the posterior remains close to what was expected, and mismatch can be resolved either by ignoring the data or by active inference that brings the data into line.

Precision is not a fixed property of either side; it is itself estimated by the brain, and the same prior or signal can be assigned different precision in different contexts. Attention is thought to be one of the mechanisms through which precision is modulated (Feldman & Friston, 2010). On this view, paying attention to a stimulus is the experiential and behavioural correlate of the brain assigning high precision to the prediction errors arising from that channel, so that those errors carry greater weight in belief updating (Clark, 2013; Hohwy, 2013). A standard

illustration is a sailor moving through dense fog: visual information becomes unreliable, the brain assigns higher precision to auditory input, and what is described as “listening carefully” is the experiential side of that precision shift (Chamberlin, 2019).

Several forms of psychopathology have been analyzed in terms of disturbed precision weighting. In schizophrenia, both over-precise high-level priors and under-precise sensory or high-level signals have been invoked to explain different symptom clusters — over-precise priors to account for delusions, and disturbed precision profiles more generally to account for the failure of certain perceptual illusions in patients (Adams et al., 2013; Sterzer et al., 2018). Autism has been described in terms of abnormally high precision on sensory input relative to top-down priors, so the world is experienced as unpredictable and over-detailed and small changes generate large prediction errors (Pellicano & Burr, 2012). Anxiety and post-traumatic stress disorder have been framed as cases of over-precise priors over threat: once formed, the trust-no-one or fear-this-stimulus prior is held with such confidence that even sustained safe evidence cannot revise it, and avoidance behaviour then prevents the agent from gathering the data that would otherwise generate the disconfirming prediction error (Wilkinson et al., 2017; McGovern et al., 2022).

The explanatory reach of precision has also drawn criticism. Because precision weights are not directly observable and are typically inferred post hoc from the very behaviour they are meant to explain, several authors have argued that precision functions as a “magic modulator” in the framework: almost any pattern of perception or action can be modeled as prediction-error minimization provided one assigns the agent a suitable set of precision weights (Miller & Clark, 2018; Litwin & Miłkowski, 2020). The objection does not refute the framework but it does discipline its use: precision is doing real explanatory work only when the precision profile attributed to the agent is independently motivated rather than retrofitted to the data.

2.6.3 Learning

The preceding sections described how the brain minimises prediction error in the moment, through either perceptual inference or active inference. However, neither one, by itself,

changes the generative model: perceptual inference shifts the moment-to-moment posterior, and active inference selects actions, but neither alters the parameters that determine what predictions the brain will make next time. If a few minutes later, I were to see another vague, long object on the floor, I may make the same prediction of a snake again.

When does learning occur? It is through repeated prediction errors gradually reshaping the parameters of the generative model itself. This can arise from either route — perceptual or active — and is implemented at the neural level through synaptic plasticity.

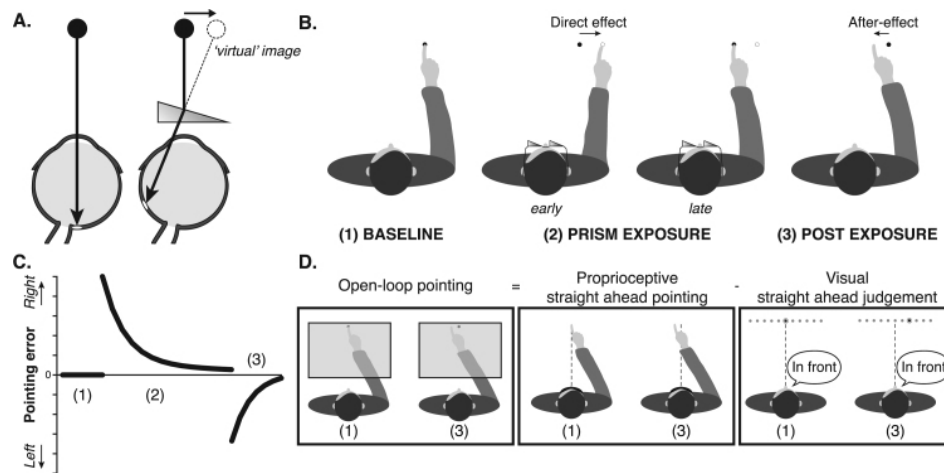
Learning through perception. Suppose I repeatedly encounter ambiguous long, coiled shapes on the ground while walking, and each time a closer look reveals a rope rather than a snake. Over many such encounters, the parameters of the generative model itself begin to shift — the prior probability of “snake” given the visual category “vague long object in this environment” is slowly downweighted in favour of “rope” (Friston & Kiebel, 2009). The same learning can be reinforced by sensory signals arriving through other channels: conversations with locals, a nature documentary, or a biologist who explains that there are no snakes in the region. In hierarchical PP, this propositional information updates higher-level beliefs about what kinds of animals populate the environment, which in turn descend as empirical priors that constrain low-level visual interpretation (Lupyan, 2015). The next time the observer sees a long ambiguous object on the path, the prior over “snake” is much weaker than before, and the percept settles on “rope” more readily.

Learning through action. A classic paradigm for studying sensorimotor learning is prism adaptation, first described by Helmholtz (1867/1962). When someone first puts on prism glasses that shift the visual field by 10° to the right, reaching for objects produces large errors — the hand appears somewhere unexpected and does not succeed in grasping the object. Initially, conscious correction may help compensate. With repeated reaching movements, however, the brain gradually updates the parameters of its generative model of the relationship between motor commands and visual feedback (Redding et al., 2005). The model itself has changed, rather than just moment-to-moment perceptual estimates, as evidenced by the aftereffect: when the prism glasses are removed, reaches now err in the opposite direction,

demonstrating that the underlying visuomotor mapping has been durably recalibrated (Redding et al., 2005; Helmholtz, 1867/1962). See Figure 5.

Figure 5

The prism glasses adaptation, showing learning through prediction error



Note. The figure depicts the three phases of prism adaptation: a baseline phase in which the participant reaches accurately for a target, an exposure phase in which prism glasses laterally shift the visual field and reaching errors are progressively corrected over trials, and a removal phase in which the prism glasses are taken off and the participant exhibits an after-effect — reaching errors in the opposite direction — that demonstrates durable recalibration of the underlying visuomotor mapping. Reprinted from “Towards a neuro-computational account of prism adaptation” (Figure 1), by P. Petitot, J. X. O’Reilly, and J. O’Shea, 2018, *Neuropsychologia*, 115, p. 189 (<https://doi.org/10.1016/j.neuropsychologia.2017.12.021>). CC BY 4.0.

2.6.4 False Inference

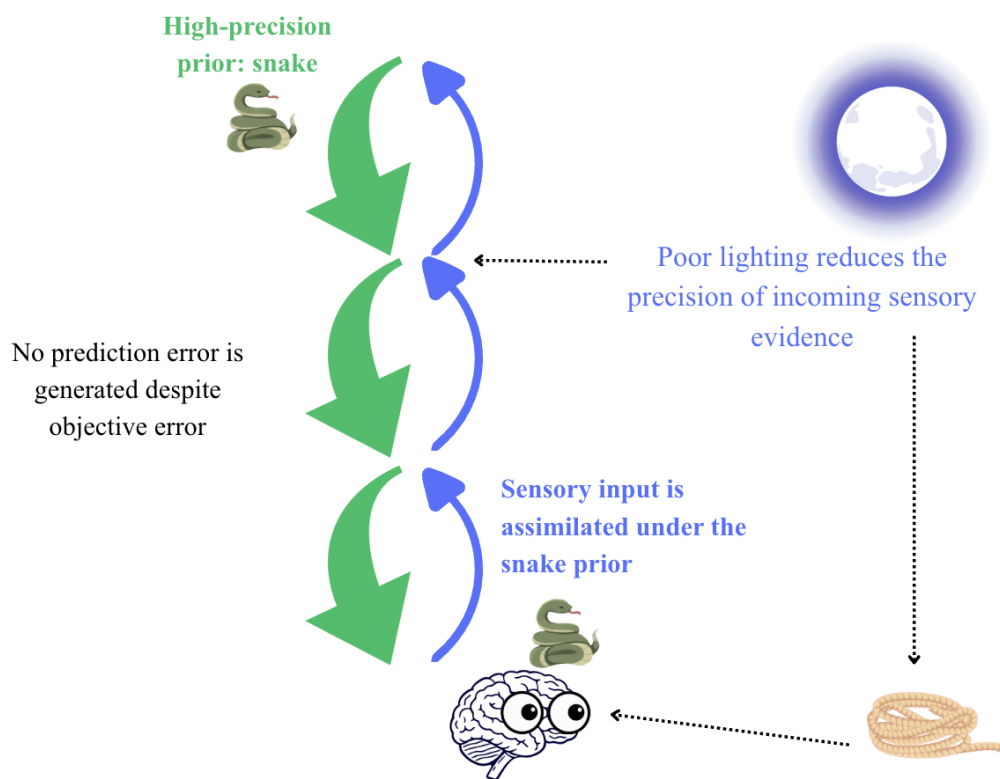
The preceding sections described how prediction error can arise and be resolved. However, one important case remains: when prediction error should occur, but does not.

Crucially, prediction error is not computed against the world as it objectively is, but against the sensory signals the brain receives. If a model predicts incorrectly, but the sensory evidence nonetheless can fit under the prediction, no prediction error will be generated, and no learning or revision will occur. Consider the rope-in-the-forest example again, illustrated in Figure 6. If lighting is poor, vision is impaired, or the object is not seen clearly, the model may settle on

the posterior belief that what was observed was, in fact, a snake. Although this would be classified as an error by objective standards, it would not generate prediction error within the model itself. On the contrary, the model will be confirmed and perhaps even reinforced. This kind of inference, in which the model selects the best-fitting posterior available to it even though that fit is poor by objective standards, is termed *false inference* (Hohwy, 2013).

Figure 6

Illustration of false inference under poor lighting conditions



Note. Original figure illustrating how the snake-vs-rope perceptual inference may settle on the snake posterior when sensory evidence is impoverished, generating no prediction error to drive revision. Created in Canva.

Importantly, false inference may occur even when sensory evidence is high in precision (foveal, well-lit, attentionally engaged). This happens when priors are so strong that the brain assimilates incoming sensory evidence under them rather than allowing the evidence to revise the model.

Example 1: The dress. A striking example from visual perception is the famous photograph of “the dress” that went viral in 2015, which some observers perceived as white and gold while others saw it as blue and black (see Figure 7a). Research suggests that these differences arose from observers’ different prior assumptions about the illumination in the scene: those who assumed a cool illuminant (like blue sky) discounted shorter wavelengths and perceived white/gold, while those who assumed a warm illuminant (like incandescent light) discounted longer wavelengths and saw blue/black (Lafer-Sousa et al., 2015; Witzel et al., 2017). When observers were shown disambiguated versions of the image with clear lighting cues — such as the dress obviously in shadow or in direct sunlight — their perception shifted predictably, demonstrating that manipulating the contextual information about illumination could “flip” the perceived colors (Witzel et al., 2017). (See Figures 7b and 7c below, covering the other image to reduce cross-influence.)

Figure 7



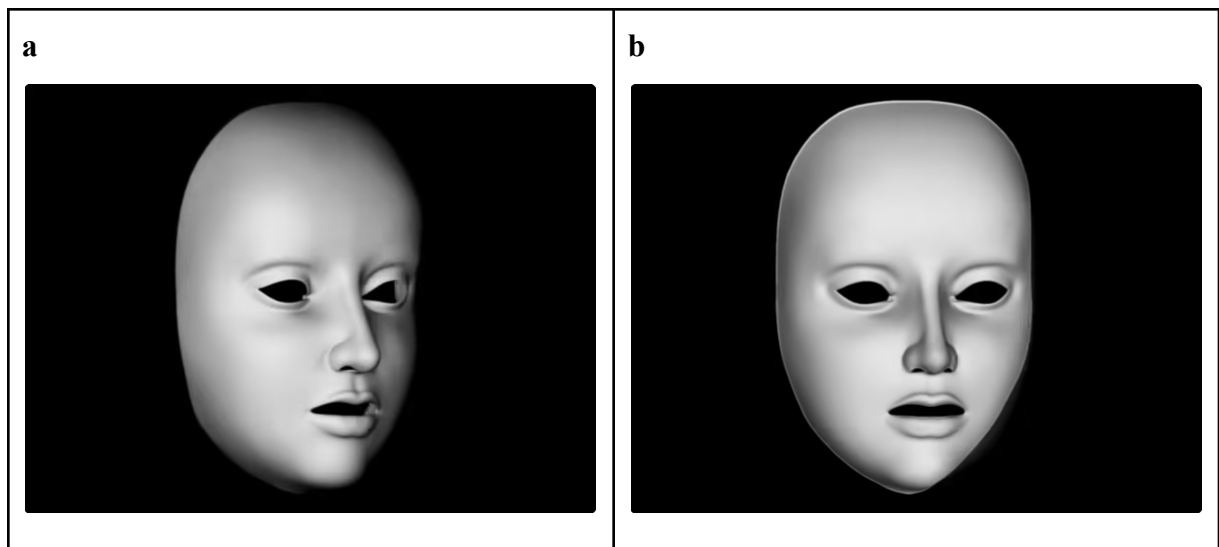
Note. (a) The dress, which many people perceived as white and gold and many others as blue and black. (b) A snippet from the dress placed in the context of a gray autumn day. (c) A snippet from the dress placed in the context of a bright spring evening. Screenshots from *How your brain predictions interfere with what you see* [Video], by G. Keller, 2017, October 18, YouTube (<https://www.youtube.com/watch?v=RgWgfdLZciM>).

Example 2: The hollow mask illusion. Another example of false inference is the hollow mask illusion, shown in Figure 8 below. This is often cited as a leading demonstration of the role of prior expectations in early vision (Gregory, 1970; Keane et al., 2016). When a convex

mask rotates to reveal its hollow interior, the visual input signals a concave face. However, the brain holds an extremely strong prior that faces are always convex — a prediction built from a lifetime of experience. Because this prior carries such high precision, the brain does not update its generative model in response to the ambiguous visual evidence. Instead, the high-precision top-down prediction dominates the lower-precision sensory input: the face appears convex (perhaps with unusual lighting or shadows) rather than concave (Dima et al., 2009; Keane et al., 2013; Schneider et al., 2002). A video demonstration is available at <https://www.youtube.com/watch?v=sKa0eaKsdA0>. Notably, individuals with schizophrenia, who may have reduced precision weighting of top-down predictions, are often less susceptible to this illusion and perceive the mask more veridically as concave (Dima et al., 2009; Keane et al., 2013; Schneider et al., 2002).

Figure 8

The hollow mask illusion



Note. (a) The mask is convex (curving outward like a normal face). (b) The mask is concave (curving inward like the inside of a bowl). Screenshots from *The rotating mask illusion* [Video], by eChalk, 2012, YouTube (<https://www.youtube.com/watch?v=sKa0eaKsdA0>).

These principles of top-down constraint are not limited to low-level visual perception. They have been documented across multiple domains. Pain perception can be generated by predictive models without tissue damage, as in the well-known case of a builder who

experienced excruciating pain from a nail piercing his boot, despite the nail having passed harmlessly between his toes (Clark, 2023; Fisher et al., 1995). Taste and product evaluations are reliably shaped by labels and contextual cues (e.g., Allison & Uhl, 1964; McClure et al., 2004), and Lee et al. (2006) showed that this influence can alter the taste experience itself, not merely retrospective judgment.

Particularly relevant to the present work, analogous effects appear in social perception. Prior beliefs about a person have long been shown to shape how that person is perceived: identical lecturers are rated differently depending on whether they are described as warm or cold (Kelley, 1950), and identical academic performance is evaluated differently as a function of the perceived socioeconomic background of the child (Darley & Gross, 1983). More recent neuroimaging work suggests these effects extend to the underlying perceptual representations themselves: stereotypes warp face processing in the fusiform cortex (Stolier & Freeman, 2016), conceptual knowledge about emotions shapes how emotional faces are perceived (Brooks & Freeman, 2018), and individual person-knowledge alters how a face is perceptually represented rather than merely judged (Oh et al., 2021).

False inference is particularly relevant to the present thesis, because strong priors are a candidate mechanism by which a patient may fail to update their model even when contradictory evidence is available, and remain locked into unhelpful thinking patterns.

2.7 Multi-Agent Predictive Processing

The preceding sections have described how PP operates within a single brain. Yet, unless one subscribes to solipsism, little of human life is spent in isolation. Most of it unfolds in interaction with other agents, each of whom, on the predictive-processing account, is running their own generative model of the world.

This raises two interlocking problems for the framework. First, if every agent is running a separate generative model, how do agents come to predict each other at all — how does one agent’s model of the other align with what the other is in fact doing? Second, even given some form of mutual prediction, how does any one agent plan and select actions in an environment populated by other inferring agents whose responses must themselves be anticipated? The

predictive-processing literature has engaged with both questions over the past decade. The first has been answered through the proposal that agents can come to share a generative model; the second through the extension of counterfactual processing into a nested, dyadic form.

2.7.1 Shared Generative Models

The most direct response in the predictive-processing literature to the multi-agent problem is the proposal that two or more brains can come to share a generative model through ongoing communication. The canonical statement of this position is Friston and Frith (2015), which addresses the apparent infinite regress that arises when each agent’s generative model has to include a model of the other agent’s model: agent A modeling B modeling A, and so on. Friston and Frith show that this regress dissolves when both agents are formally similar and use the same generative model to predict both their own and the other’s behaviour. Under these conditions, the agents’ internal states couple through reciprocal sensory exchange and enter what is mathematically termed *generalized synchronization*, in which the state of one agent reliably predicts the state of the other. The image they offer is of a duet: a song that “does not belong to you or me — agency just contextualises its expression” (Friston & Frith, 2015, p. 401). On this account, successful communication is not the transfer of information from a privately modeled speaker to a privately modeled listener, but the emergence of a shared dynamical structure that both agents participate in producing.

Subsequent work has developed this framework along several axes. Vasil et al. (2020) apply the framework specifically to cooperative communication, proposing that humans hold an evolved adaptive prior that their mental states are aligned with those of conspecifics, and that cooperative communication is the principal means by which evidence for this belief is gathered and a shared narrative is established. Veissière et al. (2020) introduce a variational account of culture they call *Thinking through Other Minds*: human agents acquire shared norms and expectations through immersive participation in patterned cultural practices, and information about other people’s expectations becomes a primary statistical regularity that humans use to predict and organize behaviour. Constant et al. (2019) model social conformity as the result of agents operating under such shared *regimes of expectations*. Bolis and

Schilbach (2018) turn the framework toward selfhood, arguing that the self is not a stable individual entity but is itself a historical product of repeated *dialectical attunement* with others over time. Lehmann et al. (2024) frame nested counterfactual processing in dyadic contexts as the mechanism by which one agent’s policy selection becomes informed by anticipated beliefs on the other side.

A formal extension of the dyadic case to multiple agents is offered by Friston et al. (2024), whose account of *federated inference* shows that multiple agents who broadcast their beliefs to one another can carry out inference collectively, producing a shared posterior that is more precise than any individual posterior. Thestrup Waade et al. (2025) go further still, formalizing the conditions under which a collective of active-inference agents (in the sense of the active inference framework) can itself be treated as a single inferring system at a higher level of organization — by virtue of maintaining a group-level *Markov blanket*. On this reading, the question of whether a family or other group has its own generative model is not a metaphor but a formal claim about the system’s organization.

2.7.2 Counterfactual Processing in Multi-Agent Contexts

In the single-agent case, the generative model simulates possible actions and their downstream consequences in order to select the action whose expected sensory trajectory best matches the agent’s prior preferences. The deep tree of exploration is already dense in the single-agent case. When multiple agents are added, its structure becomes substantially more complex.

In a multi-agent system, the agent must also infer, for each possible action, the mental states that action would produce in the other agent and the actions the other agent would in turn select in response. The deep tree thereby becomes *nested*: each candidate action generates not only a downward branch of expected sensory consequences for the acting agent but also a lateral branch of expected sensory consequences for the partner (Lehmann et al., 2024). For example, an agent who glances at their watch during a conversation generates predictions not only about their own subsequent sensations and actions but also about how the partner might receive that gesture — whether the partner fails to notice, notices and is reminded of their own time commitments, or notices and infers that the agent is disengaged. Each of these

counterfactually imagined sensory states for the partner has its own downward predictions about the actions the partner would then take: continuing the conversation, checking their own watch, or commenting on the perceived disinterest.

These kind of lateral partner-state predictions occur for each node in the original tree. A different candidate action — for instance, resisting the urge to glance at the watch and continuing to engage with the conversation — generates a different cascade of predictions about both the agent’s own state (e.g., increasing impatience, unconscious behavioural signals) and the partner’s response to that state (e.g. the partner notices and feels empathetic, notices and feels offended, or fails to notice).

The structure becomes more complex still when the partner is treated, accurately, as a similarly active inferring agent. The agent’s generative model must include not only what the partner perceives and how the partner is likely to act, but also what the partner believes the agent is perceiving, what the partner predicts the agent will do, and what the partner is inferring about the agent’s inferences. Lehmann et al. (2024) describe this as the *nested recursive* structure of policy selection in dyadic contexts: each agent’s policy selection is informed by the agent’s beliefs about the partner’s beliefs about the agent — and so on, in principle without limit.

The model is also continuously revised. Social interactions are inherently dynamic: topics shift, the partner’s beliefs evolve with each exchange, and the action opportunities available at one moment may not be those available at the next. An utterance that would have been relevant a moment earlier may no longer fit after the conversation has taken an unexpected turn. Moreover, each action contributes to what Lehmann et al. (2024) call *mutual historicity*: the shared interactional history that constrains subsequent possibilities. A strong opinion voiced earlier in a conversation, for example, shapes how every subsequent remark on the same topic is heard, even if the speaker’s view has since softened. Mutual historicity in this sense is closely related to what Bolis and Schilbach (2018) call *dialectical attunement* — the accumulated history through which two interacting agents come to constrain one another’s expectations over time. The deep tree of counterfactual exploration is therefore continuously

recomputed as the interaction unfolds, with each branch's expected outcomes sensitive to the trajectory the interaction has just taken.

2.8 Predictive-Processing Accounts of Other Psychotherapies

Over the past decade, the predictive-processing framework has been applied to a growing number of psychotherapy modalities. This section reviews the main contributions and identifies the unifying themes that emerge across them, in order to position the present analysis of Milan systemic therapy against the existing literature.

Integrative frameworks. Several authors have proposed PP as an organizing framework for psychotherapy as a whole. Holmes and Nolte (2019) frame psychopathology as a failure of free-energy minimization in one of two forms — unbound energy in trauma and agency-inhibition, or rigid top-down narratives — and treat psychotherapy as the restoration of agency and the loosening of those narratives so that experience can again be assimilated into the patient's generative model. Krupnik (2019) proposes *Strategic Modification of Priors* (SMOP) as a fifth integrative strategy for psychotherapy alongside the four classical strategies described by Norcross (2005): on this view, what unifies different modalities is that they all modify the patient's priors, and they differ in how they do so.

Therapeutic alliance. A related strand of work focuses on the *alliance* as the relational mechanism through which prior update becomes possible. Krupnik (2023) frames the alliance as an active-inference strategy (in the framework sense) that decreases the free energy of each participant's generative model of the other and of the social exchange, with *trust* and *self-efficacy* as the central variables. Hauke and Lohr (2022) propose that questioning a long-held prior is itself a short-term free-energy increase — a stress event — and that the alliance's mechanistic role is to *buffer* this increase, raising perceived safety so that the patient can remain in the exploratory mode long enough for the prior to update. Campbell et al. (2024) extend this with the argument that the active ingredient shared across psychotherapies is the restoration of *epistemic trust* — the trust that licenses social information to be taken up as personally relevant. On their account, adversity disrupts the developmental capacity for

mentalizing and produces *epistemic disruption*; psychotherapy reopens the channel by which social information can again become a prior.

Modality-specific analyses. Several papers map specific therapy modalities onto PP mechanisms in detail. Chamberlin (2019) offers a PP model of Eye Movement Desensitization and Reprocessing (EMDR), in which the therapy’s procedural elements drive prediction error against the predictions generated by the traumatic memory, triggering memory reconsolidation and updating the memory. Vanderschoot and Van Dessel (2022) complement this with a goal-directed PP analysis of EMDR for PTSD. Chamberlin (2023) extends his framework to Coherence Therapy, introducing the concept of *Bayes Optimal Pathology* — symptoms as the optimal output of inference given a suboptimal prior — and treating therapy as a dyadic act of therapist-guided active inference (in the framework sense) that renders the implicit prior conscious and juxtaposes it with contradictory experience.

Hauke and Lohr (2022) and Smith et al. (2021) apply the framework to cognitive-behavioural therapy (CBT). Berwian et al. (2025) propose a complementary computational account that combines reinforcement learning with Bayesian latent-cause inference, suggesting in particular that exposure therapy may be best understood not as the overwriting of fear associations but as the inference of a new, safe latent context in which the feared stimulus no longer signals threat. Fujimoto (2026) proposes a four-mechanism PP framework for body-oriented trauma intervention, in which interoceptive attention raises the precision of bodily signals, somatic safety experiences generate prediction errors against threat priors, eye-movement interventions modulate subcortical defense circuits, and *letting-go* practices reduce excessive prior precision on threat predictions.

Cross-cutting themes. Despite their modality-specific focus, these contributions share a common structure. Each conceptualizes the target pathology as a prior that has become over-precise, miscalibrated, or fitted to a suboptimal model of the world; each describes therapy as a procedure for surfacing that prior and generating sufficient prediction error to update it; and each treats the therapeutic relationship as the relational scaffold that enables this update. Memory reconsolidation, precision modulation, and the restoration of openness to new

evidence appear repeatedly as candidate mechanisms across modalities (Ecker, 2015; Hauke & Lohr, 2022; Krupnik, 2023).

3. Foundations of Milan Systemic Therapy

Systemic therapy is the product of a rich and continuously evolving tradition, with many different theoretical influences from different researchers and fields having been added, adapted, and subtracted from it over time. New variants have emerged repeatedly as practitioners have branched off and developed their own versions of the systemic approach. The result is a field with a substantial vocabulary and a layered conceptual history, which a reader unfamiliar with the literature may find difficult to navigate.

The clearest entry point into Milan systemic therapy (MST), in the author's experience, is the *story* of its evolution: how and why successive ideas became integrated into it, where those ideas came from, and which theoretical schools joined or departed from the systemic tradition. A summary of that history is provided in Appendix A and may be useful to read first for any reader unfamiliar with the field. The appendix focuses on the narrative arc of systemic thinking — its societal and cultural context, the principal researchers involved, and the chronological order in which key ideas entered the tradition.

The present chapter turns to the conceptual content of MST itself. §3.1 introduces two case studies from the Milan team that serve as running examples throughout. §3.2 and §3.3 then examine the theoretical concepts that organize MST, and §3.4 examines the methodology — the specific techniques through which those concepts are enacted in clinical practice.

3.1 Case Studies

Philosophical and abstract concepts can be intellectually interesting yet difficult to picture in practice. To make the concepts central to MST more concrete, this thesis uses two case studies as running illustrations and points out specifically how each concept and technique appears in them.

As Appendix A shows, systemic therapy was in continuous evolution, and the concepts themselves were progressively refined. For this reason, the case studies presented have been chosen from different periods of the tradition: one from the late 1970s and one from the early 1990s. Working with two cases rather than one carries the cost of slightly higher cognitive load for the reader, but it allows the full range of MST concepts to be seen across more than one clinical problem and at more than one stage of the tradition's development.

3.1.1 John, “The Crying Boy” (late 1970s)

This case is the first case study presented in *Milan Systemic Family Therapy* (Boscolo et al., 1987), under the heading “The Crying Boy.” It concerns a family interviewed by the original Milan team during a workshop in an English-speaking setting, as a one-off teaching consultation requested by the psychologist who had been treating the family. The transcript and the authors' accompanying reflections offer unusually detailed access to the Milan team's reasoning.

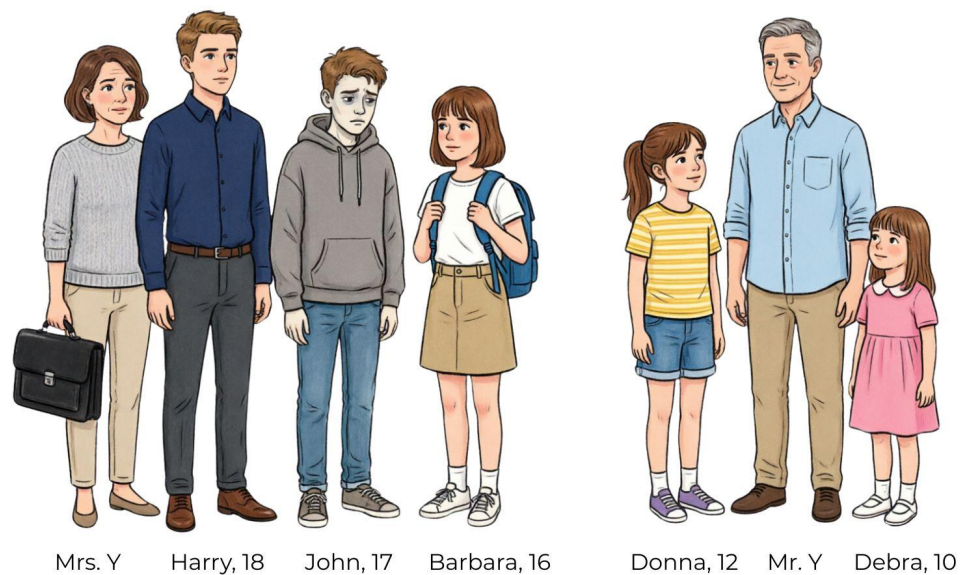
The Y family is a blended family. The parents, in their mid-forties, are five years into their second marriage. Mrs. Y has three children from her first marriage: Harry (18), John (17), and Barbara (16). Mr. Y has two daughters from his first marriage: Donna (12) and Deborah (10). An illustration of the family composition is provided in Figure 9 to aid the reader. Mrs. Y's first husband was alcoholic and violent, and had mistreated his children; Mrs. Y herself was deeply affected by that first marriage, and had been at one point suicidal. Mr. Y had also had a difficult first marriage, though he emerged from it less damaged, and in the new family he has taken on a strongly patriarchal role.

The family came to the clinic because of John, the seventeen-year-old. Two weeks before the consultation, John had been discharged from a hospital after a two-week admission for what the emergency-room physician had called “*an acute schizophrenic attack.*” The episode occurred on the first day of the school year. There was some indication of drug use — possibly cannabis, possibly LSD — and the treating psychiatrist had refused to speak with the family. The reasons for this refusal were never clarified; the team's intake suggested two possibilities: that John may have asked the doctor to keep the drug use confidential from his parents, or that

the doctor had concluded the family was “putting too much pressure on the boy to succeed” and was making the situation worse. The parents had therefore been left without an understanding of the problem or what behaviours to expect from John. Since coming home, John had been depressed and anxious, sleeping through the afternoons. Mrs. Y had called the clinic feeling confused and uncertain, and wanted primarily to know whether John was schizophrenic, and whether or not he should be medicated.

Figure 9

An illustration of John's family.



Note. Original figure, created with Qwen AI.

The outcome of this case is unknown. The authors report that the family did not return after the consultation, leaving the team to assume that the family “had handled the crisis without further therapy” (Boscolo et al., 1987, p. 90). The case is presented not as evidence of the success of systemic practices but as an illustration of their principles; however, the reader should know that very similar interventions have been documented to be strongly successful, one of which is mentioned in the beginning of the same book (Boscolo et al., 1987). As one suggestive marker, the team notes that John hugged one of the therapists after the session ended.

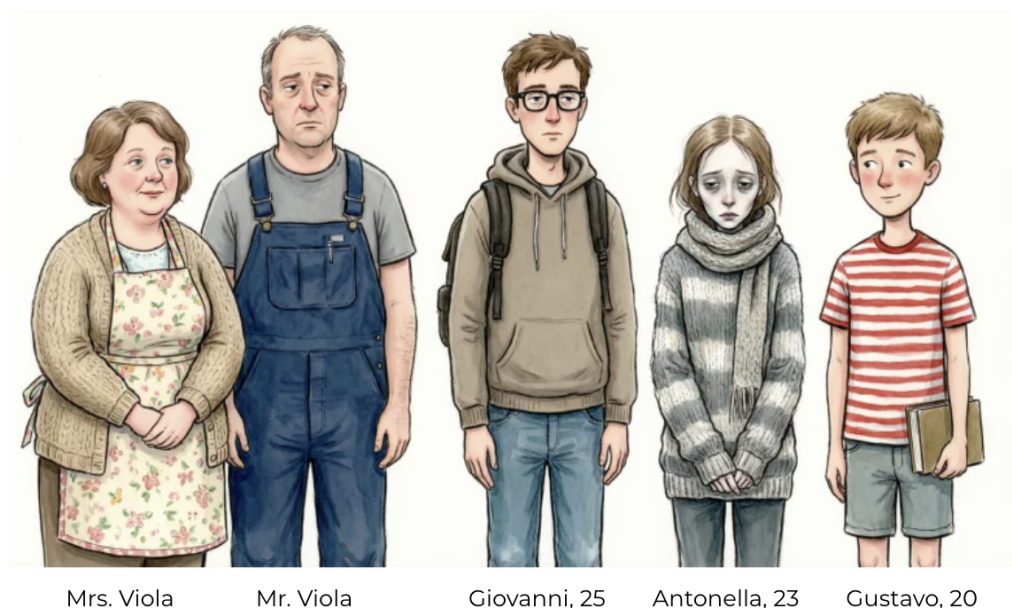
3.1.2 Antonella, “The Sweater Girl” (early 1990s)

This is the second of the three clinical cases reported in detail in *The Times of Time* (Boscolo & Bertrando, 1993, ch. 9). This was a full course of clinical therapy conducted at the Milan centre over six sessions.

Mr. Viola is a 54-year-old artisan; Mrs. Viola is a 48-year-old housewife. They have three children: Giovanni (25), an engineering student; Antonella (23), studying philosophy; and Gustavo (20), studying commerce and economics. An illustration of the family composition is provided in Figure 10. The family had been referred to the centre by the psychiatrist who had been treating Antonella for three years.

Figure 10

An illustration of Antonella’s family.



Note. Original figure, created with Qwen AI.

The family came to the centre because of Antonella, the middle child. Antonella’s problems had begun three years earlier with painful gastric symptoms — burning, nausea, heaviness, and swelling — followed by a decision that she was overweight and the development of an

anorexic syndrome. After the stomach disorders cleared up with hospital treatment, a series of phobias developed: fear of food, fear of losing control, fear that something terrible was about to happen. What tormented her most was a fear that changes in temperature might cause the bacteria in her mucous membranes to “turn virulent” and destroy her body. To control this risk she wore layered sweaters — cotton in summer, wool in winter — putting them on and taking them off at the slightest variation in temperature. She had left university and stopped seeing her friends two years earlier in order to remain at home, where the temperature was more constant and predictable. When the family arrived for the first session, Antonella’s mother described her as suffering from “nervous anorexia and psychosis” (Boscolo & Bertrando, 1993, p. 236).

The outcome of this case, unlike that of the first case study, is documented. Across three sessions, Antonella’s condition changed substantially. By the third session, held after the summer break, Antonella had started going out again, had been to the seaside, had started seeing her friends again, and her “former good spirits and sense of humor had returned.” Her “temperature phobia had disappeared, so she no longer wore her sweaters” (Boscolo & Bertrando, 1993, p. 254). The case is thus presented in the book not only as an illustration of the team’s principles but as an example of how those principles can produce documented therapeutic change.

3.2 Theoretical Foundations of Systemic Therapy

Therapy can be characterized in part by the observable behaviours of the therapist and patient during the session, but observable behaviours alone do not capture how MST works. MST rests on a set of core principles that are not enacted as discrete actions during the session but rather pervade the team’s approach and shape the intervention as a whole. This section examines those principles and how each shaped systemic therapy in practice. They are presented in the chronological order in which they entered systemic thinking, following the historical account in Appendix A.

3.2.1 General Systems Theory

What is General Systems Theory?

General Systems Theory (GST) is a transdisciplinary framework articulated principally by the Austrian biologist Ludwig von Bertalanffy in the mid-twentieth century (Bertalanffy, 1950, 1968). Bertalanffy observed that science was converging — across physics, biology, psychology, and the social sciences — on a problem that classical, reductionist science had been unable to handle: the problem of *wholeness*. Where classical science had tried to “resolve all natural phenomena into a play of elementary units, the characteristics of which remain unaltered whether they are investigated in isolation or in a complex,” the new view was that “the basic problems in modern physics are problems of organization” (Bertalanffy, 1950, p. 134).

GST is therefore the attempt to articulate principles that hold for systems of any kind — biological, mechanical, social, informational — in terms general enough to cross disciplinary boundaries. The basic premise, as Wittenborn et al. (2020) summarize it, is that “a system is a unified whole formed by interrelated parts in which a change in one part of the system affects the rest of the system.” A system has three constituents: *objects* (the parts), *attributes* (their characteristics), and *relationships* (how the parts interact within an environment) (Hecker et al., 2015, p. 45).

A handful of GST’s basic tenets — those most consequential for systemic therapy — can be stated briefly (adapted from Hecker et al., 2015, p. 45, after Minuchin, 1985):

- Any system is an organized whole; its elements are interdependent.
- The whole is greater than the sum of its parts.
- Systems are composed of *subsystems* and nested within *suprasystems*.
- Patterns within a system are *circular*, not linear.
- Systems have homeostatic mechanisms that tend to preserve their organization.
- Evolution and change are inherent in *open systems* — systems that exchange matter, energy, or information with their environments. (Closed systems, like machines, tend toward entropy.)

Implications of GST in MST

The most consequential practical implication is that to understand a system, one cannot study its parts in isolation. This represented a substantial shift in how therapy was approached: individual therapy and the psychoanalytic model that was prevalent before GST had treated the mind as “the source of mental illness” (Hecker et al., 2015, p. 43). However, from a GST perspective, “we cannot study families by looking at individual members. In order to understand families, we must study the family members in relationship to one another” (Hecker et al., 2015, p. 44). Therefore, family therapy “focuses on the family system as the source of problematic behaviours” (Hecker et al., 2015, p. 43). Problems and solutions are both seen as relational.

One aspect of GST became particularly central to MST: the notion of *circular causality* (Selvini Palazzoli et al. 1980). Nichols and Davis (2020) illustrate it with a simple case: a young woman’s low self-esteem might initially look like it is linearly caused by her mother’s fault-finding, but a circular reading reveals that “the young woman’s moping around the house might be a response to her mother’s fault-finding — and the mother’s finding fault might be a response to the young woman’s moping around the house. The more the mother criticizes, the more the young woman withdraws, and the more the young woman withdraws, the more the mother criticizes” (p. 5). Concretely, in the consulting room, the systemic therapist asks questions about who interacts with whom, in what sequence, around the presenting symptom — not what is wrong with the identified patient. Treatment targets the relational pattern. As Nichols and Davis (2020) note, “Learning to think in circles rather than lines empowers us to look at the half of the equation we can control” (p. 5).

Examples in the Case Studies

John, the crying boy

A reductionist reading would treat John’s symptom as an individual pathology. However, the Milan team read the symptom as a property of the whole family system. The team hypothesized that the family was organized around a shared myth — that this second marriage, for both parents, had to be perfect, so that each could give their children the “good”

parent they did not have in the first marriages. Sustaining the myth required each of the five children to participate. Two of them, Harry and Barbara, did so by being “perfect”; the other two, John and Donna, by being problems — giving the parents children to worry about and to care for, evidence of their indispensability as parents. John’s apparent psychosis was, on this reading, a function of the family system’s organization. (Boscolo et al., 1987, p. 88).

Antonella, the sweater girl

The team again approached the problem by looking at the system as a whole. They described the family as “divided into three subsystems: the mother and sons (who would often indirectly criticize and make fun of their father while their mother tacitly encouraged them to do so); the father, a sad, rather taciturn man; and Antonella, now completely isolated, who came to life only when speaking about her problems” (Boscolo & Bertrando, 1993, p. 237). Antonella’s isolation, phobic search for a constant temperature, and her fluctuating relationship to food were interpreted as expressions of her position in the family’s organization, and her search for an answer to the question “How does my father see me, how do my mother and brothers see me?” (p. 241). Strikingly, the first closing intervention of prescription did not address Antonella or her symptoms, but rather the tendency of the sons to show hostility towards the father and side with the mother. The unit of intervention was thus a part of the family that had not even been the presenting problem — yet the positive improvements witnessed in Antonella over the next three months, before the next session, showed the GST principle that a change in one part of a system affects all the others, in full effect. What then remained was her still-low self-esteem, which the therapist again addressed at the level of relationship, instructing Antonella and her mother to spend one day a week alone together, and once again, a long-term improvement was subsequently seen.

3.2.2 Cybernetics

What is Second-Order Cybernetics?

Cybernetics was originally developed by Norbert Wiener and his collaborators in the 1940s, defined as “the theory of control and communication in machines and animals alike” (Wiener, 1948, as cited in Boscolo & Bertrando, 1993, p. 66). Its central concept is the *feedback loop*: a

circular arrangement in which the output of a process is fed back into the system as input, influencing subsequent behaviour and allowing the system to regulate itself (Nichols & Davis, 2020).

Cybernetics distinguishes two kinds of feedback. Negative feedback detects a deviation from a desired state and acts to correct it, returning the system to its baseline. A familiar example is a home heating system: when the temperature in the room drops below the set point, the thermostat triggers the furnace, the room warms back up, and once the desired temperature is reached, the thermostat shuts the furnace off again (Nichols & Davis, 2020, p. 26). Positive feedback, by contrast, amplifies a deviation rather than correcting it. An example of this would be a driver whose car begins to skid on an icy road and, in panic or by accident, presses the accelerator: rather than regaining control, the engine speeds up, the wheels lose traction further, and the car spins out worse than before (Nichols & Davis, 2020, p. 26). Negative feedback maintains the status quo; positive feedback drives change.

Interestingly, cybernetics actually emerged in World War II aircraft, when Wiener was asked to help develop anti-aircraft fire-control systems to “minimise aiming error” (Boscolo & Bertrando, 1993, p. 66). In that engineering context, *negative feedback* and *homeostasis* were the principal concerns; *positive feedback* was a problem to be avoided, because amplifying loops led to “destructive runaways” (Nichols & Davis, 2020, pp. 26–27). The early systemic-therapy literature inherited this emphasis: the family was conceived primarily as a homeostatic system, and symptoms as devices that maintained its equilibrium, even when detrimental. (Hoffman, 1981; Jackson, 1957). Only later did cyberneticians like Walter Buckley and Ross Ashby recognise that “positive feedback loops aren’t always bad; if they don’t get out of hand, they can help systems adjust to changed circumstances” (Nichols & Davis, 2020, p. 27). Maruyama (1963) gave the formal account of this idea under the name “the second cybernetics.” From that point on, family therapists recognised that families need both kinds of feedback: negative to preserve their integrity, positive to adapt to changing circumstances (Maruyama, 1968, as cited in Nichols & Davis, 2020).

This formulation of cybernetics is sometimes called *first-order cybernetics* — the cybernetics of *observed* systems. It treats the system under study (a thermostat, a body, a family) as an

object whose feedback loops can be described from outside, by a neutral and external observer.

In the late 1960s and 1970s, principally through the work of Heinz von Foerster, this last assumption was challenged. Von Foerster (1974, 1981) argued that whenever we describe a system, we are also part of the act of describing it — every description is the description *of someone*. He called the new framework *second-order cybernetics*: the “cybernetics of cybernetics,” or the cybernetics of *observing* systems (a distinct term from Maruyama’s “second cybernetics”). As Boscolo et al. (1987) summarise: “Von Foerster (1981) states that the observer enters into the description of that which is observed in such a way that objectivity is not at all possible” (p. 14).

Second-Order Cybernetics in Systemic Therapy

The most-cited formulation of second-order cybernetics’ practical impact on systemic therapy is the Hoffman–Goolishian dictum, which Campbell (2003) takes as defining the paradigm shift: “the problem creates a system” (Hoffman, 1985, as cited in Campbell, 2003). What does this mean in practice? Cybernetic principles can be seen reflected in several recurrent moves systemic therapists make in the session.

First, in the way they treat the unit of intervention. Systemic therapists do not treat the family as an object they observe from outside; they treat the family-plus-themselves as the operative system. Their presence, their biases, their hypotheses are constituents of the observing system, not contaminants of an otherwise pure observation (Tomm, 1987; Cecchin et al., 1992). This is why the team behind the mirror is treated as a working part of the therapy, and why the therapist often invokes the team explicitly in session.

Second, in the way they iterate hypotheses through feedback. As Boscolo and Bertrando (1993) explain, systemic therapists form a hypothesis from the information gathered so far, send it into the next round of questioning, register the family’s response, and revise the hypothesis accordingly. The hypothesis is itself a node in the cybernetic feedback loop the session constitutes, not a stable belief held above it.

Third, and following from the second, in the way they conduct the interview. The session is itself a feedback loop, structured around the working hypothesis: every question is shaped by the family's response to the previous one, and every response in turn reshapes the therapist's next move. This is heavily reflected in the group's core tenet of circularity, described as the "capacity of the therapist to conduct his investigation on the basis of feedback from the family in response to the information he solicits about relationships and therefore about difference and change" (Boscolo et al., 1987, p. 11).

And fourth, in the way they read symptoms. A symptom is not interpreted as the private property of an individual but as a node in a feedback loop within the system. That loop may be predominantly negative — the symptom maintains the family's existing organisation, contributing to a homeostatic stability the family has settled into (Jackson, 1959; Boscolo & Bertrando, 1993) — or predominantly positive — the family's attempts to address the symptom paradoxically amplify it, in a runaway pattern that Watzlawick et al. (1974) named "more of the same".

Examples in the Case Studies

John, the crying boy

The classic Milan session setup operationalizes the observer-included commitment: the therapist in the room and the team behind the one-way mirror are treated as a single observing system, with the family completing the unit. The team's initial hypothesis is tested against each new piece of information the family provides — Mrs. Y's account of her first marriage, John's hospitalization, the children's varying degrees of engagement — and iterated through feedback into the final hypothesis that each of the five children supports the myth in their own way. The interview itself is a feedback loop: circular questions get one child to comment on another, the next round of questions follow from the answers. Furthermore, John's "schizophrenic" symptom is read not as his own pathology but as a homeostatic element of the family's organisation: it gives the parents their role as the good parents who care for their children, and so sustains the very myth the family has built itself around.

Antonella, the sweater girl

Cybernetic principles are even more visible in this case. The team behind the mirror is explicitly named in the room: the therapist twice leaves the session to consult them and reports back invoking their voice — “according to my colleagues...” (Boscolo & Bertrando, 1993, p. 241); “one of my colleagues said I should ask if this idea makes any sense to you” (p. 242). The observing system is thereby explicitly invoked as a collaborator. The “research” metaphor that the therapist offers for Antonella’s temperature-monitoring is also iterated through feedback in real time: the therapist proposes that Antonella is looking for “a constant”, Antonella refines it to “an average”, the mother adds “there’s no end to it because every day is a new day,” and Antonella again adds “every day is newborn, as if the previous day has been totally useless!” (Boscolo & Bertrando, 1993, pp. 239–240).

3.2.3 Information as a Difference

What is Information as a Difference?

Gregory Bateson, who informed a large part of Milan systemic theory, drew a clear distinction between *data* and *information*. Data are isolated statement, such as “I am depressed,” “John is schizophrenic,” “it was in the summer” (Boscolo et al., 1987, p. 96). Information, by contrast, is “a difference that establishes a relationship” (Boscolo & Bertrando, 1993, p. 86).

Information can be a difference between two entities — “I am more depressed than my sister” — or between the same entity at two points in time — “I am more depressed than I was a year ago” (Boscolo & Bertrando, 1993, p. 82, footnote). Either way, the difference is always a relationship, since a difference cannot be drawn without two entities being compared. Tomm (1984) goes further to say that “a relationship is always circular by virtue of its reciprocity” (p. 119).

Bateson argued that information, so defined, is what truly matters. He gave the example of a piece of chalk: we can only really know it by registering the differences between it and other things in the universe — and the differences between its molecules at one moment in time and space and at any other (Bateson, 1972, p. 458). He illustrated the same point with the example of *micronystagmus*, the involuntary movement of the eye which is “vital to sight, by which the

eye shifts quickly back and forth picking up differences in shape, color, and brightness” (cited in Boscolo et al., 1987, p. 12). Without that constant shifting, the visual field collapses; what the eye actually registers is not objects but the differences between them.

Information as a difference in systemic therapy

If a system can only be acted on through the introduction of differences, then for the systemic therapist to produce change, they must introduce information in Bateson’s sense of the word. Families typically arrive at therapy with data: diagnostic labels (“John is schizophrenic,” “Antonella has anorexia”), isolated symptoms (“I am depressed”), and non-relational descriptions (“he is crazy”). Such descriptions are “a compilation of facts that do not describe a relationship” (Boscolo et al., 1987, p. 97), and as such offer little for the system to act upon — those descriptions have already been absorbed into the family’s existing organization. The original Milan team considered the conversion of data into information to be “an active process that can only be done while interviewing the family” (Boscolo et al., 1987, p. 107). The principal technique through which the team operationalised this conversion is *circular questioning*, which generates contrast and surfaces differences. Circular questioning is described in §3.4.1.

Examples in the Case Studies

John, the crying boy

The family arrives with data: “acute schizophrenic attack” and a list of disturbing behaviours, including hyperactivity. The Milan team’s interviewing strategy is to convert this data into information by surfacing relational differences. They ask circular questions that compare children: who is more upset when John is hyper? Who was hyper before John was? The label is spread across the system, and hyperactivity becomes a property of a difference between siblings rather than a property of one of them. The closing intervention then makes this conversion explicit at the level of the family’s myth: the team names “the perfect marriage” myth as a difference between the parents’ tragic first marriages and the redemptive present one, and traces how each of the five children’s behaviour is positioned in relation to it — Harry and Barbara as perfect, John and Donna as preoccupations for perfect parents to

manage. Each child's behaviour becomes information about the others' behaviour and about the parents' myth, and none of it stands as raw datum any longer.

Antonella, the sweater girl

Similarly, Antonella's mother arrives with what is initially data: her daughter has "nervous anorexia and psychosis," she wears multiple sweaters, she has a phobia about temperature, and she does not eat (Boscolo & Bertrando, 1993, p. 236). The therapist's questions transform these descriptions into information by surfacing differences across people, across time and relationships: when did Antonella first start wearing the sweaters? What was happening in the family at that time? Who is closest to Antonella? Who is most worried about her not eating? Once again, the datum is reframed as a difference. The intervention also points out a consequential difference: between the mother's current emotional supply ("sugar") from her sons, and the future shortage she will experience when her sons marry and leave her home.

3.2.4 Structural Determinism and Perturbation

What is Structural Determinism and Perturbation?

Maturana provided many ideas to systemic therapy — chief among them, the concepts of structural determinism and perturbation.

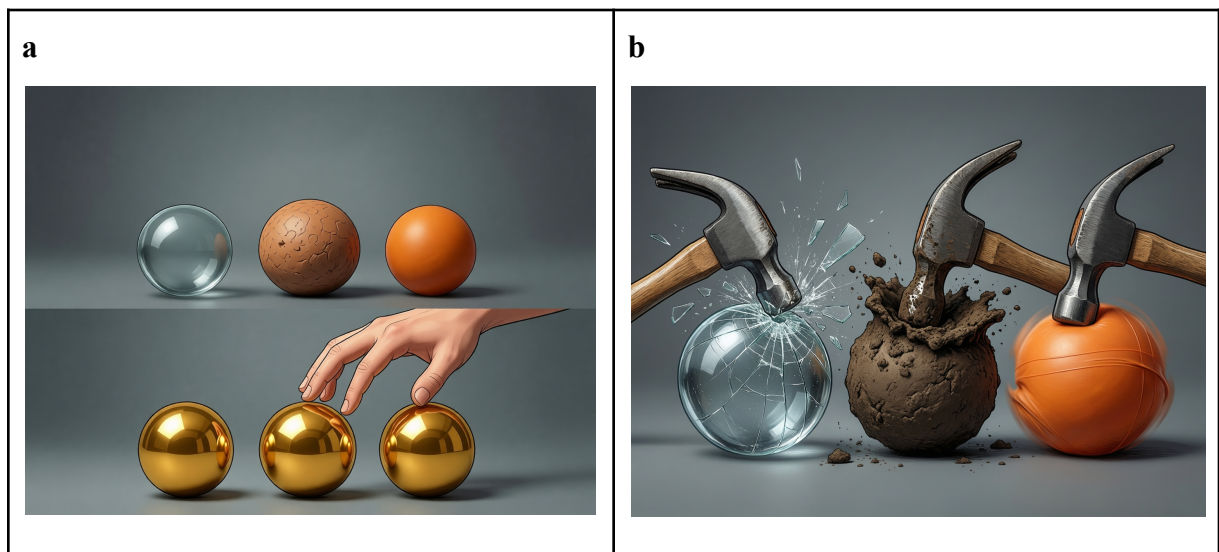
Structural determinism is the principle that a system's own structure (its components) and organization (the relations between the components) fully determines both which interactions can affect it and how it will respond to any given interaction (Dell, 1985). Therefore, the environment (including a therapist's interventions) does not dictate a system's behaviour; it can only trigger responses that the system's structure already makes possible.

A good way to understand the logic behind this principle is to consider what would be the case if we considered its opposite — which Maturana termed *instructive interaction*. Under the latter, one system can directly specify the behaviour of another. Dell illustrated this with the example of King Midas: in the Greek myth, he could turn any object to gold simply by touching it, regardless of the object's own properties (Dell, 1985). Humans tend to assume that communication works this way too: i.e., that information can determine how a living system

will respond. Yet if this were true, a professor’s lecture would produce identical understanding in every student, and a therapist’s intervention would yield the same outcome with every client — which is clearly not the case. As Dell explains, the lecture does not *determine* the students’ responses; it merely *selects* from among the responses their individual structures make possible. It is the students’ own organisation — their prior knowledge, cognitive patterns, and current state — which determines what they are capable of understanding (Dell, 1985). Figure 11 shows an illustration of instructive interaction and structural determinism.

Figure 11

An illustration of Maturana’s structural determinism and comparison to instructive interaction



Note. **(a)** Instructive interaction, as conceptualised by Maturana. King Midas’s touch causes any entity to turn to gold, regardless of that entity’s properties. The instruction determines the outcome. **(b)** Structural determinism, as conceptualised by Maturana. A hammer produces different results when it hits different objects, determined by that object’s structure. The outcome is deterministic, but based on the object itself and not on the instruction. Original figure, created with Qwen AI.

The implication is that one cannot directly instruct a system (i.e. a human or system of humans) to change; one can only perturb it. However, perturbation is still necessary — without any lecture, the students cannot have any reaction to it at all, whether it is increased understanding or a different change. Therefore, external perturbation is necessary in order to

initiate change in a system; and the system's structure and organisation constrains and defines what changes are possible within the system, from which the perturbation then selects.

It is worth noting the distinction Maturana drew between a system's organisation and its structure. The organisation is the pattern of relations that defines what kind of system it is — systems of the same kind have the same organisation. The structure is the actual manifestation of a particular example: its concrete components and their actual interactions (Maturana & Varela, 1980; Mingers, 1991, pp. 320–321). For a family, for example, the organisation is the relational pattern that makes it a family of that type (a blended family, a single-parent family, and so on), while the structure is the specific people involved and the concrete relations they enact. A system's structure can change while the system remains the same kind of system — but if its organisation changes, the system ceases to exist *as that system* and is replaced by another (Mingers, 1991, p. 320).

Structural Determinism and Perturbation in Systemic Therapy

Given this premise, systemic therapists do not attempt to “instruct” a family on how it should be organized, or how individual members should behave - Cecchin even called this “therapeutic violence” (Cecchin, 1987, fn. 1). Rather, “the therapists must act as a stimulus, a perturbation that activates the family's capacity to generate its own solutions.” (Boscolo et al., 1987, p. 100). The team expressed a belief in the family's “capacity to heal itself” (Boscolo et al. 1987, p. 102).

Systemic therapists attempt to perturb through new information, in the Batesonian sense (see §3.2.3). The idea is that familiar information cannot produce any change because the system has already accommodated it; therefore only genuinely novel information can trigger change (Boscolo et al., 1987). The contemporary CMFT school's approach is particularly purposeful about this: “We are very careful not to share the same point of view of the problem-determined system—all the people who share the same vision of the problem—and are very careful never to collude with the dominant ideas, in order to introduce dissent and differences” (Barbetta & Telfener, 2020, p. 14).

The problem the family brings can be understood as an *organisation* — a particular set of relations among the family members that holds the problem in place. Anderson et al. (1986) called this a “problem-determined system,” and the Milan team adopted the same idea: *the problem creates the system* rather than the system creating the problem (Boscolo et al., 1987, p. 14). When the family’s organisation (i.e. the relationships that constitute it) changes, the problem ceases to exist, because the configuration that made the problematic behaviour is no longer there.

Examples in the Case Studies

John, the crying boy

The Milan team did not tell the Y family how a blended family ought to be organised, did not prescribe how the parents should treat any of the children or instruct them on John’s care, and did not give John or the other children directions about how to behave toward John or toward each other. The intervention consisted of introducing new information into the family’s view of itself, through circular questioning and positive connotation (see §3.4.1 and §3.4.2. From there, the family was left to reorganise itself.

Antonella, the sweater girl

The Milan team did not instruct Antonella to eat, did not prescribe medication, and did not tell the parents how to respond to her temperature phobia or her sweater-wearing. Instead, the team perturbed the family system by addressing the sons’ tendency to side with the mother against the father, and prescribed a ritual in which Antonella and her mother were to spend one day a week alone together. From there, the family was left to reorganise itself. And, as we can see in the case study’s outcomes (see §3.1), it did.

3.2.5 Constructivism

What is Constructivism?

Constructivism, as is most relevant for systemic therapy, arose as a concept in the 1980s, carried by three main figures: Humberto Maturana with the biology of cognition, Heinz von

Foerster with second-order cybernetics, and perhaps most notably Ernst von Glasersfeld with radical constructivism (Boscolo et al., 1987). Constructivism argues that knowledge is not an accurate representation of an objective reality but rather the experiencing organism's own construction. Therefore, "we each bring our own assumptions to the same situation; we each construct reality differently" (Goldenberg & Goldenberg, 2009, p. 28). Where a realist asks whether knowledge *matches* the world, the constructivist asks only whether it *fits* — whether it lets the organism navigate, predict, and act (von Glasersfeld, 1984).

Constructivism in Systemic Therapy

Constructivism had a substantial impact on systemic therapy starting from the mid 1980s (Boscolo & Bertrando, 1996). Whereas pure systemic thinking focused on behaviour, constructivism "shifted the focus to the assumptions people have about their problems" (Nichols & Davis, 2020, p.29). In fact, the very idea of systems is constructivist — the criterion for determining the boundaries of a system is not that our definition somehow corresponds to a particular truth, but rather that it is pragmatically useful (Fryszler & Schwing, 2014). Therefore, the goal of therapy "changed from interrupting problematic patterns of interaction to helping clients find new perspectives on their lives" (Nichols & Davis, 2020, p.29).

As a result, the position of the therapist as being an impartial authority was displaced. As Mingers explained: "There can be no objectively correct outside description of the family's reality. The family and its members construct, through their linguistic interactions and conversations, the reality that they experience. All such realities are equally valid, although not necessarily equally desirable, and there can be no outside privileged viewpoint" (Mingers, 1991, p. 332).

Therefore, more attention started to be paid to what actually happens in sessions: "to exchanges of information, emotions, and meanings between therapists and clients" (Boscolo & Bertrando, 1993, p. 89). Also, more attention was put to the observing system (the therapists), as "clients observe therapists as much as therapists observe clients" (Boscolo & Bertrando, 1993, p. 89). In other words, the focus became on the therapeutic process itself rather than on

a final intervention. Indeed, the Milan team has explained that they started conducting their typical “prescription” interventions less and less frequently, relying instead on the interview process itself to facilitate change (Boscolo & Bertrando, 1993).

Therapists also changed their approach to formulating hypotheses. The early Milan team attempted to find a single hypothesis that would fit the family; the post-constructivist team holds multiple hypotheses simultaneously, each from a different observer’s position — a “stereoscopic” picture replacing a “flat, monoscopic painting” (Boscolo & Bertrando, 1993, p. 90). This also shows up in the team’s careful use of language — substituting the word *is* with *behaves* (“Mother behaves in such a way that appears overprotective” rather than “Mother is overprotective”) so that labels are recognised as descriptions made by an observer, not as properties of the person (Boscolo et al., 1987, p. 99).

Furthermore, family therapy developed new techniques during this period, such as *reframing*. This is relabeling a person’s behaviour to shift how other family members respond to it. Nichols and Davis (2020) elaborate with an example:

“The dispirited parents of a rebellious 10-year-old will feel better about themselves if they become convinced that, rather than being ‘ineffective disciplinarians,’ they have an ‘oppositional child’. The first diagnosis suggests that the parents should get tough but also that they probably won’t succeed. The second suggests that coping with a difficult child requires strategizing. The point isn’t that one description is more valid than the other, but rather that if whatever label a family applies to its problems leads to ineffective coping strategies, then perhaps a new label will alter their viewpoint and lead to a more effective response.” (p. 29)

Examples from the case studies

The Crying Boy case study occurred in the late 1970s — at this point, some constructivist ideas were already present in Milan systemic therapy through the influence of second-order cybernetics, but they were not so much in the forefront yet as the influence of Maturana and von Glasersfeld had not yet arrived in the field. The Antonella case study, on the other hand, happened in the early 1990s, when the ideas were much more prominent and well formulated.

Therefore, we can use these two case studies to compare the therapists' approaches with and without constructivist influence.

John, the crying boy

Some constructivist-adjacent moves are already visible — the team treat their working hypothesis as revisable, and their circular questions and closing positive connotation reframes the idea that “John is schizophrenic.” But the commitments are not yet fully theorised. The team holds a single hypothesis at a time: their initial hypothesis was replaced by a new single hypothesis when contradictory information emerged, rather than held alongside others. And the centre of gravity of the session is the final intervention — the consultation builds toward the positive connotation and the closing ritual as the climactic moment.

Antonella, the sweater girl

This case study shows constructivism influences more clearly. The therapists generate multiple hypotheses simultaneously — Antonella as the isolated daughter of a divided family, the family as organised into three subsystems, the sons as protective of their mother — held alongside one another rather than serially replaced. The team's language is constructivist throughout: the mother's description of “nervous anorexia and psychosis” is accepted as her construction without being adopted as the team's. There is a clear example of re-framing: the family arrives describing Antonella as “sick” and obsessive compulsive, but the therapists instead suggest “You're perhaps a little like a university researcher who's studying the effect of temperature on body cells, a very systematic kind of research” (Boscolo & Bertrando, 1993, pp. 238–239).

3.2.6 Social Constructionism

What is Social Constructionism?

Social constructionism is associated principally with Kenneth Gergen (1985), and it crystallised as a position in family therapy in the late 1980s and early 1990s. Adjacent figures include Harlene Anderson and Harold Goolishian (the “not-knowing” position), Tom

Andersen (the reflecting team), and Lynn Hoffman, who from 1991 onward becomes one of its main voices.

Social constructionism arose largely as a response to criticism of constructivism, which explains why they are frequently confused even by some researchers in the field themselves (Hoffman, 1991, p. 4). Like constructivism, social constructionism holds the idea that no “real world” exists and knowledge is subjective; but while constructivism asserts that reality is constructed inside the individual nervous system of the observer, social constructionism locates it instead in conversation between people (Goldenberg & Goldenberg, 2009; Hoffman 1991). As Hoffman put it, knowledge “evolves in the space between people, in the realm of the ‘common world’ or the ‘common dance’” (Hoffman, 1991, p. 4).

Social Constructionism in Systemic Therapy

Like the earlier constructivism, social constructionism brought several shifts in perspective and approach to systemic therapy. In Cecchin’s formulation, the team’s earlier view that “families stayed together to control relationships” was replaced by the view that “they stay together to make sense with one another,” and once the therapist is added to that equation, “therapy becomes a process of the social construction of meaning” (Cecchin et al., 1992, cited in Campbell, 2003, p. 21).

The therapist is no longer a discoverer of family functioning but a co-author of new stories — and the co-authorship is treated explicitly as such: the therapist’s hypotheses and biases are named to the family, the team is mentioned explicitly, and the therapist talks about her own values, gender, and theoretical commitments (Cecchin, 1992). Cecchin (1992) gives a memorable worked example of could be said to a family:

“I cannot avoid thinking that many problems in your family stem from the fact that your behaviour seems to be ruled by a patriarchal pattern that tends to oppress women. Some of the stories you told me succeeded in convincing me of this interpretation. Therefore, I will give you some instructions with the hope of breaking this pattern. Some of my colleagues behind the mirror, however, warned me that it is not proper to interfere with how families are organized, no matter

how inappropriate we might think the organization is. I had a long argument with them and we came to the conclusion that I will follow my conviction, but only for five therapeutic sessions.” (Cecchin, 1992, p.)

The Milan-lineage adoption of social constructionism is, however, only partial. Boscolo, Cecchin, and Bertrando explicitly refuse the strong “not-knowing” position of Anderson and Goolishian. As Campbell (2003) notes, “while the observer’s lens widened to include the therapist and client as a co-constructing unit, the Milan group never relinquished their view that therapists are the ‘conductors’ of the session and as such, must claim responsibility for their own thoughts and actions... [H]ypothesising was never abandoned.” What the team did adopt was Cecchin’s later stance of *irreverence* — taking one’s own hypotheses seriously enough to use them, lightly enough to drop them (Cecchin et al., 1992). The therapist still intervenes, prescribes, and reframes — but does so transparently, time-limitedly, and with explicit acknowledgement that the interventions are local constructions, not universal truths.

Examples in the Case Studies

John, the crying boy

As this case study occurred too early to see influences of social constructionism, it is useful mostly as a backdrop demonstrating what the approach may look like without this concept. The team behind the mirror conceals its discussions from the family and the team’s biases are not named in the room. What can be seen, however, is the team treating its own understanding as something assembled in real time, with team members visibly disagreeing (Boscolo et al., 1987).

Antonella, the sweater girl

The Antonella case shows several social-constructionist moves more clearly. The “researcher” reframe is built collaboratively across several turns of conversation: the therapist offers the metaphor, Antonella partly accepts and partly resists (“Yes, no, well usually yes”), the mother is drawn in, and the metaphor is elaborated together. The therapist also explicitly consults the team mid-session, in front of the family: “Excuse me, I’d just like to ask my colleagues behind

the mirror if one of them has had any experience of this kind of research” (Boscolo & Bertrando, 1993, p. 239). This naming of the team to the family — a step toward the reflecting-team format developed by Andersen — is a social-constructionist move made before the Milan team had fully articulated its commitment to the position.

3.3 Key Concepts in Systemic Therapy

The concepts in this section differ from the previous one in that they are not theories borrowed from other researchers and fields, but ideas developed by the systemic therapists themselves. Yet, they are still not concrete, specific actions, but rather concepts that pervade the entire systemic therapeutic practice. As such, they are examined separately from the methodology in §3.4.

3.3.1 Resistance

Resistance is a notion that began with very strong influence on systemic therapy, as the early Milan team adopted the concept from the MRI and the strategic tradition. The family was assumed to be actively defending its current organisation against the therapist’s attempts to change it — even when the symptoms that organisation produced were painful. Jay Haley in particular developed a way to use the family’s “resistance” to their advantage: the practice of “paradoxical prescription”, where the symptom was positively connoted and the family was instructed to continue with the problematic behaviour. In Haley’s reasoning, in this case, to resist the therapist would mean giving up the symptom — therefore, resistance becomes therapeutic (Haley, 1963, as cited in Boscolo et al., 1987, p. 6).

The early Milan team’s clinical vocabulary reflected this combative stance. As Boscolo et al. (1987, pp. 6–7) record, the therapeutic encounter was described in terms of “secret battles,” “denied coalitions,” “moves,” “counterattacks,” “escalations,” “tactics,” and “ploys”:

“Original Milan team discussions sounded like councils of war. The therapists took care, however, not to challenge the family openly. They preferred to operate like guerrillas, using creative deceptions that would bypass clients’ ‘resistance’.

Much time and thought was spent on ways to keep the family from rendering the team powerless.” (Boscolo et al., 1987, p. 7)

The shifts in the 1970s and 1980s undermined this agonistic framing: second-order cybernetics meant that resistance could not be a property of just the family, as it was more likely to be a property of the therapist-family system as a whole, including the therapist’s expectations. And secondly, Maturana’s idea of structural determinism meant there could be no instructive interaction, and therefore the family cannot strictly speaking “resist” instruction.

After these shifts, the team did not entirely discard the concept of resistance, but they substantially reinterpreted it. First, non-compliance with a directive became something to learn from rather than something to outmanoeuvre. As Boscolo et al. (1987) describe the shift:

“The ignoring of a directive is no longer taken as a ‘maneuver’ to defeat the therapy team, but rather as new evidence about how the family system works (and how far off the therapists were in their attempts to understand it). The team then weaves this information into its ongoing, evolving hypothesis.” (p. 22)

Second, the team came to recognise their own role in producing what they had previously called resistance. They asked, “What are we doing now to create resistance?” — treating it as a relational event between therapist and family, not as a property of the family alone (Boscolo et al., 1987).

By the 1990s, resistance in this sense of the word had largely disappeared from systemic therapy. The recognition that systems can be conservative or slow to change is retained, however, under softer vocabulary. Bertrando (2007) writes about clients “blocked in the idea that their present state is the only possible one” (p. 47), and about therapeutic systems that get “drawn within the family system, sharing most of the family’s values and acting in a way that maintains the family status quo” (p. 120). Barbetta and Telfener (2020) write that “pathological systems are stuck more than ‘ill’ and for this reason are admirable, since they find incredible capabilities to cope with life, which is always changing.” On this contemporary reading, what looked to the early Milan team like adversarial resistance has been re-described

in two related ways — as adaptive coping that has become limiting, and as a co-produced stuckness in which the therapist has her own part.

Examples in the Case Studies

John, the crying boy

This case study still shows influence from the strategic period. We can see this in their use of positive connotation and paradoxical prescription, explained in §3.4.1 and §3.4.2. The therapists believed the family would naturally resist change — therefore, to use this to their advantage, they told the family to continue with the behaviour they considered problematic and reframed it as something benefiting the family. (See the aforementioned sections for further details).

We can already see the team's shift in their view of resistance in the authors' comments on the case study, provided before the book's publication in 1987:

“A better intervention in this family would have been to change the premise and not accept it as we did. We should have said ‘We understand why the parents have created this myth of having to be perfect. They have this myth because they suffered too much. But now we see the myth is not necessary. It will take a lot of time, two or three years perhaps, for them to get over this myth and forget their old partners. And while they’re waiting for this moment, the children should meet together to see what they can do to help their parents.’” (Boscolo et al., 1987, p. 90)

Antonella, the sweater girl

Compared to The Crying Boy, this case study offers a striking difference in the therapists' treatment of resistance. Most strikingly, the team did not use paradoxical prescription or positive connotation. Rather, they said to Antonella that they were not worried about her because they knew she could decide to change her behaviour at any moment (Boscolo & Bertrando, 1993, pp. 250–251). Rather than regarding the family as resistant and trying to “outmanoeuvre” it, the team expressed trust in the family members' capacity to change at will.

3.3.2 Neutrality

Neutrality was the third of the Milan team's three guiding principles (alongside hypothesising and circularity), as articulated in their 1980 paper "Hypothesizing — Circularity — Neutrality" (Selvini Palazzoli et al., 1980). In its original formulation, neutrality is a pragmatic effect the therapist's behaviour has on the family rather than an inner state: at the end of a session, no member of the family should be able to say whose side the therapist was on. This requires the therapist to "obtain and maintain a different level (metalevel) from that of the family" (Selvini Palazzoli et al., 1980, p. 6) — to operate at a logically higher level rather than as another participant inside the family's disputes.

This original formulation came under pressure as the team absorbed second-order cybernetics and constructivism, and it is the tenet that has arguably undergone the greatest revision since its conceptualisation. Boscolo and Bertrando (1993) describe the shift in retrospect:

"As formulated in the seventies in first cybernetics terms, the concept assumed that observers and observed were separate. It was possible, then, for the therapist to place oneself 'above' one's clients. Second-order cybernetics changed all this: separation was now impossible, the system had to include the observer and the observed, so the therapist could never be really 'neutral' because one could never be neutral about oneself, one's biases, one's own ideas." (p. 88)

If the observer is part of what is observed (see §3.2.2 on second-order cybernetics), then there is no metalevel from which to be neutral; if every claim the therapist makes is an observer-distinction (see §3.2.5 on constructivism), then it is impossible to withhold judgment about reality. As such, neutrality was reformulated as "a state of curiosity in the mind of a therapist," where the therapist refuses to commit to any single description of the family as the correct one, moving continuously among hypotheses, viewpoints, and relational positions (Cecchin, 1987, p. 405). Thus, it was reframed as an active stance rather than a withholding one. The neutral-as-curious therapist is not detached but is actively generating alternative interpretations, none of which is treated as final. Tomm (1987, p. 4) makes an important clarification: neutrality is "physically and logically impossible to remain absolutely," as

certain questions or comments can be seen as aligning with one family member and may therefore break neutrality. Rather, it is “the result, over time, of the interaction between the therapist and the family” (Boscolo et al., 1987, p. 149). As Boscolo et al. put it (1987, p. 12), neutrality is “closer to multipositional than nonpositional”; the therapist is “able to move through these conflicting claims like a charmed person without becoming ensnared by them.”

Importantly, the stance has clear limits. Neutrality is not moral indifference, and the Milan team are explicit that they do not condone destructive behaviour: “We do not condone socially destructive behaviour. We only introduce this different way to punctuate it” (Boscolo et al., 1987, p. 152). Hoffman (1985) explains: “one cannot be neutral and be a parent, a teacher or a policeman.” Where social-control or safety considerations demand a position (“Call the police. Use the hospital. Set up a suicide watch” (Hoffman, 1985)), the therapist takes one, and recognises that they are no longer in the neutral or curious stance for that moment. The contemporary Milan-Bertrando lineage prefers to speak of “a *tendency* towards neutrality, a therapist and team ideal which, by definition, can never be fully attained” (Boscolo & Bertrando, 1993, p. 88).

Therapists achieve neutrality through several specific practices. Firstly, the therapist asks questions rather than making statements, since “to make a statement means you’re not neutral; you’re defining something” (Boscolo et al., 1987, p. 153). Secondly, they constantly switch their time and attention between family members, asking everyone for their input and opinion in turn, so that they cannot ally with any one member for a longer time. In fact, in the second case study presented in Boscolo et al. (1987), the therapists mentioned they made the mistake of largely leaving the father out of the dialogue, so they inadvertently excluded him and formed more of an alliance with the mother and daughters of that case.

Examples in the Case Studies

John, the crying boy

This case study was conducted before the curiosity reformulation, therefore it demonstrates neutrality is in its original 1980 form. The team moves circularly through the family, asking Donna how Mother behaves when John “becomes nervous and dizzy,” asking which of the

parents is more concerned about John's symptoms, and asking the siblings what they think would happen if John got better. Each question equalises the team's alliance: by the end of the session, no family member could name whose side the therapist was on. In the closing intervention, the team's positive connotation of the family myth is constructed so that no member is implicitly cast as a villain (see §3.4.2). The team also do not accept John's hospital diagnosis of an "acute schizophrenic episode" as a fact about him — they treat it as a label to investigate rather than to confirm.

Antonella, the sweater girl

In this case, by contrast, the stance is the post-1987 curiosity reformulation. The therapist works actively through the session, generating one alternative interpretation after another — Antonella as the isolated daughter, as a "researcher" studying temperature, as the family's mediator — and treating each as a possibility to be tested in conversation. There is constant continuity between the different stances: the therapist does not treat any one description of the family as final.

3.3.3 Non-Normativity

Non-normativity is the refusal to define what a "correct" or "healthy" family should be like; instead of specifying changes, the therapist should observe how the family actually interacts and creates the context for change (Hoffman, 1985). This means when a family comes, the therapist does not impose a diagnosis or treat anything as a problem that the family itself does not identify as a problem. Cecchin gives the example of a family where the women are more energetic and the men are quieter (Boscolo et al., 1987). As he explains, "you would never say 'Look at this man. He doesn't take initiative, he's totally destroyed by his wife,' because that implies they shouldn't be this way" (p. 69). Instead, the non-normative formulation would be "I see in this family the women have tremendous energy and the men are very quiet. When did you decide to be quiet like that? Or energetic like that?" (p. 69).

This principle has been present in MST from its earliest days, as it was imported from the strategic MRI approach. Over time, this emphasis only grew:

“In dialogue with clients we increasingly avoided words associated with illness, like ‘symptom,’ ‘session,’ and ‘diagnosis,’ thus avoiding the use of the language of pathology, to be consistent with the most recent theoretical developments (second-order cybernetics, constructivism, constructionism, post-modern views).” (Boscolo & Bertrando, 1993, p. 106)

In its contemporary expression, this commitment has developed into a position of radical inclusiveness. Bertrando and Lini (2019) describe their contemporary stance:

“There is no language, no way of speaking or staying in the dialogue, that is ‘right’ or ‘wrong’ in itself — except the attempt to silence or the inability to listen to the others. Every language and every way of being should be respected. We also believe that any participant in a therapeutic dialogue has something relevant to say about themselves and the others, and should get the opportunity to say it” (p. 178).

The CMFT today likewise continues to follow a non-normative stance:

“We do not adhere to a preconceived notion of right or wrong, healthy or sick; rather, we are concerned with the inefficiency of relational functioning, specifically how the emotional and/or other costs become burdensome for the system. ... Today, depathologization is a characteristic more frequently associated with narrative and dialogic therapies than with systemic therapies; however, it remains a central element of the systemic approach of the Milan school.” (Telfener & Barazzetti, 2025, p. 9-10)

Examples in the Case Studies

John, the crying boy

The clearest non-normative move in this case is the team’s treatment of the diagnosis John has been given. The family arrives carrying the hospital physician’s verdict — “an acute schizophrenic episode” — together with the treating psychiatrist’s refusal to speak with the parents (Boscolo et al., 1987, p. 38). The team neither adopts the diagnosis nor offers a

counter-diagnosis; it treats the label as something the family has come to live inside rather than as a fact about John. Hoffman's commentary on the case makes the principle explicit:

“Labelling plays an important part in this case. People give and accept labels in the hope that they will supply meaning to something poorly understood and frightening, such as mental illness. ... Once a label has been accepted, all subsequent behaviours become related to that label and every action becomes testimony to schizophrenia: he sleeps too much, he sleeps too little; he talks too much, he talks too little” (Boscolo et al., 1987, p. 33–34).

The team's antidote is circular questioning that redistributes the label across the family — “Who was the most ‘hyper’ or ‘jittery’ before John was?” — so that “‘hyper’ becomes a description of shared or potential behaviours in the system and not testimony to the illness of one member” (Boscolo et al., 1987, p. 33). The closing intervention is consistent with this stance: rather than naming John as the sick one to be treated, the team positively connotes all five children's behaviour as work in service of the family's shared myth that the second marriage must be perfect. At no point does the team define what a well-functioning version of this family would look like; the intervention is meant to perturb the system and allow the family to reorganize itself.

Antonella, the sweater girl

The Antonella case shows the same commitment in the post-1987 frame. The family arrives carrying not one but two chronic diagnoses delivered by the previous psychiatrist over three years of individual treatment: “nervous anorexia and psychosis” (Boscolo & Bertrando, 1993, p. 236). Rather than working within this diagnostic frame, the therapist reframes Antonella's temperature-regulation behaviour using the metaphor of a “researcher” studying body cells in collaboration with her engineer brother. Boscolo and Bertrando's own gloss on this sequence is the clearest single statement of non-normativity at work in the case:

“The therapist is using metaphorical discourse to introduce other possible worlds: most importantly, a world in which Antonella is redeemed from being branded as sick or deviant. The ‘research’ metaphor not only restores her status as a university student but

also implies collaboration with other members of the family” (Boscolo & Bertrando, 1993, p. 240).

The same non-normative stance shapes the final intervention: the therapist sets out for Antonella “all the possible options [she] has: remain as she is, change, get married, not get married, make a career for herself, lead an ‘ordinary’ life,” explicitly to “broaden Antonella’s range of options, and also to avoid losing his neutrality by forcing her into a prescribed solution” (Boscolo & Bertrando, 1993, p. 250). The principle is summarised in the authors’ own retrospective on the second session: “The family came to therapy with a diagnosis: Antonella is suffering from anorexia and psychosis. The therapist restated the problem in terms of existential rather than clinical suffering” (Boscolo & Bertrando, 1993, p. 252).

3.4 Key Methods in Milan Systemic Therapy

3.4.1 Circular Questioning

Circular questioning is the distinctive interviewing technique the Milan team developed in the late 1970s and continued to refine throughout the 1980s. In Cecchin and Boscolo’s own retrospective account, the technique was crystallised not through deliberate design but under training pressure:

“Toward the end of the 1970s, Boscolo and Cecchin found that their students were not interested in the elaborate paradoxical prescriptions the trainers would devise but asked instead ‘Why did you ask that question?’ ‘Why did you address the mother instead of the daughter?’ ... The focus on therapist instead of family behaviour clarified the development of a specific style of questioning which the team called circular questioning” (Boscolo et al., 1987, p. 10).

The technique was first formalised in the 1980 “Hypothesizing—Circularity—Neutrality” paper around two Batesonian axioms: *information is a difference*, and *difference is a relationship (or a change in a relationship)* (Selvini Palazzoli et al., 1980). Every circular question is shaped by these two axioms: it asks about a *difference* — between people, between

time-frames, between behaviours, or between perceptions — because differences are the units of information, and because differences are always relational.

Tomm (1987) extends the metaphor explicitly to therapy:

“Just as the eye, in order to see, needs to move back and forth with a continual micro-nystagmus in order to distinguish ‘differences’ in the patterns of light falling upon the retina, so therapists must continually probe family members by asking questions, paraphrasing their answers, and noting their verbal and nonverbal responses in order to draw distinctions about their experiences” (p. 3).

The technique is also, in this sense, the operational counterpart of the Milan principle of *circularity*: it is the means by which the therapist’s investigation is conducted “on the basis of feedback from the family in response to the information he solicits about relationships and therefore about difference and change” (Selvini Palazzoli et al., 1980).

Operationally, the most distinctive form is *triadic*: rather than asking one family member directly about their own experience, the therapist asks them about the relationship between two others. Penn (1982) calls this “gossiping in the presence” and notes that “when a question of relationship and difference is asked, ‘Who in the family is the first to notice when your sister is depressed?’ or ‘Who worries with mother when she worries?’ one cannot *not* give a relational description as an answer” (p. 269). A direct question (“Mrs Y, how do you feel about your son’s symptoms?”) typically elicits the family’s official version of itself — well-rehearsed material that most family members have heard before and do not learn anything new from. A triadic question (“Donna, how does your mother behave when John becomes nervous?”) forces each family member into a new position simultaneously: Donna must describe her mother’s relationship with John in her mother’s presence; the mother must hear herself described by her daughter; John must hear the description of how others respond to his behaviour; and the whole family becomes, briefly, a “community of observers” of itself (Boscolo & Bertrando, 1993, p. 86).

The most detailed early typology of circular questions was developed by Karl Tomm (1985), who distinguishes two principal axes.

The first axis concerns the *content* of the question, with four question subtypes:

“Content” type circular questions:

1. *Category-difference questions* compare persons (“Who gets more upset when she runs away, mother or father?”), relationships (“Is your wife closer to Sandra or to Bill?”), perceptions, or actions.
2. *Temporal-difference questions* compare states across time (“Was there more fighting before or after mother’s stroke last year?”; “Was father closer to Sandra when she was a little girl or is he closer now?”).
3. *Categorical-context questions* explore connections between meanings and actions (“When you think he’s trying to control you, what does he do?”).
4. *Temporal-context questions* trace interactional sequences (“When you are at dinner and your sister is not eating, what do your parents do?” — followed by “And then what does your mother do?”, “Then does she eat?”, etc.) until a whole circular pattern between mother, daughter, and father becomes visible.

The second axis is the question’s *intention*, with two main subtypes. The distinction between the two types lies not so much in their formulation as in their timing and the therapist’s intent and assumptions: the same question can serve either function (Tomm, 1985, 1988).

“Intention” type circular questions:

1. *Descriptive questions* are asked to elicit information that refines the therapist’s hypothesis about the system. Most of the content-axis questions above function as descriptive questions when asked early in a session, before a working hypothesis has been built (“Who gets more upset when she runs away, mother or father?” asked simply to map the family’s reactivity).
2. *Reflexive questions* are asked “to deliberately trigger a change in the system being investigated” (Tomm, 1985, p. 35). Several further sub-types of reflexive questions appear in the Milan literature:

- a. *Hypothetical questions* invite the family to consider what would happen under a counterfactual scenario (“If your mother decided to stop worrying about you, what would your father do?”).
- b. *Future-oriented questions*, which Penn (1985) developed in detail as *feed-forward*, ask the family to simulate states it has not yet entered (“If this arguing were to continue, what do you expect will become of your relationship in five years from now?”).
- c. *Embedded-suggestion questions* subtly propose a possibility while still asking (“If you were to bring her a rose after an argument, like some men do in order to heal a relationship, would you be violating your family tradition?”).

Future-oriented questions were regarded as being particularly powerful, eliciting much stronger responses, particularly in nonverbal behaviour, than questions about the past or the present (Boscolo et al., 1987). As the team explains, this is likely “related to the fact that the family, over time, has developed a deterministic ideology. But when you ask about the future, you can’t be deterministic” (Boscolo et al., 1987, p. 134). Further, “It’s a way of introducing an idea about change without saying they should change” (Boscolo et al. 1987, p. 134).

The technique of circular questioning has been continuously refined since its first formulation. In the 1970s it functioned primarily as an information-gathering procedure: the team used the questions to build a systemic hypothesis on which a “final intervention” — a prescription or ritual — would later be based, and that final intervention was considered “the crux of the encounter between team and family, the only way of bringing about change” (Boscolo & Bertrando, 1993, p. 89). After the team’s 1980 split and the post-1980 absorption of second-order cybernetics and constructivism, the emphasis shifted. Cecchin (1987) reframed circular questions in the register of *curiosity* — questions whose purpose was to “undermine the family belief system that is based on ‘truths’” and so to “create opportunities for new stories” (p. 409). The contemporary CMTF lineage summarises this trajectory: “Questions were no longer requests for linear information but rather tools used to ‘perturb’ the therapeutic system and thus allow differences to emerge” (Barbetta & Telfener, 2020, p. 5). The technique

was also adapted to individual therapy by bringing absent others into the therapeutic dialogue through the patient's descriptions of them (Bertrando, 2002).

By the late 1980s, the team had come to view circular questioning as so effective in its own right that it nearly displaced the formal end-of-session intervention, with circular questioning becoming “interventions in microcosm throughout the interview process” (Boscolo et al., 1987, p. 97). Boscolo and Bertrando (1993, p. 86) attribute this shift directly to the technique's information-generating capacity: “Circular questions draw attention to differences, to new connections between ideas, meanings and behaviours that can change the epistemology (personal premises, unconscious assumptions [Bateson, 1972]) of the various members of the family.” As previously mentioned, future questions were particularly potent:

“By the end of the interview, these future questions often introduce so much new information that they preclude giving the family a task or prescribing the situation or using a positive connotation. The family is free to make what it will of these new ingredients that have been stirred into their system. They will organize themselves in some kind of relation to them, without any formal request from the therapist.” (Boscolo et al., 1987, p. 71)

The contemporary lines of systemic therapy continue to use circular questioning. The systemic-dialogical lineage of Bertrando and Lini (2019) describes their approach:

“We ask many questions during therapy, but we keep them open and nonstrategic: we never ask a question if we think we know the answer. ... Our questions are ways to widen perspectives and to help clients discover options they were not originally aware of, trying to create a context where clients may surprise the therapist — and themselves. Questions acquire in this way a Socratic flavour, where the therapist is aware that she knows that she does not know.” (p. 183)

The contemporary CMTF school also makes their stance clear:

“We believe that ideas must circulate, and that the circularity of questions alone can generate a dimension of complexity that linear questions fail to produce.” (Telfener & Barazzetti, 2025, p. 3)

Examples in the Case Studies

John, the crying boy

- **Category-difference:** Selvini to Barbara: “Who was more preoccupied about this — your mother or your father?” (Boscolo et al., 1987, p. 76)
- **Temporal-difference:** Cecchin to Harry: “So there was a period in which you were close to each other? ... When was it?” (p. 59)
- **Categorical-context:** Selvini to the family: “When John began this behaviour two days before school, what was the explanation you gave for this behaviour?” (p. 47)
- **Reflexive — hypothetical:** Selvini to the parents: “For example, if this strange behaviour had not been exhibited by John but by another child, in your opinion who would have been the other child?” (p. 51)
- **Reflexive — future-oriented:** Cecchin: “If Barbara decides to go to college or leave home, who will take her place in the family?” (p. 69)
- **Reflexive — embedded-suggestion:** Cecchin to John: “Do you agree with them? Do you think your mother is right? I know many kids who don’t agree with this kind of talk.” (p. 64)

Antonella, the sweater girl

- **Category-difference:** Boscolo to Antonella: “In your family, who is your most faithful assistant in the research you do?” (Boscolo & Bertrando, 1993, p. 239)
- **Categorical-context (meaning/action via metaphor):** Boscolo to Father: “Who might have the sugar to sweeten all the bitter pills you’ve had to swallow, and are still swallowing?” (p. 245)
- **Reflexive — hypothetical/future-oriented:** Boscolo to Antonella: “When your brothers get married and start families of their own, which family would you like to

live with, if you really do decide to stay with the family? With Giovanni, with your parents, or with Gustavo? Which is the one you would choose?” (p. 247)

- **Reflexive — embedded-suggestion:** Therapist to Antonella: “Let’s imagine your brothers get married and you go to live with one of them. How do you see your parents, in a few years’ time, on their own... Mightn’t your mother start giving your father a little sugar, I don’t know...” (p. 250)

3.4.2 Positive Connotation

Positive connotation was the signature intervention of the early Milan team. Its definition depends on the specific time period we look at, as this technique underwent significant evolution since its inception. The earliest form, sometimes called the *sacrifice intervention*, framed the symptomatic member as sacrificing for the rest of the family — e.g. “your son decided to stay at home ... because were he to go out and lead his own life, his parents would be left in a state of intolerable loneliness” (Boscolo et al., 1987, p. 15). This often produced rapid symptom reduction, but the team noted that “the recipients of such generosity rebelled angrily and the giver of it ducked to avoid reprisals” — and “the family often vanished from treatment too” (Boscolo et al., 1987, p. 15). It would also implicitly negatively connote the other family members: “It was as if there had to be a cause to pin the blame on, and if one exonerated the symptomatic member, other family members would feel at fault” (Boscolo et al., 1987, p. 7).

In the post-1980 reformulation, the connotation was placed in the service not of one person’s well-being but of a *shared* premise, value, or myth — for example, “in this family there is a strong belief that family members will always come to the rescue when someone needs help. ... The real problem in this family would be if there were no more problems” (Boscolo et al., 1987, p. 16). The vocabulary also changed to what the team called *logical connotation*: a message that no longer claims the symptom is good or needed, only that “people have gotten used to it and that such habits are hard to break ... how meaningful it is in context, how understandable and by now perhaps necessary” (Boscolo et al., 1987, p. 16).

The intervention is typically delivered as a closing message at the end of the session. A canonical worked example is the case of Ernesto, an eight-year-old whose grandfather had recently died and who had begun behaving like “a caricature of a little old man,” hallucinating that his grandfather was following him on family walks. The team’s closing message was as follows:

“We are closing this first session with a message to you, Ernesto. You’re doing a good thing. We understand that you considered your grandfather to be the central pillar of your family ... he kept it together, maintaining a certain balance. Without your grandfather’s presence you were afraid something would change, so you thought of assuming his role, perhaps because of this fear that the balance in the family would change. For now you should continue in this role that you’ve assumed spontaneously. You shouldn’t change anything until the next session, which will be January 21, five weeks from now.” (Selvini Palazzoli et al., 1978, p. 81)

In response, Ernesto “jumped suddenly from his chair” and ran to one of the therapists to protest that his school grades would suffer, that he was “a disaster” at school and might be held back (Selvini Palazzoli et al., 1978, p. 81). By the second session he had given up the old-man behaviour and had begun “taking advantage of school”; the hallucinations of his grandfather had also stopped (Selvini Palazzoli et al., 1978, p. 82).

The refinement of circular questioning largely displaced this technique, as explained in §3.4.1. In contemporary Milan-lineage practice, positive connotation is partly retained and partly abandoned, depending on the lineage. The contemporary CMTF position, articulated by Telfener and Barazzetti (2025), preserves the underlying idea — treating the symptom as adaptive and meaningful — but does so in a reflective, non-manipulative way, and one that communicates acceptance and support: “You stay as you are, and I will be here with you” (p. 7). The Bertrando–Lini systemic-dialogical lineage, by contrast, has moved further from the original technique: in their approach, “most of the technical equipment rooted in the systemic tradition disappears: all directive strategies, most rules of setting, reframings, prescriptions, or rituals” — replaced by the ‘endless flow of dialogue’” (Bertrando & Lini, 2019, p. 177).

Examples in the Case Studies

John, the crying boy

The closing intervention delivered to the Y family is a canonical example of post-1980 positive connotation in its shared-premise form. The team had identified an unspoken family myth — that this second marriage for both parents had to be “perfect” to redeem the bad first marriages — and connoted all five children’s behaviour as serving that myth. The “perfect” children, Harry and Barbara, were doing so through their academic success and good conduct, which let each parent feel they had given the children “a good father” and “a good mother”; the “problem” children, Donna and John, were doing so by giving the parents something to be concerned about, “something more to do for their children, so as to be perfect parents” (Boscolo et al., 1987, p. 87–88). No one was positioned as the cause.

Antonella, the sweater girl

In this second case study, we can see positive connotation in an attenuated form already throughout the second session: the therapist reframes Antonella’s temperature-regulation behaviour using the metaphor of a “researcher” studying body cells — a reframing that “restores her status as a university student” and rescues her from being “branded as sick or deviant” (Boscolo & Bertrando, 1993, p. 240). At the end, the team notes that Antonella’s brothers had been giving the mother a lot of “sugar” (agreement and accommodation), and delivers this positive connotation for Antonella’s symptoms:

“At present, we see that Antonella is trying in some way to get her mother used to swallowing one or two bitter pills. No matter how hard you try, Mrs. Viola, you can’t persuade Antonella to change, so in this sense you’re swallowing bitter pills, at least for the moment. ... These bitter pills are good for you, they’ll make you stronger, immunize you like your husband. Immunity in life can also come from adversity.” (Boscolo & Bertrando, 1993, p. 251–252)

3.4.3 Prescription

Because problems in the Milan view are relational rather than individual (§3.2.1 on general systems theory and §3.2.2 on cybernetics), interventions must act on relationships and feedback loops, not on isolated members. The classical vehicle for this was the *prescription* — a directive given by the therapist or team at the end of a session, telling the family to perform a specific behaviour, sequence, or task between sessions.

Two main sub-types appear most often in the literature. The first is the *paradoxical prescription*, sometimes called *symptom prescription*, in which the therapist instructs the family to continue the problem behaviour itself. This was inherited from the MRI strategic tradition (Weakland et al., 1974) and designed to bypass family resistance: as Haley (1963) explained, “there was no way a family could resist a therapist who told them to continue with the behaviours they were already engaged in unless they gave up those behaviours” (in Boscolo et al., 1987, p. 7). If the family complies, then they are doing deliberately what they had been doing helplessly, and the meaning around the symptom changes. If they refuse, they have to stop. Paradoxical prescription flows logically from positive connotation: if the symptom and its surrounding behaviours are described as meaningful in context, then the family should — at least until the next session — continue exactly what they are doing. The closing message of the Ernesto case quoted in §3.4.2 is a great example of how positive connotation and prescription go together: “For now you should continue in this role that you’ve assumed spontaneously. You shouldn’t change anything until the next session” (Selvini Palazzoli et al., 1978, p. 81).

The second sub-type is the *ritual* — a behavioural prescription with a temporal or sequential structure. Boscolo et al. define it simply as “an ordering of behaviour in the family either on certain days [odd days, even days] or at certain times [after dinner, in the morning]” (Boscolo et al., 1987, p. 4). The canonical ritual form is the *odd-day / even-day ritual*: on odd days of the week, one rule applies; on even days, its opposite; on the seventh day, the family is to act spontaneously (Selvini Palazzoli et al., 1978; Boscolo et al., 1987). A mother who has been simultaneously asked to be “wife to her husband and mother to her daughter,” for instance,

might be directed to be wife on odd days and mother on even days, dissolving the double bind by putting time into the paradox (Boscolo et al., 1987, p. 17).

The technique underwent substantial evolution. By the time of *Milan Systemic Family Therapy* (Boscolo et al., 1987), the team was using prescriptions less and less, for three converging reasons. First, the *sacrifice intervention* form was found to negatively connote the rest of the family and was deliberately abandoned (Boscolo et al., 1987). Second, the team noted that “the entire vocabulary of ‘paradox’ is being used less and less,” partly because paradoxical prescriptions were difficult to deliver sincerely under a second-order frame in which the therapist is part of the system (Boscolo et al., 1987, p. 16). Third — and perhaps most decisively — the refinement of circular questioning increasingly displaced the need for a closing prescription at all, as mentioned in §3.4.1. Furthermore, non-compliance with a prescription, originally read as resistance to be overcome, came to be read instead as information about how the system worked (Boscolo et al., 1987).

In contemporary Milan-lineage practice, prescription has been transformed and made less standardized. The contemporary CMTF states that in their approach, “rituals are not necessarily performed at the end of each session, as in the counter-paradox prescriptions of earlier times” (Telfener & Barazzetti, 2025, p. 14). The Bertrando–Lini systemic-dialogical lineage, by contrast, drops the technique entirely, as already mentioned in §3.4.2 (Bertrando & Lini, 2019). However, a recent case-report literature argues that prescription can be re-delivered in a “more collaborative, less expert-positioned” form (Salamino & Gusmini, 2025).

Examples in the Case Studies

John, the crying boy

The closing intervention to the Y family, after the team delivered positive connotation, was a clear example of a paradoxical prescription:

“Have a meeting together, you siblings, once a week or every two weeks, at a time you choose, and stay together, you five children, and discuss how to continue

this work you are doing.... The parents should be left out because it's a problem the five of you have: to assess if you are successful or not successful." (Cecchin, in Boscolo et al., 1987, p. 88)

We can also see elements of ritual in the prescription through the inclusion of a temporal schedule (weekly or biweekly meetings, at a time the children choose). John, the identified patient, is given a specific role — calling the meetings — repositioning him from the focus of the problem to a manager of the family's collective task.

Antonella, the sweater girl

In the second session, the therapist had closed with the "sugar and bitter pills" metaphor: the mother had been living on a constant supply of "sugar" — agreement, accommodation, and deference from her family — and was at risk of "relational diabetes" once her sons married and that supply dropped. The team prescribed that the sons begin "giving [her] less sugar," weaning her gradually so she would not crash later (Boscolo & Bertrando, 1993, pp. 251–252).

This produced visible improvements but had not resolved the strained mother-daughter relationship, so the third-session intervention took the form of a ritual:

"We feel that there is a strong bond between Antonella and her mother and vice versa, even though they have been disappointed with each other over recent years. We would like to ask you, Antonella, and you, Mrs. Viola, to devote one day a week, Wednesday, to trying to understand each other. On that day, the other members of the family should avoid becoming involved with you, and you should keep how you have spent the day a secret." (Boscolo & Bertrando, 1993, p. 254)

As the authors explain, the ritual is intended to encourage a new relationship between the mother and daughter, through a temporal separation from the rest of the family.

PART 2

A Predictive Processing Analysis of Milan Systemic Therapy

With sufficient background on predictive processing (PP) and Milan systemic therapy (MST), let us now take a look at how MST produces change from a PP perspective.

The purpose of psychotherapy is to produce change. In the predictive-processing paradigm, learning and change are achieved through prediction error, which updates priors and adjusts the generative model (Friston, 2010; Hohwy, 2013; Parr et al., 2022). Therefore, if PP is to successfully explain why systemic therapy works, we must identify how systemic therapy creates prediction errors, and how those prediction errors can lead to updates in priors and in the generative model.

This second part of this thesis offers such an explanation in three steps: a characterisation of the problem system (Chapter 4); a PTPN-framework reading of how the Milan techniques generate prediction error and enable prior updating (Chapter 5); and an extension of the analysis to the multi-agent level, where the priors at issue are distributed across the family rather than held in any one person (Chapter 6).

4. The Problem System

To show how a systemic therapist produces change in a system, we must set out the context of how that system works.

In both case studies presented in this thesis, and certainly in the vast majority of therapeutic cases, there is a presenting problem with which the family comes. The family expresses opposition to this problem and normally openly opposes the problem in an attempt to get rid of it. However, this problem has usually been around for some time — it is not a passing symptom, but something that has reached some form of stability in the system. It exerts some kind of counteracting force against the family's opposition, which keeps it in place. An easy way to conceptualise this is two people both pushing against a rock with equal force, thus making the rock stay in place.

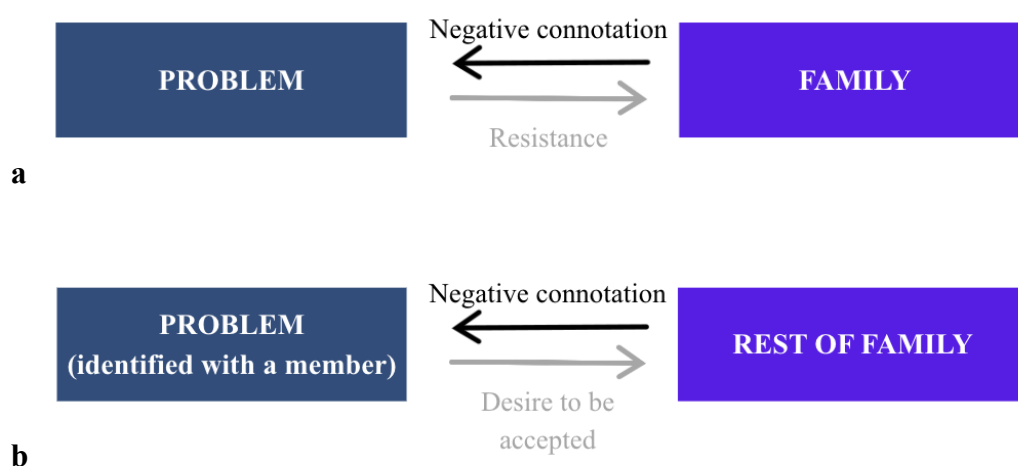
One specific instantiation of the problem dynamic is shown in Figure 12b. In this case, the problem is identified with a particular person (or several). In the Milan literature this person is called the *identified patient* — the family member onto whom the family's difficulty has been concentrated, and on whose behaviour the family's request for help is focused (Boscolo et al., 1987). In the case studies examined in this thesis, John (the “schizophrenic” son) and Antonella (the “anorexic and psychotic” daughter) are the identified patients of their respective families.

Two psychological consequences can be observed from this concentration of the problem on a single person. First, as was discussed in Chapter 3, the family describes the problem in dispositional terms — as a property of the person, rather than as a property of the relational pattern that produced it. Second, the identified patient comes to internalise the label as part of who they are. PP accounts of self-concept formation argue that the self is constructed and maintained through priors — beliefs about what kind of person one is — that the brain treats with unusually high precision (Mokady & Reggev, 2022). Mokady and Reggev (2022) make the strong claim that, in contrast to the standard PP assumption that priors are updated by prediction error, self-relevant priors *resist* updating: they are precision-elevated stable constructs, and incoming evidence is preferentially used to confirm them. Fisher and Hohwy

(2024) make a closely related argument from the free-energy-principle side, showing that self-evidencing systems are constitutively biased toward outcomes congruent with their preferred-state priors — what they call the universal optimism of the self-evidencing mind. Both accounts converge on the same point: priors that constitute the self are protected, not exposed, when they are challenged.

Figure 12

An illustration of the homeostatic problem



This protection becomes important when an identified patient is asked to change. Rationally, the person being labelled may agree with the description of his behaviour — Antonella expresses opinions that would seem to accept her labels at face value. However, humans also have a fundamental desire to be accepted and to maintain a coherent sense of self (Steele, 1988; Cohen & Sherman, 2014). Being identified with a negative trait often causes them to become defensive and show opposition. Self-affirmation theory (Steele, 1988) and three decades of subsequent research (Cohen & Sherman, 2014) have documented this dynamic in detail: when a person’s self-integrity is threatened, the response is to reject the threatening information, derogate its source, or counter-argue.

This is the central paradox of the problem system. Because John may identify with the problem as part of who he is, he subconsciously pushes back against the family’s opposition, thereby protecting the problem’s existence. The same dynamic appears throughout the

empirical psychology literature: psychological reactance theory (Brehm & Brehm, 1981) shows that direct attempts to restrict a person's freedom of behaviour produce a motivational state that resists the restriction; ironic-process theory (Wegner, 1994) shows that direct attempts to suppress a thought or behaviour increase its accessibility. The implication is that the problem cannot be removed by pushing harder against it. To produce change, the therapeutic intervention must be designed in a way that does not threaten the system's organisation but instead creates conditions under which the system can update from within.

5. Milan Systemic Therapy from a Predictive Processing Perspective

This chapter develops the central claim of the thesis: that the interventions of MST can be re-described as a coordinated set of mechanisms for generating prediction error against the problem-system priors identified in Chapter 4, while simultaneously creating the conditions under which the patient's generative model can update from within. The argument is summarised in Figure 13.

First, the therapist's interventions introduce a substantial amount of information into the system, aimed at dislodging the negative connotation of the problem and its identification with the associated person (the red portion of Figure 13). A breakdown of the new information added to the system is offered in §5.1.

These interventions and the new information delivered through them produce prediction error (shown in yellow in Figure 13), leading to an update in priors that in turn drives further revision to the negative connotation (shown in purple in Figure 13). This step is crucial: therapists may attempt to convey new ideas or ways of thinking to their patients, but new information does not necessarily result in learning, in an update of priors, or in a change of the generative model issuing in new behaviour. A framework for why the present analysis argues that MST interventions do produce this change is provided in §5.2.

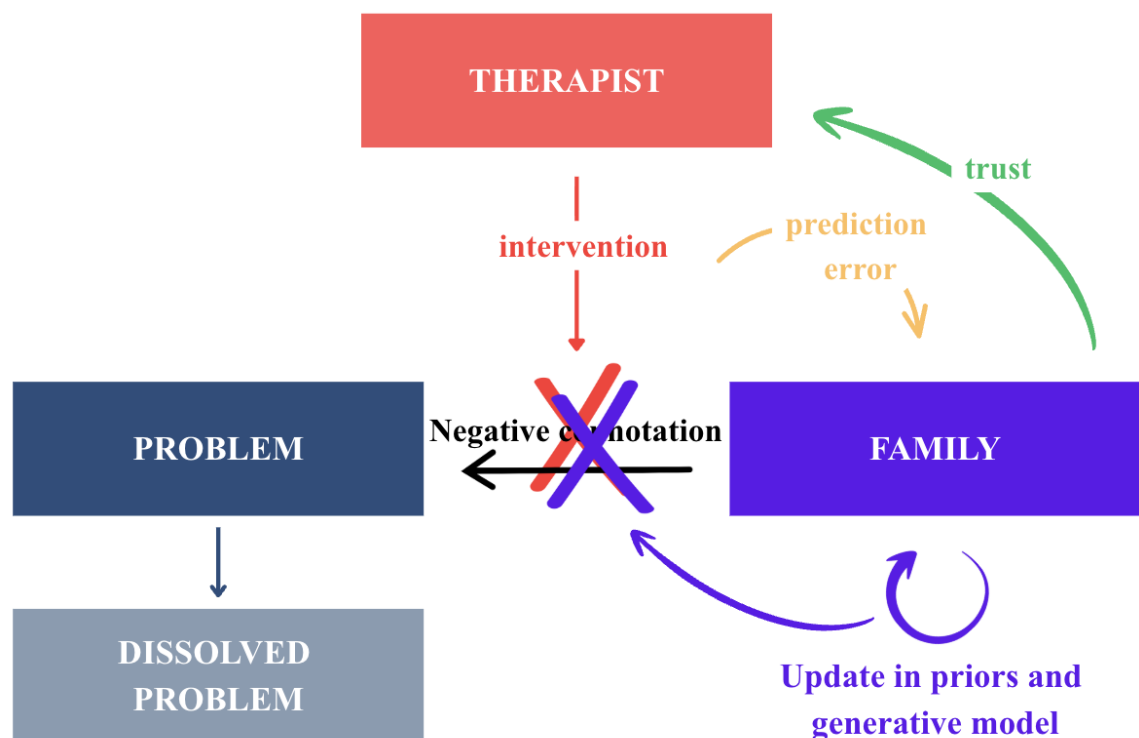
One central part of the aforementioned framework for enabling learning is *trust* (shown in green in Figure 13). The family arrives at the therapy session with some level of trust already,

but the interventions delivered by the therapist likely increase the family's trust across the course of the encounter, thereby further increasing the precision of the prediction error delivered through the interventions. An elaboration of this mechanism is provided in §5.2.2.

Lastly, the problem gradually changes form over time and transforms (shown on the left side of Figure 13). As priors update across the system, the configuration that held the problem in place loses coherence, and the problem ceases to exist as such (see §3.2.4 on Maturana's structure–organisation distinction). This transformation is further supported by the multi-agent character of family therapy, and is explored in Chapter 6.

Figure 13

A big picture view of Milan systemic therapy from a predictive processing perspective



5.1 How Systemic Therapy Delivers New Information

The main interventions of systemic therapy, as examined in this thesis, are circular questioning and positive connotation. Before getting into the specifics, it should be noted that the mechanism by which circular questioning shifts thinking is, in the Milan team's account, *cumulative*. It is not necessarily a single question or answer that produces a discrete prediction error or update in priors (although that is theoretically possible). Rather, it is the accumulated juxtaposition of answers — what Boscolo calls “gravestone rubbing” (Boscolo et al. 1987, p. 18).

5.1.1 Questions Themselves Create Prediction Error Through New Information

The question itself perturbs the family's model even before any content is exchanged. As Cecchin explains, “When you ask a question you introduce a difference. The answer is secondary because you have already introduced something new” (Boscolo et al. 1987, p. 110).

One specific way a question does this — regardless of the answer — is by signalling what the therapist treats as important. Clients come to therapy expecting the therapist to help with their problem; they therefore assume that whatever the therapist asks about must be something the therapist believes is somehow connected to the problem. In PP terms, the client is doing theory-of-mind inference about the therapist's intentions behind the questions (Frith & Frith 2012; Diaconescu et al. 2014). Karl Tomm makes essentially the same argument in his *interventive interviewing* framework: every question is also an intervention, because it communicates the interviewer's hypothesis about what is relevant to the problem (Tomm 1987).

Boscolo gives this example in their 1987 book: “We tend to avoid talking about the extended family in the beginning of the first session. We start from the center and go out. If, in the beginning, we ask questions about the grandfather, the grandmother, these questions imply historical causality, that there is some connection between the problem and their grandmother and their grandfather” (Boscolo et al. 1987, pp. 110-111).

As the clients takes in this implicit information, it may raise the probability of the associated idea in their generative model's posterior distribution. This in turn may shape what associations come to mind for subsequent questions and how the clients answer them. It also shapes how they perceive and interpret the people involved in the problem in the days that follow. This is a familiar finding outside therapy: classic work on leading questions has shown that the wording of a question can reshape what people later recall about events they witnessed (Loftus & Palmer 1974). The PP version of the same point is just the general principle that top-down expectations constrain what is perceived (Clark 2013); recent work on how priors specifically shape social perception is developed by McGovern & Otten (2024).

5.1.2 Circular Questioning Reframes the Problem as Relational

Circular questions presuppose that a behaviour is distributed across the family, that the problem exists within relationships rather than in any one person (see §3.2.1 on general systems theory and §3.4.1 on circular questioning). Asking a circular question therefore implicitly asks the family member to start thinking of the problem as relational (even if only temporarily), as they cannot produce a coherent answer without considering the relational frame.

This change in the family's generative model is gradual. A single relational question is unlikely to produce enough prediction error to shift the model, because the original frame — “the problem is John” — is typically held at high precision, and PP predicts that high-precision priors move only a small amount per observation (Mathys et al., 2014). What does the work is the cumulative effect of many questions. Each circular question is diagnostic: the answer it elicits is more probable under a relational reading of the family than under an individual-pathology reading, so each question contributes Bayesian evidence in favour of the relational frame. The family member is also not just being exposed to this evidence — they are producing it themselves. The generation effect (Slamecka & Graf, 1978) shows that self-generated information is more strongly encoded than information received passively, and Bem's (1972) self-perception theory shows that people infer their attitudes from observing their own behaviour. Therefore, when a family member answers ten or fifteen relational

questions in sequence, they are simultaneously receiving repeated diagnostic evidence and watching themselves operate inside the relational frame.

The family member may not register this consciously — they are probably not thinking “I see, the problem might be relational.” Implicit-learning research shows that people can extract complex statistical patterns from repeated exposure without being able to articulate what they have picked up (Reber 1989). What they do register is a metacognitive feeling that something has shifted. PP gives a formal account of this experience: metacognitive feelings (tip-of-the-tongue, the feeling of knowing, the sense that something is off) arise from changes in the dynamics of prediction error, before the agent can articulate what is changing (Fernández Velasco & Loev 2025). This kind of pre-articulate sensitivity is well-attested outside therapy too — people can judge that a set of words is semantically coherent before they can retrieve the common associate (Topolinski & Strack 2009). In the consulting room, this is what allows the fixed-label thought to begin feeling somehow wrong even if nobody can explicitly explain how.

For example in the Crying Boy, the therapists asked this series of questions when the family complained that John had suddenly become cooperative, where previously he had not been:

- “Who was cooperative before he became cooperative?”
- “But you said during the summer he was more cooperative. What about Harry, was he cooperative?”
- “Was he more cooperative than John?”
- “What about the girls? Are they cooperative?”
- “Who’s the less cooperative?”
- “Who? The girls, you say.”
- “More so, more cooperative? More cooperative with you?”

Earlier in the session, the parents had described John’s recent cooperativeness as something puzzling — a property they had not expected of him, and which they were treating as part of his disturbance (“How do you go to a doctor and say ‘There is something wrong with my son, he’s doing what he’s asked?’”, p. 43). The label “cooperative” was anchored to John in a way

that made his having it feel like another symptom. Through the seven questions above, Cecchin and Selvini redistribute the label across the family. “Cooperative” is then no longer a property belonging diagnostically to John but a quality that varies across all five children. When the parents return to the fixed-label thought “John is cooperative now, and that’s a problem,” there are many associations around that idea that create the feeling that something is off about that idea, even if they cannot explain exactly what it is.

5.1.3 Circular Questioning Introduces Multiple Perspectives on Reality

Circular questioning asks each person to share their opinion and perspective on different issues, bringing up the idea that each family member has their own reality.

Patients typically come into the session with the assumption that there is a single objective reality, shared across the family. For example, John’s parents told the therapists that John was the problem, treating this as an objective truth. However, the therapists asked a series of questions implying that other family members may see things differently. Examples include:

- “Everybody said that this was a problem. Do you agree?”
- “How do you see it? I asked ‘What’s the problem?’ They said ‘Oh, it’s John.’ So they say you’re a problem. Do you see it differently?”
- “How do you see it? As a problem, or is it something else?”
- “Who was more preoccupied about this — your mother or your father?”
- “They think the same way because they have agreed to think the same way, or do you think they really think that way?”
- “Now I want to ask John, how do you see your situation? How do you see this episode in September that scared everyone? You went to a hospital and a diagnosis was made. How do you understand that situation now?” (Boscolo et al., 1987,)

The very fact that these questions are asked of different family members introduces the idea that everybody may have a different opinion, and that all opinions are equally important. Even the labeling of the event as a problem is open to debate, as the first three questions show.

This relates to the concept of constructivism discussed in §3.2.5, and is very compatible with the perspective of PP. Within PP, perception is not a copy of the world but the brain's best inference about hidden causes, generated by a generative model that has been shaped by the individual's own history of observations (Clark, 2013; Hohwy, 2013). Two family members looking at the “same” situation are not running the same model; they are each running a separate model, with separate priors, on the data each of them happens to attend to. From this view, disagreement between family members is not a question of who has the facts right, but of which generative model is being used.

Similarly to the idea in the previous section, as several of these questions are asked throughout the therapeutic session, the family members present implicitly absorb the information that there may be many viewpoints within the family system.

5.1.4 Answers as a Further Source of Prediction Error

As the family members answer these questions, their answers may be genuinely surprising to the others. For example, consider the second-last question listed in §5.1.3. Suppose that the mother and father did not really have the same opinion, but were trying to present a united front in front of the children, and believed that they were succeeding in this. If one of the children answered that question by saying that the parents did not in fact have the same opinion, it would become clear to the parents that their efforts were not as successful as they thought. This further reinforces the new information discussed in the previous section: that other people do not necessarily have the beliefs that a person thinks they do.

5.1.5 Circular Questioning Widens the Space of Possible Outcomes

Families typically arrive with what Boscolo describes as a deterministic frame: “The family usually has the sensation of being stuck in an unending game. They have the idea that there is no way out” (Boscolo et al., 1987, p. 134). In PP terms, this is a generative model with a very high-precision prediction that the same situation will continue, with very few alternatives and little variance. The therapeutic problem is therefore not to prescribe a new state — which the model would see as highly unlikely and would probably not act upon — but to raise the precision on the possibility of difference.

Future questions, which were described in §3.4.1, are “a way of introducing an idea about change without saying they should change... it’s a way of saying that there could be a difference” (Boscolo et al. 1987, p. 134). To answer the question, the family member has to generate a hypothetical state — a future configuration of the family that does not match the current one. The act of producing this simulation, even tentatively, is itself an addition to the model: the alternative state has been represented and is now available as a referent for subsequent inference. Decades of work in social cognition show that recently activated constructs remain more accessible and bias how subsequent input is interpreted and categorised (Higgins 1996). The same dynamic is visible in semantic priming, where processing one concept reliably facilitates the processing of related ones (Meyer & Schvaneveldt, 1971). The future-tense question, on this reading, is doing something structurally similar to a prime — it raises the precision of a representation (here, an alternative future configuration of the family) that the model will then carry forward into the next moments of the session and potentially into subsequent family interactions.

In PP terms, two mechanisms can be offered for how these changes in precision occur. The first is precision-weighting through attention. On standard accounts of the active inference framework, attention is the optimisation of the precision assigned to prediction errors generated under a particular hypothesis (Feldman & Friston 2010). When the family member is asked a future question, the alternative state has to be attended to in order to be generated as an answer at all; attending to it raises its precision, and a hypothesis with raised precision may be more readily available and selected in subsequent inference. The second possible mechanism operates from the opposite direction: rather than raising precision on an alternative hypothesis, future questions may lower precision on the dominant hypothesis. Before the future question is asked, the family’s generative model is dominated by a single high-precision hypothesis — that things are as they are and will continue to be. Few alternative possibilities in the model. Entertaining even a tentative alternative spreads probability mass across additional candidate states; the variance of the posterior over possible futures widens, and the precision concentrated on the previously dominant hypothesis is correspondingly diluted.

5.1.6 Positive Connotation Delivers a Non-Threatening Reframe

Where circular questioning works by delivering many small perturbations to the family's model across the course of a session, *positive connotation* (described in §3.4.2) delivers a single coherent reframe at the end of the session. This shift in form — from many small perturbations to a single organizing one — gives positive connotation a distinctive role in the predictive-processing account of how MST produces change.

The circular questions “prime” the new ideas, so that the positive connotation is more readily processed and accepted. This priming claim has two complementary readings within PP.

The first is a familiar accessibility-of-constructs effect (Higgins, 1996; Meyer & Schvaneveldt, 1971). By the time the closing intervention is delivered, the family has spent the session generating answers that are coherent with a relational, distributed reading of the problem. The relevant constructs — the symptom as a function of family organization, the patient's behaviour as connected to the parents' history, the label as a description rather than a fact — have all been activated. The positive connotation arrives into a generative model in which these constructs are already raised in precision. A closing message that integrates them lands as a coherent organizing hypothesis rather than as an unfamiliar imposition.

A third feature is worth noting, because it explains why positive connotation works where direct confrontation fails. As argued in Chapter 4, self-relevant priors are precision-elevated and resist updating when they are challenged directly (Mokady & Reggev, 2022; Steele, 1988; Cohen & Sherman, 2014). A connotation that says “the symptom is the patient's failure” or “the family is the cause” puts pressure on exactly these priors and provokes the autopoietic defense described in Chapter 4. Positive connotation circumvents the defense by not making the symptom or any family member's behaviour the target of a negative description in the first place. The reframe is therefore receivable: the model can update without first having to defend itself.

5.1.7 The Intervention as a Whole Functions as a Surprise

The intervention as a whole is quite surprising to most clients in the affective sense of the word: the therapist's interventions act to destabilize the negative label that is applied to the problematic family member, thus disrupting the homeostatic force keeping it in place. These are usually unexpected by the family: the family came to therapy in order to solve a problem and therefore expected the therapist to treat the problem as a problem and join the family in exerting more pressure against it.

Instead, the therapist does not adopt the family's frame of the problem (see §3.3.3 on non-normativity). The questions presuppose that the difficulty is relational rather than individual; the closing message connotes the symptom as something that serves the family rather than something to be eliminated; and the family member who was identified as "the problem" is treated as a participant rather than as the cause. Each of these moves contradicts the family's expectations about what therapy is and what the therapist will do.

Therefore, these interventions are surprising to the family in the literal sense, and arguably create a strong prediction error at the level of a high-order belief: specifically, the family's generative model of the therapist's intentions and of the purpose of therapy itself.

This may be especially consequential because surprise is a known modulator of memory encoding. There is now a substantial body of work showing that events that violate prior expectations are preferentially encoded into episodic memory (Greve et al., 2019; Quent et al., 2021; Henson & Gagnepain, 2010). The mechanism, on contemporary predictive-processing accounts, is that prediction error functions as a learning signal not only for online perception but for the consolidation of episodic traces: when an event is more surprising than the model predicted, the hippocampal–dopaminergic system flags it for enhanced encoding, increasing the probability that the agent will retain and later be able to retrieve it (Lisman & Grace, 2005; Kafkas & Montaldi, 2018). Emotionally charged events show the same enhancement (McGaugh, 2000), and surprise in therapy carries both cognitive and emotional weight at once.

The implication is that the moments of the session in which the therapist's interventions most violate the family's expectations are also the moments most likely to be remembered. The relational reframe, the positive connotation, and the closing prescription thus enter the family's memory not as ordinary content but as marked, accessible content — available for reactivation in the weeks and months after the session, and therefore available to shape subsequent reinterpretation of the family's situation. The intervention's lasting effect on the family's generative model is supported, on this account, not only by the in-session updating described in the preceding sections but by the privileged encoding that surprise affords.

5.2 How Milan Systemic Therapy Creates Prediction Error That Updates Priors

As discussed in §5.1, during a therapy session the patient may receive a lot of new information, but that does not mean it will update the generative model. If every new piece of information triggered an update, generative models would be perpetually unstable; in practice, models are biased toward preservation and revise only selectively (Mathys et al., 2014; Sharot et al., 2023; Hauke & Lohr, 2022).

As therapists are agents of change, the key question for them is: what distinguishes information and interventions that update the patient's priors and generative models from those that do not?

This thesis proposes that, for a therapeutic intervention to produce the desired update in a patient's generative model, four conditions must be minimally satisfied. The four conditions give the framework its name (PTPN):

PTPN: The Four-Factor Therapeutic Prior Update Framework

1. **Perturbation:** The information must perturb the system enough to change it, but not enough to break it.
2. **Trust:** The information must come from a trustworthy source.
3. **Preference:** The information must be favorable with respect to the patient's phenotype-derived preferences over outcomes.

4. **Non-confrontation:** The information must be delivered in a form that does not directly challenge the patient's self-relevant priors.

Each of the four PTPN components has been examined separately in the predictive-processing and belief-updating literatures, but to the author's knowledge no existing model combines them as a set of conditions specifically governing when a therapeutic intervention results in a desired update of priors.

The following sections examine each component in turn.

5.2.1 Perturbation

The first condition for information to create a prior update is that the incoming information must function as a perturbation: a novelty that contradicts what the existing model predicts. Without anything novel, no change can occur — the existing model is simply reconfirmed, and the system's generative inference proceeds unchanged.

This requirement has been demonstrated empirically across several largely independent literatures. In the neuroscience of memory reconsolidation, Pedreira et al. (2004) provided the foundational experimental demonstration that mismatch — prediction error between what is expected and what occurs — is a necessary condition for a stored memory to be updated rather than re-confirmed. Subsequent work from the same lab (Fernández et al. 2016) consolidated the finding into a unified framework: a stored memory enters the labile state in which it can be updated only if its reactivation is accompanied by a discrepancy between predicted and observed outcomes. Ecker (2015), translating this neuroscience into clinical language, specifies further that the mismatch must be contradictory to the existing memory, not merely different. A novel-but-compatible experience confirms the existing model; a novel-and-contradictory experience destabilises it. Sinclair and Barense (2019) extend the requirement on the magnitude axis: partial mismatches — incomplete reminders that only partly replicate the original encoding context — are the most reliable triggers of destabilization. Total mismatch tends to be parsed as a different category of event entirely, generating a new memory rather than updating the old one.

Literature in psychotherapy offers insights on the magnitude of error needed for a belief update. Chamberlin (2019), in his predictive-processing model of EMDR, explicitly names a *Prediction Error Window* as a condition for memory update: prediction error must be in a moderate range — too small, no update; too large, a new memory is created. This is also the principle behind Hauke and Lohr’s (2022) account of CBT in free-energy terms: questioning a prior triggers a short-term free-energy increase (i.e. stress), and beyond a certain threshold the patient will close down rather than update. A parallel formulation appears in adult learning theory: Wergin (2019) calls the same phenomenon *constructive disorientation* and explains that for transformative learning to occur, the disorientation “must be strong enough to encourage suspension of assumptions and beliefs and the willingness to entertain fresh perspectives, but not so strong that a response to the preexisting order is triggered, as well as various immunities to change” (p. 70).

Examples in MST

In both case studies presented in this thesis, the therapist does not argue against anything that any person has said. Such direct contradiction would likely be rejected: the family members’ own experiences carry high precision relative to assertions from a therapist who has just met them and was not present during the described interactions. Instead, the therapist elicits information from the family members themselves that is related to the belief, and adds more context which may change its meaning.

In *The Crying Boy*, the therapist does not attempt to deny that John is jittery, nor to argue that being jittery is not so bad after all. Rather, the therapist asks the family “who else is jittery besides John?” This does not invalidate the proposition that John is jittery; it adds the information that other family members are jittery too, which reframes jitteriness as something not unique to John or “his problem” but as something also present in family members not conceived as problematic. This contextual addition shifts the meaning the family ascribes to jitteriness and the role they perceive it as playing in the presenting problem.

The final intervention — positive connotation paired with paradoxical prescription — extends this same logic. The therapists do not attempt to persuade the family to see the problem as

good; such a reframe would fail to integrate with the family's generative model, which clearly labels the problem as bad. They concede explicitly that John's behavior creates problems, but they situate this problematic behavior in a larger context in which it serves to maintain the family system. The family's existing beliefs remain intact alongside the new information, so the new information can be integrated into the generative model without triggering resistance.

The second case study follows the same pattern. As a particular example, when Antonella describes her frustration and distress with her problem, the therapist neither denies that this is what she does nor argues against it. Instead, he comments, "you're kind of like a researcher." This comment fulfills Ecker's novel-and-contradictory requirement, because viewing the behaviour as research rather than a problem obsession puts a positive interpretation on it which contradicts the previous view of it being a problem or obsession. Yet, it is a partial mismatch because the therapist is still talking about the same behaviour, and this interpretation is possible for such behaviour. Moreover, the therapist never says that this "research" carries only positive consequences or is free of stress; the new frame remains compatible with the family's lived experience of distress.

5.2.2 Trust

The second condition for therapeutic prior update is that the information must come from a source the patient regards as trustworthy. More formally, trust amplifies the precision of incoming evidence so that it actually drives belief update rather than being discounted; distrust attenuates evidence so that even high-quality information passes through the system without leaving a trace.

The formal mechanism is well developed in the PP literature. The Hierarchical Gaussian Filter (HGF; Mathys et al., 2014) is a Bayesian model of how an agent updates its beliefs in a noisy, uncertain world. Its central insight is that a prior is revised in proportion to a precision ratio — the precision of the incoming evidence divided by the sum of the precisions of evidence and prior. The same prediction error, weighted by a higher precision, drives a larger update; weighted by a lower precision, a smaller one. Diaconescu et al. (2014) extend the HGF into the social domain, demonstrating formally that observers infer the intentions of advisers to

estimate their trustworthiness, and then weigh advice from those advisers according to the precision they assign to the advisers' inferred intentions. Hauke and Lohr (2022) show that the same mechanism operates in the therapeutic relationship: trust in the therapist buffers the short-term free-energy increase that questioning a prior triggers. Without trust, the patient withdraws and the perturbation has no purchase on the model; with trust, the patient remains in exploration mode long enough for the new information to update the model.

Empirically, source effects on belief update are large. The classic Hovland and Weiss (1951) experiments showed that identical messages produced substantially more attitude change when attributed to a high-credibility source than when attributed to a low-credibility one. The effect is not one of persuasion at the moment of message receipt; it shapes the interpretation of the message itself. McGovern and Otten (2024) review the PP literature on how prior knowledge about social categories shapes intergroup perception not just at the level of judgment but at the level of perception — the same behaviour observed in an in-group versus an out-group member is perceived differently at the perceptual level. In therapy, this means that the same intervention will be interpreted differently depending on the patient's prior beliefs about the therapist; reputation, credentials, referral source, and the perceived warmth and competence of the therapist all act as priors that modulate the perceived precision of the therapist's contributions. A substantial portion of trust in a therapist is, in this sense, inherited: it comes from what the patient has heard about the therapist from other sources whom the patient already trusts. Reputational priors shape how the patient interprets the therapist's statements long before direct evidence about the therapist's reliability has accumulated. The same statement, made by a therapist the patient has been told is exceptional versus by one the patient suspects is mediocre, will be parsed quite differently — not just evaluated differently, but registered with different precision and thus given different weight in belief update.

This second condition is partially supported by the third (the information being favourable to the patient's preferred outcomes). There is robust evidence that people extend more trust to sources whose communications are positively valenced towards them. Aronson and Linder's (1965) classic *gain-loss theory of attraction* demonstrated that people feel greater attraction to evaluators who praise them than to those who do not. The in-group favouritism literature (Brewer, 1999; Tajfel & Turner, 1979) shows that trust is preferentially extended to those who

share the perceiver's beliefs or group memberships — sources who confirm the existing model are easier to trust than sources who challenge it. Sharot and Garrett (2016) and Sharot et al. (2023) have shown that good news is incorporated more readily than bad news, suggesting a more general principle that favourable information has privileged access to belief update; by extension, sources who consistently deliver favourable information are likely to be afforded higher precision over time. This creates a self-reinforcing loop in therapy: a therapist who can deliver some favourable information early in the alliance is more easily trusted, which raises the precision of subsequent contributions.

Finally, the second condition must be extended in one further direction: the patient's own senses are themselves a source of information, and one to which the system is built to assign very high precision. In PP, sensory evidence — whether exteroceptive (sight, sound, touch) or interoceptive (bodily sensation, gut feeling, emotional tone) — is generally afforded high precision under healthy conditions (Feldman & Friston, 2010; Friston, 2010; Fotopoulou & Tsakiris, 2017). This has an important consequence for therapeutic interventions: information offered by the therapist cannot be in open contradiction with what the patient directly perceives or feels, because the patient will weigh direct perceptual evidence heavily against the therapist's contribution. A therapist who insists that the patient is safe in a context where the patient's body is sending strong threat signals will fail to update the patient's prior; the body's signal will dominate. This is one of the reasons trauma-informed therapies emphasise *somatic* attunement — the therapist works *with* the patient's interoceptive signals rather than against them, treating those signals as a high-precision source of information about the patient's state that must be respected before any therapist-supplied reframing can take hold (Fujimoto, 2026).

Examples in MST

The Milan team's interventions were widely seen by contemporaries as gaining their power partly from the team's reputation and charisma: Ugazio (1985, p. 31), writing from within the Milan Center, noted that the team's classicB interventions had “the character of a cryptic message, suited to a therapist with a peculiar charismatic power.” Boscolo and Cecchin themselves describe a case in which a school psychologist attempted exactly the kind of

positive-connotation intervention the Milan team had used to dramatic effect with a comparable family. Where the Milan team's intervention had led to the resolution of the problem almost immediately, the school psychologist's version was met with institutional ridicule, and Boscolo and Cecchin note in the same chapter that trainees in lower-status positions face this difficulty as a matter of course: "the families do not easily connect with them and often abandon treatment to find other, more prestigious, therapists" (Boscolo et al., 1987, p. 23). Additionally, many textbooks and handbooks describe the original team as highly influential and charismatic (Bray & Stanton, 2009; Carr, 2016; Wampler et al., 2020).

It is highly likely that the families in both case studies knew of the Milan team's good reputation and entered the consultations ready to regard the therapists' opinions as trustworthy. The referral context in each case provides some support for this. In the first case study (Boscolo et al., 1987), the Y family was brought in for the consultation by a female psychologist who had asked the Milan team for help "regarding the diagnosis and also the treatment plan" because Mrs. Y "had some questions in that regard, especially around medication" (p. 39). The family thus arrived through a trusted professional intermediary, and the consultation itself took place at a conference where the Milan team's status as invited experts would have been salient. Since seeking advice typically presupposes some prior trust in the source, the family's willingness to follow this referral suggests they entered the session with at least some baseline trust in the team. In the second case study (Boscolo & Bertrando, 1993), the Viola family was referred to the centre "by a psychiatrist who had been treating their daughter Antonella for three years" (p. 235) — a clinician with whom they had a long-standing relationship and who would not normally have made the referral without expecting positive results from the consultation. The first session itself appears to have inspired further trust: three months later the mother reported that Antonella had dismissed her individual psychotherapist of a year and a half "because of 'lack of results'" (p. 238) and the family requested another family-therapy session.

The sessions with both families are likely to have reinforced this trust further, since the Milan therapists' positive connotation — discussed in the preceding section — placed

every family member’s behaviour in service of the family system rather than singling out any individual as pathological.

5.2.3 Preference

The third condition for therapeutic prior update is that the new information must be favourable with respect to the patient’s preferred-state priors. Broadly, this means they must be supportive of their well-being, self-esteem, sense of security, and survival-relevant goals. Priors can of course be updated in directions that are unfavourable to the organism — “I am a bad person,” “nobody likes me,” “the world is dangerous” — and patients often arrive in therapy precisely because such updates have already occurred. Updates in the negative direction generate stress, negative affect, and a kind of physiological and psychological instability; they are, in the PP sense, surprising relative to the deep priors a healthy organism is built to maintain. Therapeutic updates that move towards *favourable* priors are therefore not only ethically preferable, they are the ones the patient’s system is structurally biased to accept.

This bias has its formal home in the active-inference machinery of the free energy principle. As elaborated in §2.5, Friston and colleagues distinguish variational free energy (the surprise of what just happened, minimised by perceptual inference) from expected free energy (the projected surprise of possible futures, which selects which action to take). Expected free energy decomposes into epistemic value (information gain) and pragmatic value (movement towards preferred outcomes), where preferred outcomes are encoded formally as priors over observations — sometimes called phenotype-defining priors, prior preferences, or the “C matrix” in the POMDP formulation (Friston et al., 2015). These preferences are not learned within the lifetime of the organism but built into its architecture by selection pressure: a mammal expects, among other things, body temperature within a narrow range, oxygen saturation above a threshold, varied sensory input, and maintained social bonds (Barrett et al., 2016; Friston, 2010, 2013). Fisher and Hohwy (2024) make this connection particularly clearly: optimism bias, they argue, is not a cognitive bug but a foundational tenet of *self-evidencing* — the formal expression of phenotype-defining priors being precise and biased towards outcomes congruent with the agent’s preferred states. Mokady and Reggev (2022) extend the same logic to the self-concept: stable self-relevant priors cultivate

belief-maintaining rather than belief-updating dynamics, with evidence confirming the predicted self-triggering subjective reward responses.

The social-psychology literature on human needs and motivations describes the same phenomenon in a different vocabulary. Humans have strong, well-documented preferences for being seen as good, being liked, belonging to social groups, and maintaining a coherent and positive self-concept (Baumeister & Leary, 1995; Ryan & Deci, 2017; Steele, 1988). Self-affirmation theory in particular (Cohen & Sherman, 2014; Steele, 1988) demonstrates that the perception of *threat to self-integrity* is a major source of defensive processing — pushing harder against a self-image-threatening message tends to rather than dislodge it. Conversely, brief affirmations that protect self-integrity reduce defensiveness and increase openness to challenging information. The mechanism is consistent with the PP account: information that contradicts the patient's preferred-state priors is treated as a threat and triggers active inference designed to defend the existing model (in the social register: defensiveness, reactance, dropout); information that is consistent with — or actively supportive of — the preferred-state priors lands more easily.

One important caveat must be added. There is a well-documented exception to the general rule that humans prefer favourable feedback — Swann's *self-verification theory* and the *cognitive-affective crossfire* (Swann, 1990; Swann et al., 1989). People who hold negative self-views tend to seek out feedback that confirms those views and to surround themselves with people who see them negatively. At first glance this appears to contradict the claim that humans prefer favourable information. However, the negative-feedback preference is *cognitive* — it satisfies the desire for psychological coherence and predictability — while *affectively*, the same people still prefer positive feedback and feel better when they receive it. Everyone prefers favourable information — but only insofar as that information does not contradict an existing high-precision prior. Where the favourable update would require overriding such a prior, the cognitive system rejects it

The key, then, is in not violating the first condition of the PTPN model: the update must not contradict the patient's own experience or a strongly held prior. For example, if a patient experienced themselves doing something harmful — say, deeply hurting another person —

and feels guilty about it, a therapist who insists that the act was “actually a good thing” is offering a favourable framing that the patient cannot accept even if they would like to, because it would require overriding a high-precision emotional memory, and the attempt is likely to be experienced as the therapist failing to understand. This is the same constraint that appeared in §5.2.2 (the patient’s own perception and embodied experience function as very high-precision sources), now applied specifically to emotional memory: the contents of episodic emotional memory are also priors, often held with very high precision, and a new favourable prior cannot simply overwrite them. Ecker (2015), in his account of clinically translating memory reconsolidation, makes the related point that destabilisation requires experiential contradiction, not merely declarative reassurance. A therapist who tells the patient their guilt is misplaced is not offering experiential contradiction; a therapist who helps the patient encounter a context in which the guilt no longer organises their behaviour is.

Examples in MST

As the therapist works to displace the negative label from the problem and dissolve its identification with a particular person, room opens for more positive views to enter the picture. If John is not the only one in the family who is jittery, but it is a characteristic that several family members exhibit, then saying “John is jittery” no longer makes sense as a description of the problem. He is relieved, even if just slightly, from the role of being the harbinger of the family’s distress. Furthermore, if his behaviour serves an important function for the family, then John is not simply creating unpleasantness for everyone around him: he is doing something the family system needs. The implication is that John cannot be all that bad, which opens up the possibility of a more positive, or at least less intensely negative definition of him.

As explained in §2.5, humans are biased towards states favourable to themselves, including being liked and holding a positive self-appraisal. The prediction error produced by the reframe is therefore more likely to be accepted by John than a direct correction would be. At the same time, none of the other family members are negatively connoted, and none of their experiences are invalidated, so they are unlikely to have cause to resist the new information either.

5.2.4 Non-Confrontation

The fourth condition is that the new information must be delivered in a way that does not threaten the patient's sense of agency or behavioural freedom. Direct instruction, persuasion, and confrontation tend to undermine therapeutic prior update rather than support it; the new information lands more reliably when the patient experiences themselves as arriving at it, or at least as free to accept or reject it. This is the difference between *prescribing* a change and *creating the conditions* under which a change becomes possible.

The clearest empirical case for this condition comes from the motivational interviewing (MI) literature. MI is built around the principle that confrontational and directive approaches to behaviour change are systematically less effective than non-confrontational ones; its central techniques — reflective listening, eliciting the patient's own arguments for change, rolling with resistance, supporting autonomy — are designed precisely to keep the patient in the role of agent rather than recipient of instruction (Miller & Rollnick, 2013). Meta-analyses across more than a hundred trials show MI to be efficacious across a range of presenting problems, including substance use, health behaviours, and treatment engagement (Lundahl et al., 2010), and the effect is greater when therapists adhere closely to the autonomy-supporting stance than when they drift into directive correction (Apodaca & Longabaugh, 2009; Apodaca et al., 2016). The early MI literature is especially clear that confrontation backfires: in Miller, Benefield, and Tonigan's (1993) comparison of directive-confrontational and client-centred styles, more confrontational therapist behaviour predicted more client resistance, which in turn predicted poorer drinking outcomes at twelve-month follow-up.

The most commonly invoked psychological mechanism for these findings is *reactance*. Reactance theory (Brehm, 1966; Brehm & Brehm, 1981) holds that when a person perceives their behavioural freedom to be threatened, they experience an aversive motivational state directed at restoring the threatened freedom — often by doing the opposite of what was advocated. A well-meaning directive to change can therefore produce a strengthening of the very behaviour the directive targeted, a phenomenon that has been replicated across health-behaviour, persuasion, and clinical contexts (Steindl et al., 2015). A related mechanism is the ironic-process effect, in which deliberate attempts to suppress an unwanted thought or

behaviour increase its subsequent occurrence (Wegner, 1994). Both mechanisms point in the same direction: information delivered as a demand on the patient's behaviour is liable to entrench the existing prior rather than update it.

In predictive-processing terms, the condition follows from the cost structure of belief update. Questioning a high-precision prior generates short-term prediction error and, with it, a transient increase in free energy — phenomenologically experienced as stress, anxiety, or destabilisation (Hauke & Lohr, 2022). Under safe and trust-supported conditions, the patient can tolerate this short-term cost long enough for the deeper update to occur. Confrontational delivery, however, raises the cost beyond what the patient can absorb in the moment: the perturbation crosses out of the prediction error zone identified in §5.2.1, the patient retreats into defence of the existing model, and the new information fails to land. Non-confrontational delivery, by contrast, distributes the cost more gradually and leaves the patient's sense of being a free agent — which is itself a deeply held prior, with high precision under healthy conditions — intact. In this sense the fourth condition is a protective constraint on the first three: it preserves the trust condition (the therapist does not act as someone trying to override the patient), it preserves the preference condition (the patient's agency is treated as one of the preferred states to be supported, not as an obstacle to be removed), and it keeps the perturbation within the magnitude band at which destabilisation triggers updating rather than rejection.

Examples in MST

The Milan team's whole interventional repertoire can be read as an extended commitment to non-confrontational delivery. The two case studies illustrate the pattern. In the Crying Boy case, the team's final intervention concludes not with a directive to John, to his parents, or to his siblings, but with the observation that “therapy is not necessary for the time being” because the children are “doing all these necessary things” for the parents' myth; the only request was that the siblings meet together periodically to discuss “how to continue this work you are doing” (Boscolo et al., 1987, pp. 87–88). The family was given nothing to react against. In the Antonella case (Boscolo & Bertrando, 1993), the third-session intervention similarly avoided any direct attack on the temperature-monitoring behaviour, instead

prescribing a weekly day for Antonella and her mother to “try to understand each other” (p. 254). In both cases the therapeutic input was a reframe or a ritual rather than an instruction, and in both cases the family retained the sense that they remained free agents in the change process.

5.3 How the Problem Dissolves in a Multi-Agent System

§§5.2.1–5.2.4 examined how systemic interventions may introduce new information that satisfies the four PTPN conditions and so updates a single agent’s generative model. But that examination was focused on the individual. MST treats family problems as residing in interactions rather than in individuals, and so a satisfactory account of the mechanism must extend the analysis from one mind to several, and from one session to the stretch of family life that follows it.

The first key observation is that all family members are present in the same session, and so all of them have the opportunity to take in new information at the same time. A problem can persist in a family for years in part because if a single member updates their priors and chooses a different course of action, the rest of the system remains entrenched in the homeostatic loop: the deviation is too small to move the system as a whole, and the system pulls the deviating member back into balance. This pattern can repeat indefinitely as long as priors are never updated simultaneously across all, or at least most, of the members. A systemic therapy session, by contrast, creates a single event in which several members’ generative models update at roughly the same time. Even when those updates are not identical in content, multiple members deviating from their usual behaviours at once means that the overall system is pushed away from its previous equilibrium and given an opportunity to settle into a new one.

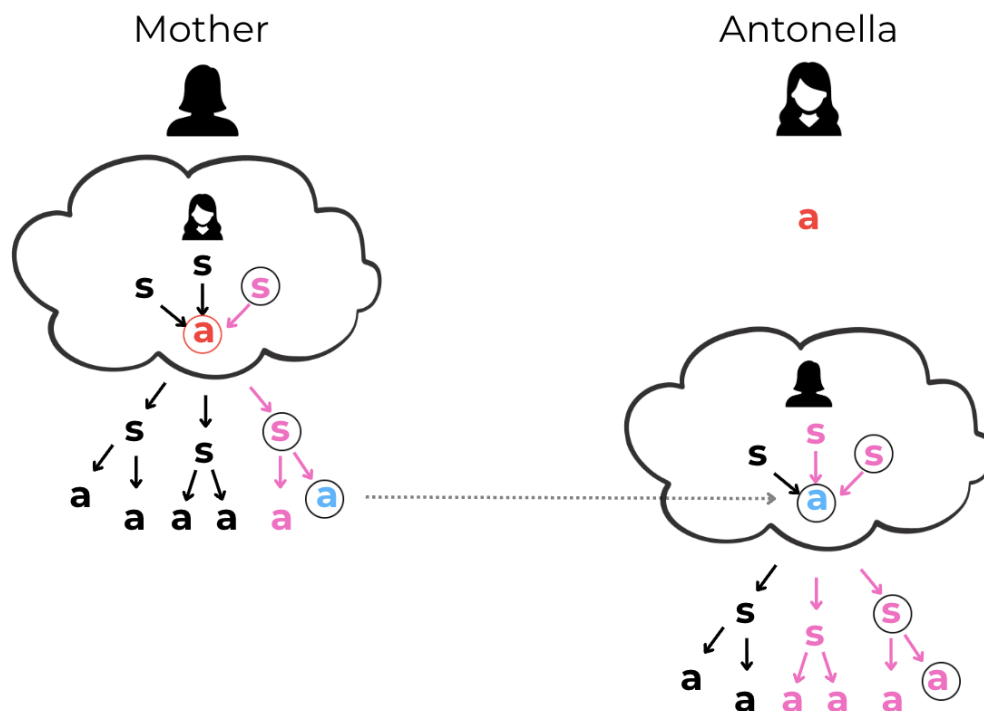
Belief update in this multi-agent setting is also fundamentally iterative. The change of priors does not begin and end in the session; the session introduces new information, and the family then carries that information forward through every subsequent interaction at home. Each interaction is a small Bayesian update against the still-developing posteriors of the previous one. Antonella does something; her mother now has new candidate interpretations available

for what that action means, and may settle on one of them as posterior; her response is, in turn, slightly different from what it would have been before; Antonella now has new candidate interpretations of her mother's response, and so on. The therapist is no longer in the room for any of this, but the conditions the therapist established — new information now resident in the shared model, sufficient trust still in place, the new framings still favourable and non-confrontationally delivered — continue to gate which candidate interpretations are selected as posteriors.

Figure 14 illustrates a single iteration of this loop for the Antonella case.

Figure 14

How the problem changes from a multi-agent PP perspective



An action a is performed by Antonella, shown on the right side in red. Her mother observes it, and her generative model produces several candidate hidden states s — intentions, feelings, motives — that could lie behind Antonella's action. Some of these candidates are familiar from her existing model; others, shown in pink, are new candidate states that the therapy

session has introduced. Because the four PTPN conditions are satisfied, one of these new candidate states is selected as the mother's posterior — the path shown by the pink line through the circles. Her resulting action, shown as the blue *a*, is therefore different from what it would have been before. Antonella now observes this changed action, and her own model produces candidate hidden states to account for it. Because her mother's action is itself already outside the usual range, the candidate states Antonella considers are themselves shifted away from her usual set, and the new — pink — interpretations are more strongly supported than they would have been in the absence of the changed input.

The diagram captures only one step of this exchange; the substance of the multi-agent claim is that this step repeats, across many turns and across the time between sessions. Each agent's update acts as a flywheel for the others'. The mother's having selected a new interpretation supports her selection of a new action, which supports Antonella's selection of a new interpretation of her mother's behavior in turn, which supports Antonella's selection of a new action, and so on. Small individual updates compound into a substantial drift away from the system's previous trajectory; what begins as a marginal change in how a single utterance is parsed can, across many iterations, reorganise the family's habitual interactional pattern.

Second, the new information has entered the family's shared generative model (see §2.7.1). It is now common ground that the symptom can be read as research, or as serving the family, or as belonging to several members rather than one. This shared common ground engages each member's theory-of-mind machinery: Antonella now models how her family may be interpreting her behaviour, and, knowing that the family has the new interpretations available, she may begin to act in ways that solicit those interpretations rather than the old ones. The therapist's reframe thus continues to do work in the family long after the session has ended, both by changing what each member predicts of the others and by changing what each member offers for the others to predict.

6. Discussion

6.1 Implications

6.1.1 Implications and Next Steps for Psychotherapy as a Whole

It is worth situating this thesis within the broader theoretical shift that PP represents. PP is not merely another candidate model of how the brain works; it inverts the standard picture. The classical view, on which most schools of psychotherapy have implicitly relied, treats the brain as a passive stimulus-driven organ; on the PP account, the same brain becomes an active probabilistic prediction machine (Villiger, 2024; Clark, 2016; Friston, 2010).

If this inversion is taken seriously, it has substantial consequences for what psychotherapy can be. On the classical, reactive picture, the natural target of intervention is the response: the behaviour or the symptom that follows from input. Therapy on this picture is the project of reshaping that response, whether by adjusting the contingencies that maintain it (the behavioural tradition), by changing the conscious thoughts that mediate it (the cognitive tradition), or by working through the historical experiences that shaped it (the psychodynamic tradition). All three traditions, in their classical forms, treat the symptom as an output that can be addressed at the level of the output. In the PP picture, the symptom is the expression of a particular generative model that the brain (or, in the systemic case, the family) is currently using to interpret its world. Treating the output without revising the model that generated it is at best a temporary intervention, because the same model will continue to generate the same expectations, the same sensory predictions, and — under the right conditions — the same symptoms. On this reading, the real work of therapy is not to react against the symptom but to help the patient (or the family) construct or revise the generative model from which the symptom follows. Agency, in this picture, is returned to the patient: rather than being trained to react differently to events that happen to them, they are supported in approaching those events from within a revised model.

This recasting of therapy as model revision also illuminates a long-standing puzzle in the clinical literature: why insight alone is so often insufficient for change. If the brain were a

reactive processor, then giving the patient a correct interpretation of their situation should be enough — the new information should change the response. In practice it rarely is, and PP offers a principled explanation for why. A single verbal interpretation, however accurate, is a low-precision signal compared with a generative model that has been built up over years of accumulated experience and that is being actively used to organise current perception. The prior wins; the interpretation slides off. For belief revision to occur, the new information has to be delivered with sufficient precision — through repeated experience, through emotionally arousing material, through structural reorganisation of the model itself — to overcome the precision of the prior it is trying to update. This is the same point that the memory-reconsolidation literature has been making for two decades (Ecker, 2015; Lane et al., 2015), and that recent PP treatments of psychotherapy have begun to formalise (Chamberlin, 2019, 2023; Hauke & Lohr, 2022). The therapeutic alliance, on a closely related reading, functions partly as a way of raising the precision the patient assigns to the therapist's interventions in the first place — without trust, the therapist's perturbations of the model carry too little weight to do anything (Krupnik, 2023).

The role of the therapist in the PP picture is correspondingly different. The therapist is not, in the first instance, the bearer of correct interpretations or the dispenser of helpful advice. The therapist is the constructor of conditions under which the patient's or the family's generative model is most likely to update — conditions in which a high-precision prediction error can be generated against the model that is producing the trouble. Milan systemic interventions, on the reading developed in this thesis, can be understood as a particularly explicit version of this constructional work. Circular questioning, positive connotation, and the family's encounter with the team's hypothesis are not direct claims about what the family is or should be; they are perturbations introduced into a system whose collective generative model would otherwise continue to organise its perception of itself in the same way. The therapeutic effect, when it occurs, is not the result of the family having been persuaded of a correct view but of the family's shared model having been pushed into a region of state-space from which it can no longer reach equilibrium on its old terms. What changes is the model, and with it the perceptions, the predictions, the affective responses, and the actions that the model was generating.

6.1.2 Implications and Next Steps for Predictive Processing

For PP, this thesis represents one more domain in which the framework has been put to work. PP is not yet a fully established account of how the brain works, and applying a developing paradigm to a new phenomenon is itself one of the ways in which such a paradigm is tested. Where PP has previously been applied to core perception, to action, and to psychopathology — most prominently to schizophrenia in the work of Friston and colleagues — its extension to systemic therapy adds a class of phenomena that is further from its original empirical home. If the analysis offered here holds, then PP gains another piece of evidence that its vocabulary genuinely applies to social and relational phenomena, not only to individual perception and inference. If it does not hold, that itself would be a useful negative result, exposing the limits of where the framework can do genuine explanatory work and where it can only re-describe.

Several of the specific claims made in this thesis are, in principle, testable. The claim that circular questioning generates prediction error against a shared family-level generative model could be examined experimentally — for instance, by measuring physiological or neural correlates of prediction-error responses during real or simulated circular-questioning sessions. The claim that positive connotation operates through precision modulation on threat-related priors could be tested against alternative accounts of how non-confrontational framings reduce defensive responding. Whether the necessary methodology yet exists, particularly for measuring shared or relational generative models rather than individual ones, is a separate question — but the claims themselves are not in principle metaphysical.

6.1.3 Implications and Next Steps for Milan Systemic Therapy

This thesis is not undertaken in order to change anything about how MST is conducted. The aim is to examine whether PP can provide an adequate explanation for why MST techniques work, rather than to propose a revised practice.

If this explanation is accepted, it could help put MST on stronger scientific footing within the wider therapeutic field. Systemic practitioners themselves have observed that the methods seem to work well but that the reasons why they work are not well understood. The field's existing theoretical apparatus is drawn largely from biological and cybernetic ideas that are

now several decades old — the kinetic accounts of the early systems thinkers, Maturana and Varela’s autopoiesis, Bateson’s ecology of mind — and these traditions, although foundational, predate the contemporary cognitive-neuroscience frontier. A framework that connects MST techniques to a current account of brain function would bring the practice closer to being recognised as a scientifically grounded approach, on a footing comparable to that of the other therapies for which predictive-processing accounts have already been proposed (Chamberlin, 2019, for EMDR; Fujimoto, 2026, for body-oriented trauma intervention).

Just as the ideas of Bateson and Maturana heavily influenced the evolution of MST in the 1970s and 1980s, PP could continue to inform its evolution into the twenty-first century. If the analysis offered here is on the right track, future refinements of systemic technique could be developed with an eye to what they are doing at the level of prediction, precision, and model update, and existing techniques could be assessed against the same machinery. None of this requires any change to current practice; the point is that the practice would then sit on a fuller theoretical foundation than it currently does.

6.2. Limitations

6.2.1 Predictive Processing’s Poor Falsifiability

One of the recurrent criticisms of the predictive-processing paradigm is that it is too flexible. The precision-weighting machinery in particular allows seemingly contradictory empirical findings to be accommodated within a single explanatory schema. The clearest case is schizophrenia, where the same patient population is invoked to support opposite precision claims: hallucinations are explained by over-precise high-level priors that override sensory data and generate percepts without external input (Adams et al., 2013; Sterzer et al., 2018), while the failure of the same patients to perceive the hollow-mask illusion is explained by under-precise high-level priors that no longer dominate depth cues (Sterzer et al., 2018). Adams et al. (2013) propose a two-stage reconciliation — a primary deficit of failure to attenuate sensory precision, with a compensatory increase in prior precision — but the same explanatory pattern recurs across other psychopathologies. Autism has been variously

characterised as involving over-precise sensory weighting (Pellicano & Burr, 2012; Friston & Frith, 2015), under-precise social priors (Palmer et al., 2015), or aberrant meta-precision (Van de Cruys et al., 2014); depression has been described in terms of over-precise negative priors (Kube et al., 2020), blunted reward-related precision, or aberrant precision-learning rates, depending on the symptom being targeted.

The general worry, sometimes called the *magic-modulator* problem (Miller & Clark, 2018, who name the worry and propose a response; see also Sprevak, 2023), is that precision parameters are typically not directly measured and instead are inferred post hoc from the very behaviour or symptomatology they are supposed to explain. Almost any pattern of perception, action, or symptomatic persistence can in principle be accommodated by assigning the right combination of precision values across the hierarchy. Litwin and Miłkowski (2020) extend this concern to the entire framework, arguing that PP-based models “are seldom empirically validated, frequently offered as mere just-so stories.” Williams (2022) makes a related but distinct point: the free-energy principle on which much of PP rests admits two interpretations — a Descriptive version that is trivially compatible with any self-organising system, and an Explanatory version that would have substantive implications for brain function but is not actually entailed by the descriptive one — with the apparent strength of the framework resulting from an unnoticed equivocation between the two. Hodson et al. (2024) come to a similar empirical conclusion from a different direction: existing studies offer only modest direct support for predictive coding, and the active inference framework has not yet been formally compared to alternative model classes such as model-free reinforcement learning, leaving its empirical priority unestablished.

Some defenders of PP have argued that this critique is partly misplaced. Miller and Clark (2018) propose that the magic-modulator worry is averted once precision is grounded in specific subcortical and neuromodulatory systems (dopamine, noradrenaline, acetylcholine), which constrain how precision can be assigned. Andrews (2021) argues that demanding falsifiability of the free-energy principle rests on a category error, treating a model structure as if it were a substantive empirical theory. These are honest replies but they shift the terms of the debate rather than meeting the falsifiability demand head-on.

The same concern applies to the application of PP to systemic therapy that the present thesis develops. I have not been able to identify a falsifiable hypothesis in systemic therapy itself. It is therefore possible that the apparent explanatory success of PP applied to systemic therapy reflects the paradigm's flexibility rather than evidence of its truth. The argument of the present thesis should be read accordingly — as offering a candidate theoretical *interpretation* of what systemic interventions are doing, not as a confirmed empirical model of why they work.

6.2.2 Mismatch in Time Scales

Predictive processing is often formulated as an account of belief update operating on a millisecond-to-second timescale, while systemic family therapy works at the scale of weeks, months, and sometimes years. The Antonella case, for example, unfolded across six sessions spread over more than a year, with the most marked behavioural changes — return to university, disappearance of the temperature phobia — appearing during the long gaps between sessions rather than within them (Boscolo & Bertrando, 1993, pp. 235–255). On the face of it, this timescale mismatch presents a difficulty for any predictive-processing account of how the therapy works.

§5.3 has begun to address this difficulty. The argument developed there is that a systemic intervention does not produce its effect in a single session but seeds a process of iterative Bayesian update that continues into the interactions the family has at home, across many short cycles, with each family member's small per-cycle updates acting as a flywheel for the others'. On this picture, the slow-clock change that the family eventually reports is the cumulative product of many fast-clock updates, none of which alone would be visible behaviourally. Because the family's priors over its own interactions are typically high-precision and well-entrenched, a single perturbation cannot move them; what moves them is the sustained inflow of mildly altered observations from family members whose own models have begun to drift. The systemic character of the problem — the fact that each member's behaviour is partly a response to the others' — explains why the propagation must be sustained over time before the family as a whole settles into a new pattern.

The hierarchical structure of the active inference framework is consistent with this account. Generative models in the framework can be extended hierarchically to represent processes that unfold at very different timescales: lower levels handle fast-changing observations (words within a sentence; the kinematics of an action) while higher levels represent slower-changing structure (the meaning of the sentence; the goal behind the action) (Friston et al., 2021; Proietti et al., 2023). A family's generative model of its own interactional patterns sits at a higher, slower-changing level than the within-conversation dynamics in which prediction error is generated. The formal machinery for relating fast within-session updates to slow between-session changes is therefore already in place; what is missing is a worked-out model of how the two timescales couple in the specific case of family therapy.

This thesis has not attempted such a model. The account in §5.3 is qualitative, and it does not specify how many iterations would be required before observable change occurs in cases of varying severity, nor how the propagation depends on the structure of the family system or on the size of the original perturbation. Computational modelling work along these lines — perhaps building on multi-agent PP simulations of the kind developed for epistemic communities (Albarracín et al., 2022) — would be a natural next step.

6.2.3 Reinforcement Learning as an Alternative Framework

An alternative explanation to the predictive-processing account presented here is reinforcement learning (RL). On a reinforcement-learning view, behaviour is shaped not by the testing and updating of prior expectations about sensory states but by the post hoc consequences of action: actions that yield reward are strengthened, those that do not are weakened, and over many repetitions the agent converges on policies that maximise expected reward. The framework has formidable computational and empirical credentials. Sutton and Barto's (1998) formal apparatus has been widely successful in machine learning, and the discovery that midbrain dopamine neurons code a reward prediction error — being activated by more reward than expected, returning to baseline for fully predicted rewards, and suppressed by less reward than expected — has been one of the most replicated findings in computational neuroscience (Schultz, 2016). On this view, the relevant signal driving learning

is the discrepancy between actual and expected reward, not the more general sensory prediction error of the PP framework.

The relationship between PP and RL has been examined directly, and the two frameworks are not strict empirical rivals. Friston et al. (2009) argued that the optimisation of behaviour does not require the notion of reward, value, or utility, demonstrating with a worked solution to the classic mountain-car problem that active-inference agents (in the sense of the framework) can reproduce policies normally derived from RL and dynamic programming. Sajid et al. (2021) extended this with a direct comparison against Q-learning and Bayesian model-based RL agents on OpenAI gym tasks, showing that active inference agents can learn behaviours in *reward-free* environments by treating reward as one observation among others that the agent has a preference over, and that they perform epistemic exploration (uncertainty-reducing action) “in a Bayes-optimal fashion” rather than relying on heuristics that RL has to engineer in separately. The most formally precise result is Da Costa et al. (2023), who proved that *sophisticated* active inference produces Bellman-optimal — i.e., reward-maximising — actions on any finite temporal horizon, while standard active inference does so only for a planning horizon of one step ahead. In this sense active inference encompasses RL as a special case, though only the recursive sophisticated form does so generally. The dopaminergic signal that Schultz interpreted as a reward prediction error has likewise been reinterpreted within PP as encoding the *precision* of prediction errors rather than scalar reward value, with implications for how aberrant dopamine signalling produces psychotic symptoms (Diederen & Fletcher, 2021; Adams et al., 2013) — though Diederen and Fletcher (2021) themselves note that direct empirical evidence for the PP reinterpretation of dopamine remains limited.

The empirical priority of either framework therefore remains unsettled. Hodson et al. (2024) review the empirical status of predictive coding and active inference and conclude that, although active-inference models often fit behavioural data reasonably well, “there has not been a focus on testing empirical validity of active inference theory per se, which would require formal comparison to other models (e.g., non-Bayesian or model-free RL models)” (p. 1). They identify direct empirical comparison between the active inference framework and model-free RL as one of the most important outstanding tasks for the field.

This debate has concrete relevance for the present thesis. Many of the therapeutic effects of Milan systemic interventions resist a clean reinforcement-learning re-description. Consider positive connotation: a straightforward reinforcement-learning re-reading runs into difficulty here. The family's prior negativity toward the symptom would, in operant terms, be a punisher of the behaviour; removing that punisher (which is what positive connotation does to the family's reactions) should, on a model-free RL account, increase rather than reduce the symptom. The observed clinical outcome is in the opposite direction.

A second feature of Milan-targeted symptoms that strains the RL reading is their persistence in the face of sustained negativity, frustration, and pressure to change from other family members — circumstances in which simple operant accounts predict weakening rather than persistence. The behaviourist response — that hidden secondary reinforcers must be maintaining the symptom — is perhaps defensible, but not in the scope of this thesis to examine. More tellingly, the most direct computational-RL attempts to model the persistence of negative thinking — including work by authors who argue strongly for the RL-of-psychotherapy programme — have themselves shown that rumination disrupts reinforcement learning rather than being well-explained by it (Hitchcock et al., 2022).

A recent computational-psychotherapy literature has begun to formalise psychotherapy mechanisms in computational terms, drawing on both RL and Bayesian/representation-learning frameworks (Berwian et al., 2025); notably, that literature itself proposes that exposure therapy is better explained by Bayesian *latent-cause inference* — the patient infers a new safe context in which the feared stimulus is non-threatening, rather than overwriting the original fear association — which is closer in structure to the PP reading than to classical model-free RL. Distinguishing the RL and PP interpretations of therapeutic effect empirically within a Milan setting would require measurements that this thesis does not undertake and that, on Hodson et al.'s Fri (2024) assessment, have not yet been undertaken anywhere in the literature. This could be the subject of future work. The argument here is framed as offering a predictive-processing *interpretation* of Milan techniques, not as a refutation of reinforcement-learning accounts of behaviour change.

6.2.4 Interventions Outside the Scope of This Analysis

Circular questioning, future-oriented questions, and positive connotation with its accompanying ritual were examined in this thesis as what I believe are the core mechanisms of change. This is supported by the Milan therapist's own beliefs and their observation that families often expressed great surprise at the positive connotation and prescription (Boscolo et al, 1987).

However, it could be that it is a different aspect of systemic therapy, such as the observing teams, the 5-part structure of the session, or common factors such as the therapist's empathy or shared goals that produce the change.

It is still possible that these other features of therapy operate through PP, therefore this would not mean that PP cannot explain why systemic therapy produces outcomes, however, it would mean that my analysis provided above is not the way it occurs, and possibly incorrect.

Conclusion

This thesis has offered a predictive-processing reading of Milan systemic therapy, proposing that its central interventions — circular questioning, positive connotation, and the closing team hypothesis — function as structured perturbations of the family's shared generative model. It is intended to serve as a theoretical starting point for future empirical and computational work, offering both a mechanism-level account of how Milan techniques may generate prediction error and the PTPN framework (Perturbation, Trust, Preference, Non-confrontation), which specifies the conditions under which such information actually updates priors. Several theoretical commitments needed to be made in constructing this account, highlighting open questions around predictive processing's empirical priority over alternative frameworks and the mismatch between the short timescale of predictive inference and the much longer course over which systemic therapy unfolds. Further research, both formal modelling of multi-agent inference under therapeutic perturbation and empirical work tracking belief revision across therapy sessions, is therefore necessary for testing the framework proposed here.

References

- Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, 4, 47. <https://doi.org/10.3389/fpsyt.2013.00047>
- Albarracin, M., Demekas, D., Ramstead, M. J. D., & Heins, C. (2022). Epistemic communities under active inference. *Entropy*, 24(4), 476. <https://doi.org/10.3390/e24040476>
- Allison, R. I., & Uhl, K. P. (1964). Influence of beer brand identification on taste perception. *Journal of Marketing Research*, 1(3), 36–39. <https://doi.org/10.2307/3150054>
- Anderson, H. (2016). Postmodern/poststructural/social construction therapies: Collaborative, narrative, and solution-focused. In T. L. Sexton & J. Lebow (Eds.), *Handbook of family therapy* (pp. 182–204). Routledge/Taylor & Francis Group.
- Anderson, H., Goolishian, H. A., & Winderman, L. (1986). Problem determined systems: Towards transformation in family therapy. *Journal of Strategic and Systemic Therapies*, 5(4), 1–13.
- Andrews, M. (2021). The math is not the territory: Navigating the free energy principle. *Biology & Philosophy*, 36(3), 30. <https://doi.org/10.1007/s10539-021-09807-0>
- Apodaca, T. R., Jackson, K. M., Borsari, B., Magill, M., Longabaugh, R., Mastroleo, N. R., & Barnett, N. P. (2016). Which individual therapist behaviors elicit client change talk and sustain talk in motivational interviewing? *Journal of Substance Abuse Treatment*, 61, 60–65. <https://doi.org/10.1016/j.jsat.2015.09.001>
- Apodaca, T. R., & Longabaugh, R. (2009). Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction*, 104(5), 705–715. <https://doi.org/10.1111/j.1360-0443.2009.02527.x>

- Aronson, E., & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of Experimental Social Psychology*, 1(2), 156–171.
[https://doi.org/10.1016/0022-1031\(65\)90043-0](https://doi.org/10.1016/0022-1031(65)90043-0)
- Ashby, W. R. (1956). *An introduction to cybernetics*. Chapman & Hall.
- Badcock, P. B., & Davey, C. G. (2024). Active inference in psychology and psychiatry: Progress to date? *Entropy*, 26(10), 833. <https://doi.org/10.3390/e26100833>
- Barbetta, P., & Telfener, U. (2020). The Milan approach, history, and evolution. *Family Process*, 60(1), 4–16. <https://doi.org/10.1111/famp.12612>
- Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160011. <https://doi.org/10.1098/rstb.2016.0011>
- Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chandler Publishing.
- Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. (1956). Toward a theory of schizophrenia. *Behavioral Science*, 1, 251–264. <https://doi.org/10.1002/bs.3830010402>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529.
<https://doi.org/10.1037/0033-2909.117.3.497>
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). Academic Press.
[https://doi.org/10.1016/S0065-2601\(08\)60024-6](https://doi.org/10.1016/S0065-2601(08)60024-6)
- Bertalanffy, L. von. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, 1(2), 134–165. <https://doi.org/10.1093/bjps/I.2.134>
- Bertalanffy, L. von. (1968). *General system theory: Foundations, development, applications*. George Braziller.

Bertrando, P. (2002). The presence of the third party: Systemic therapy and transference analysis. *Journal of Family Therapy*, 24(3), 351–368.

<https://doi.org/10.1111/1467-6427.00218>

Bertrando, P. (2007). *The dialogical therapist: Dialogue in systemic practice*. Karnac Books.

Bertrando, P., & Arcelloni, T. (2006). Hypotheses are dialogues: Sharing hypotheses with clients. *Journal of Family Therapy*, 28(4), 370–387.

<https://doi.org/10.1111/j.1467-6427.2006.00358.x>

Bertrando, P., & Lini, C. (2019). Theory and practice of systemic-dialogical therapy. *Australian and New Zealand Journal of Family Therapy*, 40(3), 176–189.

<https://doi.org/10.1002/anzf.1365>

Bertrando, P., & Lini, C. (2023). *Systemic-dialogical therapy with individuals: Finding one's place*. Routledge.

Berwian, I. M., Hitchcock, P., Pisupati, S., Schoen, G., & Niv, Y. (2025). Using computational models of learning to advance cognitive behavioural therapy. *Communications Psychology*, 3, 72. <https://doi.org/10.1038/s44271-025-00251-4>

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76, 198–211.

<https://doi.org/10.1016/j.jmp.2015.11.003>

Bolis, D., & Schilbach, L. (2018). ‘I interact therefore I am’: The self as a historical product of dialectical attunement. *Topoi*, 39, 521–534. <https://doi.org/10.1007/s11245-018-9574-0>

Boscolo, L., & Bertrando, P. (1993). *The times of time: A new perspective in systemic therapy and consultation* (S. Thorne, Trans.). W. W. Norton.

Boscolo, L., & Bertrando, P. (1996). *Systemic therapy with individuals*. Karnac Books.

Boscolo, L., Cecchin, G., Hoffman, L., & Penn, P. (1987). *Milan systemic family therapy: Conversations in theory and practice*. Basic Books.

- Bray, J. H., & Stanton, M. (Eds.). (2009). *The Wiley-Blackwell handbook of family psychology*. Wiley-Blackwell.
- Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. Academic Press.
- Breunlin, D. C. (2016). Advancing training and supervision of family therapy. In T. L. Sexton & J. Lebow (Eds.), *Handbook of family therapy* (2nd ed., pp. 517–529). Routledge.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, 55(3), 429–444. <https://doi.org/10.1111/0022-4537.00126>
- Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*, 2, 581–591. <https://doi.org/10.1038/s41562-018-0376-6>
- Campbell, C., Kumpasoğlu, G. B., & Fonagy, P. (2024). Mentalizing, epistemic trust, and the active ingredients of psychotherapy. *Psychodynamic Psychiatry*, 52(4), 435–451. <https://doi.org/10.1521/pdps.2024.52.4.435>
- Campbell, D. (2003). The mutiny and the bounty: The place of Milan ideas today. *Australian and New Zealand Journal of Family Therapy*, 24(1), 15–25.
- Campbell, D., & Draper, R. (Eds.). (1985). *Applications of systemic family therapy: The Milan approach*. Grune & Stratton.
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, 9(3), 399–431. <https://doi.org/10.1152/physrev.1929.9.3.399>
- Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain*, 133(4), 1265–1283. <https://doi.org/10.1093/brain/awq010>

- Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics. *Pharmacological Reviews*, 71(3), 316–344. <https://doi.org/10.1124/pr.118.017160>
- Carr, A. (2016). The evolution of systems theory. In T. L. Sexton & J. Lebow (Eds.), *Handbook of family therapy* (pp. 13–29). Routledge/Taylor & Francis Group.
- Cecchin, G. (1987). Hypothesizing, circularity, and neutrality revisited: An invitation to curiosity. *Family Process*, 26(4), 405–413.
- Cecchin, G., Lane, G., & Ray, W. A. (1992). *Irreverence: A strategy for therapists' survival*. Karnac Books.
- Chamberlin, D. E. (2019). The predictive processing model of EMDR. *Frontiers in Psychology*, 10, 2267. <https://doi.org/10.3389/fpsyg.2019.02267>
- Chamberlin, D. E. (2023). The active inference model of coherence therapy. *Frontiers in Human Neuroscience*, 16, 955558. <https://doi.org/10.3389/fnhum.2022.955558>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190217013.001.0001>
- Clark, A. (2023). *The experience machine: How our minds predict and shape reality*. Allen Lane.
- Clausius, R. (1867). On several convenient forms of the fundamental equations of the mechanical theory of heat. In T. A. Hirst (Ed.), *The mechanical theory of heat, with its applications to the steam-engine and to the physical properties of bodies* (pp. 327–365). John van Voorst. (Original work published 1865)

- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371.
<https://doi.org/10.1146/annurev-psych-010213-115137>
- Collins, D., & Tomm, K. (2009). Karl Tomm: His changing views on family therapy over 35 years. *The Family Journal*, 17(2), 106–117.
- Constant, A., Ramstead, M. J. D., Veissière, S. P. L., & Friston, K. (2019). Regimes of expectations: An active inference model of social conformity and human decision making. *Frontiers in Psychology*, 10, 679. <https://doi.org/10.3389/fpsyg.2019.00679>
- Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, 35(3), 32. <https://doi.org/10.1007/s10539-020-09746-2>
- Da Costa, L., Sajid, N., Parr, T., Friston, K., & Smith, R. (2023). Reward maximisation through discrete active inference. *Neural Computation*, 35(5), 807–852.
https://doi.org/10.1162/neco_a_01574
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20–33.
<https://doi.org/10.1037/0022-3514.44.1.20>
- Dell, P. F. (1982). Family theory and the epistemology of Humberto Maturana. *Family Therapy Networker*, 6, 26, 39–41.
- Dell, P. F. (1985). Understanding Bateson and Maturana: Toward a biological foundation for the social sciences. *Journal of Marital and Family Therapy*, 11(1), 1–20.
<https://doi.org/10.1111/j.1752-0606.1985.tb00587.x>
- Dell, P. F. (1989). Violence and the systemic view: The problem of power. *Family Process*, 28(1), 1–14.
- Diaconescu, A. O., Mathys, C., Weber, L. A. E., Daunizeau, J., Kasper, L., Lomakina, E. I., Fehr, E., & Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical

Bayesian learning. *PLoS Computational Biology*, 10(9), e1003810.

<https://doi.org/10.1371/journal.pcbi.1003810>

Diederen, K. M. J., & Fletcher, P. C. (2021). Dopamine, prediction error and beyond. *The Neuroscientist*, 27(1), 30–46. <https://doi.org/10.1177/1073858420907591>

Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., & Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *NeuroImage*, 46(4), 1180–1186.

<https://doi.org/10.1016/j.neuroimage.2009.03.033>

Doherty, W. J. (2020). The evolution and current status of systemic family therapy: A sociocultural perspective. In K. S. Wampler, R. B. Miller, & R. B. Seedall (Eds.), *The handbook of systemic family therapy: Vol. 1. The profession of systemic family therapy*. John Wiley & Sons. <https://doi.org/10.1002/9781119790181.ch2>

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

<https://doi.org/10.7551/mitpress/9780262042383.001.0001>

Ecker, B. (2015). Memory reconsolidation understood and misunderstood. *International Journal of Neuropsychotherapy*, 3(1), 2–46. <https://doi.org/10.12744/ijnpt.2015.0002-0046>

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 215. <https://doi.org/10.3389/fnhum.2010.00215>

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of prediction error. *Neuroscience & Biobehavioral Reviews*, 68, 423–441. <https://doi.org/10.1016/j.neubiorev.2016.06.004>

Fernández Velasco, P., & Loev, S. (2025). Metacognitive feelings: A predictive-processing perspective. *Perspectives on Psychological Science*, 20(4), 525–544. <https://doi.org/10.1177/17456916231221976>

Fisch, R. (1978). Review of *Problem-solving therapy*, by Jay Haley. *Family Process*, 17, 107–110.

Fisch, R., Weakland, J. H., & Segal, L. (1982). *The tactics of change: Doing therapy briefly*. Jossey-Bass.

Fisher, E. L., & Hohwy, J. (2024). The universal optimism of the self-evidencing mind. *Entropy*, 26(6), 518. <https://doi.org/10.3390/e26060518>

Fisher, J. P., Hassan, D. T., & O'Connor, N. (1995). Minerva. *British Medical Journal*, 310(6971), 70. <https://doi.org/10.1136/bmj.310.6971.70>

Fotopoulou, A., & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychanalysis*, 19(1), 3–28.
<https://doi.org/10.1080/15294145.2017.1294031>

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

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4(11), e1000211. <https://doi.org/10.1371/journal.pcbi.1000211>

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.

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

Friston, K. (2018). Am I self-conscious? (Or does self-organization entail self-consciousness?). *Frontiers in Psychology*, 9, 579.
<https://doi.org/10.3389/fpsyg.2018.00579>

Friston, K. J., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2021). Sophisticated inference. *Neural Computation*, 33(3), 713–763. https://doi.org/10.1162/neco_a_01351

- Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE*, 4(7), e6421. <https://doi.org/10.1371/journal.pone.0006421>
- Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behaviour: A free-energy formulation. *Biological Cybernetics*, 102(3), 227–260. <https://doi.org/10.1007/s00422-010-0364-z>
- Friston, K., & Frith, C. (2015). A duet for one. *Consciousness and Cognition*, 36, 390–405. <https://doi.org/10.1016/j.concog.2014.12.003>
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, 364(1521), 1211–1221. <https://doi.org/10.1098/rstb.2008.0300>
- Friston, K. J., Parr, T., Heins, C., Constant, A., Friedman, D., Isomura, T., Fields, C., Verbelen, T., Ramstead, M., Clippinger, J., & Frith, C. D. (2024). Federated inference and belief sharing. *Neuroscience and Biobehavioral Reviews*, 156, 105500. <https://doi.org/10.1016/j.neubiorev.2023.105500>
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214. <https://doi.org/10.1080/17588928.2015.1020053>
- Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 3, 130. <https://doi.org/10.3389/fpsyg.2012.00130>
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313. <https://doi.org/10.1146/annurev-psych-120710-100449>
- Fryszler, A., & Schwing, R. (2014). *Systemic therapy: A practical guide*. Vandenhoeck & Ruprecht.
- Fujimoto, M. (2026). A predictive processing framework for body-oriented trauma intervention: A hypothesis illustrated by Body Connect Therapy. *Frontiers in Psychology*, 17, 1781289. <https://doi.org/10.3389/fpsyg.2026.1781289>

- Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40(3), 266–275. <https://doi.org/10.1037/0003-066X.40.3.266>
- Goldenberg, H., & Goldenberg, I. (2009). The revolution and evolution of family therapy and family psychology. In J. H. Bray & M. Stanton (Eds.), *The Wiley-Blackwell handbook of family psychology* (pp. 19–37). Wiley-Blackwell.
- Gregory, R. L. (1970). *The intelligent eye*. Weidenfeld & Nicolson.
- Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, 148(2), 325–341. <https://doi.org/10.1037/xge0000498>
- Haley, J. (1963). *Strategies of psychotherapy*. Grune & Stratton.
- Hauke, G., & Lohr, C. (2022). Piloting the update: The use of therapeutic relationship for change — A free energy account. *Frontiers in Psychology*, 13, 842488. <https://doi.org/10.3389/fpsyg.2022.842488>
- Hecker, L. L., Mims, G. A., & Boughner, S. R. (2015). General systems theory, cybernetics, and family therapy. In J. L. Wetchler & L. L. Hecker (Eds.), *An introduction to marriage and family therapy* (2nd ed., pp. 43–64). Routledge/Taylor & Francis Group.
- Helmholtz, H. von. (1962). *Helmholtz's treatise on physiological optics* (J. P. C. Southall, Ed. & Trans., Vols. 1–3). Dover. (Original work published 1867)
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20(11), 1315–1326. <https://doi.org/10.1002/hipo.20857>
- Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. D. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33(2), 398–446. https://doi.org/10.1162/neco_a_01341

- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). Guilford Press.
- Hitchcock, P. F., Forman, E. M., Rothstein, N., Zhang, F., Kounios, J., Niv, Y., & Sims, C. R. (2022). Rumination derails reinforcement learning with possible implications for ineffective behavior. *Clinical Psychological Science*, 10(4), 714–733.
<https://doi.org/10.1177/21677026211051324>
- Hodson, R., Mehta, M., & Smith, R. (2024). The empirical status of predictive coding and active inference. *Neuroscience and Biobehavioral Reviews*, 157, 105473.
<https://doi.org/10.1016/j.neubiorev.2023.105473>
- Hoffman, L. (1981). *Foundations of family therapy: A conceptual framework for systems change*. Basic Books.
- Hoffman, L. (1985). Beyond power and control: Toward a “second order” family systems therapy. *Family Systems Medicine*, 3(4), 381–396.
- Hoffman, L. (1991). A reflexive stance for family therapy. *Journal of Strategic and Systemic Therapies*, 10(3 & 4), 4–17.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, 35(2), 209–223. <https://doi.org/10.1111/mila.12281>
- Holmes, J., & Nolte, T. (2019). ‘Surprise’ and the Bayesian brain: Implications for psychotherapy theory and practice. *Frontiers in Psychology*, 10, 592.
<https://doi.org/10.3389/fpsyg.2019.00592>
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635–650. <https://doi.org/10.1086/266350>

- Jackson, D. D. (1957). The question of family homeostasis. *Psychiatric Quarterly Supplement*, 31(Part 1), 79–90.
- Jackson, D. D. (1959). Family interaction, family homeostasis, and some implications for conjoint family psychotherapy. In J. H. Masserman (Ed.), *Individual and familial dynamics* (pp. 122–141). Grune & Stratton.
- Jackson, J. B., & Landers, A. L. (2020). Structural and strategic approaches. In K. S. Wampler (Editor-in-Chief), R. B. Miller, & R. B. Seedall (Vol. Eds.), *The handbook of systemic family therapy: Vol. 1. The profession of systemic family therapy* (pp. 339–364). Wiley.
- Kafkas, A., & Montaldi, D. (2018). How do memory systems detect and respond to novelty? *Neuroscience Letters*, 680, 60–68. <https://doi.org/10.1016/j.neulet.2018.01.053>
- Keane, B. P., Silverstein, S. M., Wang, Y., & Papathomas, T. V. (2013). Reduced depth inversion illusions in schizophrenia are state-specific and occur for multiple object types and viewing conditions. *Journal of Abnormal Psychology*, 122(2), 506–512. <https://doi.org/10.1037/a0032110>
- Keane, B. P., Silverstein, S. M., Wang, Y., & Papathomas, T. V. (2016). A new depth-inversion illusion paradigm provides evidence of weakened top-down processing in schizophrenia. *Schizophrenia Research*, 178(1–3), 110–115.
- Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, 18(4), 431–439. <https://doi.org/10.1111/j.1467-6494.1950.tb01260.x>
- Khorrami Banaraki, A., Toghi, A., & Mohammadzadeh, A. (2024). RDoC framework through the lens of predictive processing: Focusing on cognitive systems domain. *Computational Psychiatry*, 8(1), 178–201. <https://doi.org/10.5334/cpsy.119>
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(11), e1000209. <https://doi.org/10.1371/journal.pcbi.1000209>

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

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719. <https://doi.org/10.1016/j.tins.2004.10.007>

Krupnik, V. (2019). Bayesian approach to psychotherapy integration: Strategic modification of priors. *Frontiers in Psychology*, 10, 356. <https://doi.org/10.3389/fpsyg.2019.00356>

Krupnik, V. (2023). The therapeutic alliance as active inference: The role of trust and self-efficacy. *Journal of Contemporary Psychotherapy*, 53(2), 207–215. <https://doi.org/10.1007/s10879-022-09576-1>

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry*, 87(5), 388–398. <https://doi.org/10.1016/j.biopsych.2019.07.017>

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by “the dress” photograph. *Current Biology*, 25(13), R545–R546. <https://doi.org/10.1016/j.cub.2015.04.053>

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, 38, e1. <https://doi.org/10.1017/S0140525X14000041>

Lebow, J., & Sexton, T. L. (2016). The evolution of family and couple therapy. In T. L. Sexton & J. Lebow (Eds.), *Handbook of family therapy* (pp. 1–12). Routledge/Taylor & Francis Group.

Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you’ll like it: The influence of expectation, consumption, and revelation on preferences for beer. *Psychological Science*, 17(12), 1054–1058. <https://doi.org/10.1111/j.1467-9280.2006.01829.x>

- Lehmann, K., Bolis, D., Friston, K. J., Schilbach, L., Ramstead, M. J. D., & Kanske, P. (2024). An active-inference approach to second-person neuroscience. *Perspectives on Psychological Science*, 19(6), 931–951. <https://doi.org/10.1177/17456916231188000>
- Lini, C. (2022). Positional responsibility in systemic-dialogical therapy. *Journal of Family Therapy*, 44(2), 339–350. <https://doi.org/10.1111/1467-6427.12390>
- Lini, C., & Bertrando, P. (2020). Finding one's place: Emotions and positioning in systemic-dialogical therapy. *Journal of Family Therapy*. Advance online publication. <https://doi.org/10.1111/1467-6427.12267>
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, 46(5), 703–713. <https://doi.org/10.1016/j.neuron.2005.05.002>
- Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, 44(7), e12867. <https://doi.org/10.1111/cogs.12867>
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589. [https://doi.org/10.1016/S0022-5371\(74\)80011-3](https://doi.org/10.1016/S0022-5371(74)80011-3)
- Lundahl, B. W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing: Twenty-five years of empirical studies. *Research on Social Work Practice*, 20(2), 137–160. <https://doi.org/10.1177/1049731509347850>
- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, 6(4), 547–569. <https://doi.org/10.1007/s13164-015-0253-4>
- Maruyama, M. (1963). The second cybernetics: Deviation-amplifying mutual causal processes. *American Scientist*, 51(2), 164–179.

- Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. *Frontiers in Human Neuroscience*, 8, 825. <https://doi.org/10.3389/fnhum.2014.00825>
- Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel Publishing.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioural preference for culturally familiar drinks. *Neuron*, 44(2), 379–387. <https://doi.org/10.1016/j.neuron.2004.09.019>
- McGaugh, J. L. (2000). Memory: A century of consolidation. *Science*, 287(5451), 248–251. <https://doi.org/10.1126/science.287.5451.248>
- McGovern, H. T., De Foe, A., Biddell, H., Leptourgos, P., Corlett, P., Bandara, K., & Hutchinson, B. T. (2022). Learned uncertainty: The free energy principle in anxiety. *Frontiers in Psychology*, 13, 943785. <https://doi.org/10.3389/fpsyg.2022.943785>
- McGovern, H. T., & Otten, M. (2024). Priors and prejudice: Hierarchical predictive processing in intergroup perception. *Frontiers in Psychology*, 15, 1386370. <https://doi.org/10.3389/fpsyg.2024.1386370>
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227–234. <https://doi.org/10.1037/h0031564>
- Miller, M., & Clark, A. (2018). Happily entangled: Prediction, emotion, and the embodied mind. *Synthese*, 195(6), 2559–2575. <https://doi.org/10.1007/s11229-017-1399-7>
- Miller, W. R., Benefield, R. G., & Tonigan, J. S. (1993). Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology*, 61(3), 455–461. <https://doi.org/10.1037/0022-006X.61.3.455>
- Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd ed.). Guilford Press.

- Mingers, J. (1991). The cognitive theories of Maturana and Varela. *Systems Practice*, 4(4), 319–338. <https://doi.org/10.1007/BF01062008>
- Mokady, A., & Reggev, N. (2022). The role of predictions, their confirmation, and reward in maintaining the self-concept. *Frontiers in Human Neuroscience*, 16, 824085. <https://doi.org/10.3389/fnhum.2022.824085>
- Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current Opinion in Neurobiology*, 23(2), 195–201. <https://doi.org/10.1016/j.conb.2013.01.020>
- Nichols, M. P., & Davis, S. D. (2020). *Family therapy: Concepts and methods* (12th ed.). Pearson.
- Norcross, J. C. (2005). A primer on psychotherapy integration. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed., pp. 3–23). Oxford University Press. <https://doi.org/10.1093/med:psych/9780195165791.003.0001>
- Oh, D., Walker, A. C., & Freeman, J. B. (2021). Person knowledge shapes face identity perception. *Cognition*, 217, 104889. <https://doi.org/10.1016/j.cognition.2021.104889>
- Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2020). On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, 486, 110089. <https://doi.org/10.1016/j.jtbi.2019.110089>
- Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Consciousness and Cognition*, 36, 376–389. <https://doi.org/10.1016/j.concog.2015.04.001>
- Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behaviour*. MIT Press. <https://doi.org/10.7551/mitpress/12441.001.0001>
- Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory*, 11(5), 579–585. <https://doi.org/10.1101/lm.76904>

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–510.

<https://doi.org/10.1016/j.tics.2012.08.009>

Penn, P. (1982). Circular questioning. *Family Process*, 21(3), 267–280.

Penn, P. (1985). Feed-forward: Future questions, future maps. *Family Process*, 24(3), 299–310.

Petitot, P., O’Reilly, J. X., & O’Shea, J. (2018). Towards a neuro-computational account of prism adaptation. *Neuropsychologia*, 115, 188–203.

<https://doi.org/10.1016/j.neuropsychologia.2017.12.021>

Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behaviour. *Biological Psychology*, 186, 108741. <https://doi.org/10.1016/j.biopsycho.2023.108741>

Prochaska, J. O., & Norcross, J. C. (2018). *Systems of psychotherapy: A transtheoretical analysis* (9th ed.). Oxford University Press.

Proietti, R., Pezzulo, G., & Tessari, A. (2023). An active inference model of hierarchical action understanding, learning and imitation. *Physics of Life Reviews*, 46, 92–118.

<https://doi.org/10.1016/j.plrev.2023.05.012>

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, 179, 107382.

<https://doi.org/10.1016/j.nlm.2021.107382>

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>

Ray, W. A., & SimmEs, M. (2016). Embracing cybernetics: Living legacy of the Bateson research team. *Cybernetics and Human Knowing*, 23, 29–57.

- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219–235. <https://doi.org/10.1037/0096-3445.118.3.219>
- Redding, G. M., Rossetti, Y., & Wallace, B. (2005). Applications of prism adaptation: A tutorial in theory and method. *Neuroscience & Biobehavioral Reviews*, 29(3), 431–444. <https://doi.org/10.1016/j.neubiorev.2004.12.004>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
- Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: Demystified and compared. *Neural Computation*, 33(3), 674–712. https://doi.org/10.1162/neco_a_01357
- Salamino, F., & Gusmini, E. (2025). Case report: Keeping the Milan approach legacy alive? Paradox and counterparadox working therapeutically with non-suicidal self-injury. *Frontiers in Child and Adolescent Psychiatry*, 4, 1657395. <https://doi.org/10.3389/frcha.2025.1657395>
- Schneider, U., Borsutzky, M., Seifert, J., Leweke, F. M., Huber, T. J., Rollnik, J. D., & Emrich, H. M. (2002). Reduced binocular depth inversion in schizophrenic patients. *Schizophrenia Research*, 53(1–2), 101–108. [https://doi.org/10.1016/S0920-9964\(00\)00172-9](https://doi.org/10.1016/S0920-9964(00)00172-9)
- Schrödinger, E. (1944). *What is life? The physical aspect of the living cell*. Cambridge University Press.
- Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, 18(1), 23–32. <https://doi.org/10.31887/DCNS.2016.18.1/wschultz>
- Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H. B., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, 8, e41703. <https://doi.org/10.7554/eLife.41703>
- Selvini Palazzoli, M. (1978). *Self-starvation: From individual to family therapy in the treatment of anorexia nervosa*. Jason Aronson.

- Selvini Palazzoli, M., Boscolo, L., Cecchin, G., & Prata, G. (1978). *Paradox and counterparadox: A new model in the therapy of the family in schizophrenic transaction* (E. V. Burt, Trans.). Jason Aronson. (Original work published 1975)
- Selvini Palazzoli, M., Boscolo, L., Cecchin, G., & Prata, G. (1980). Hypothesizing—Circularity—Neutrality: Three guidelines for the conductor of the session. *Family Process*, 19(1), 3–12.
- Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synaesthesia. *Cognitive Neuroscience*, 5(2), 97–118. <https://doi.org/10.1080/17588928.2013.877880>
- Seth, A. (2021). *Being you: A new science of consciousness*. Faber & Faber.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007. <https://doi.org/10.1098/rstb.2016.0007>
- Sexton, T. L., & Lebow, J. (Eds.). (2016). *Handbook of family therapy*. Routledge.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences*, 20(1), 25–33. <https://doi.org/10.1016/j.tics.2015.11.002>
- Sharot, T., Rollwage, M., Sunstein, C. R., & Fleming, S. M. (2023). Why and when beliefs change. *Perspectives on Psychological Science*, 18(1), 142–151. <https://doi.org/10.1177/17456916221082967>
- Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, 42(10), 727–739. <https://doi.org/10.1016/j.tins.2019.08.007>

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592–604.

<https://doi.org/10.1037/0278-7393.4.6.592>

Smith, R., Moutoussis, M., & Bilek, E. (2021). Simulating the computational mechanisms of cognitive and behavioural psychotherapeutic interventions: Insights from active inference. *Scientific Reports*, 11, 10128. <https://doi.org/10.1038/s41598-021-89047-0>

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89–96. <https://doi.org/10.1111/j.1467-7687.2007.00569.x>

Sprevak, M., & Smith, R. (2023). An introduction to predictive processing models of perception and decision-making. *Topics in Cognitive Science*. Advance online publication. <https://doi.org/10.1111/tops.12704>

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261–302). Academic Press.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, 8, 666.

<https://doi.org/10.3389/fnhum.2014.00666>

Steindl, C., Jonas, E., Sittenthaler, S., Traut-Mattausch, E., & Greenberg, J. (2015). Understanding psychological reactance: New developments and findings. *Zeitschrift für Psychologie*, 223(4), 205–214. <https://doi.org/10.1027/2151-2604/a000222>

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84(9), 634–643.

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, 19(6), 795–797.

<https://doi.org/10.1038/nn.4296>

- Stratton, P. (2016). *The evidence base of family therapy and systemic practice*. Association for Family Therapy. <http://www.aft.org.uk/>
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.
- Swann, W. B., Jr. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 408–448). Guilford Press.
- Swann, W. B., Jr., Pelham, B. W., & Krull, D. S. (1989). Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. *Journal of Personality and Social Psychology*, 57(5), 782–791. <https://doi.org/10.1037/0022-3514.57.5.782>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.
- Telfener, U., & Barazzetti, A. (2025). The Milan approach today. *Family Process*, 64(1), e13075. <https://doi.org/10.1111/famp.13075>
- Thestrup Waade, P., Lundbak Olesen, C., Ehrenreich Laursen, J., Nehrer, S. W., Heins, C., Friston, K., & Mathys, C. (2025). As one and many: Relating individual and emergent group-level generative models in active inference. *Entropy*, 27(2), 143. <https://doi.org/10.3390/e27020143>
- Tomm, K. (1984). One perspective on the Milan systemic approach: Part I. Overview of development, theory and practice. *Journal of Marital and Family Therapy*, 10(2), 113–125.
- Tomm, K. (1985). Circular interviewing: A multifaceted clinical tool. In D. Campbell & R. Draper (Eds.), *Applications of systemic family therapy: The Milan approach* (pp. 33–45). Grune & Stratton.
- Tomm, K. (1987). Interventive interviewing: Part I. Strategizing as a fourth guideline for the therapist. *Family Process*, 26(1), 3–13.

Tomm, K. (1988). Interventive interviewing: Part III. Intending to ask lineal, circular, strategic, or reflexive questions? *Family Process*, 27(1), 1–15.

<https://doi.org/10.1111/j.1545-5300.1988.00001.x>

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, 138(1), 39–63.

<https://doi.org/10.1037/a0014678>

Tschantz, A., Barca, L., Maisto, D., Buckley, C. L., Seth, A. K., & Pezzulo, G. (2022). Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. *Biological Psychology*, 169, 108266.

<https://doi.org/10.1016/j.biopsycho.2022.108266>

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism.

Psychological Review, 121(4), 649–675. <https://doi.org/10.1037/a0037665>

Vanderschoot, M., & Van Dessel, P. (2022). EMDR therapy and PTSD: A goal-directed predictive processing perspective. *Journal of EMDR Practice and Research*, 16(3), 108–122.

<https://doi.org/10.1891/EMDR-2022-0009>

Vasil, J., Badcock, P. B., Constant, A., Friston, K., & Ramstead, M. J. D. (2020). A world unto itself: Human communication as active inference. *Frontiers in Psychology*, 11, 417.

<https://doi.org/10.3389/fpsyg.2020.00417>

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J. (2020). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences*, 43, e90.

<https://doi.org/10.1017/S0140525X19001213>

Villiger, D. (2024). An integrative model of psychotherapeutic interventions based on a predictive processing framework. *Journal of Contemporary Psychotherapy*, 54(4), 295–304.

<https://doi.org/10.1007/s10879-024-09637-7>

- von Foerster, H. (Ed.). (1974). *Cybernetics of cybernetics: Or, the control of control and the communication of communication*. Biological Computer Laboratory, University of Illinois.
- von Foerster, H. (1981). *Observing systems*. Intersystems Publications.
- von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality* (pp. 17–40). W. W. Norton.
- von Sydow, K., Beher, S., & Retzlaff, R. (2024). Systemic psychotherapy: An introduction to its theoretical foundations and clinical practice. *Deutsches Ärzteblatt International*, 121, 783–792. <https://doi.org/10.3238/arztebl.m2024.0194>
- Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, 1464(1), 242–268. <https://doi.org/10.1111/nyas.14321>
- Wampler, K. S., Miller, R. B., & Seedall, R. B. (Eds.). (2020). *The handbook of systemic family therapy: Vol. 1. The profession of systemic family therapy*. John Wiley & Sons.
- Watzlawick, P., Weakland, J. H., & Fisch, R. (1974). *Change: Principles of problem formation and problem resolution*. W. W. Norton.
- Weakland, J. H., Fisch, R., Watzlawick, P., & Bodin, A. M. (1974). Brief therapy: Focused problem resolution. *Family Process*, 13(2), 141–168.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101(1), 34–52. <https://doi.org/10.1037/0033-295X.101.1.34>
- Wergin, J. F. (2019). Constructive disorientation. In *Deep learning in a disorienting world* (pp. 79–104). Cambridge University Press. <https://doi.org/10.1017/9781108647786.004>
- Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. MIT Press.

Wilkinson, S., Dodgson, G., & Meares, K. (2017). Predictive processing and the varieties of psychological trauma. *Frontiers in Psychology*, 8, 1840.

<https://doi.org/10.3389/fpsyg.2017.01840>

Williams, D. (2022). Is the brain an organ for free energy minimisation? *Philosophical Studies*, 179(5), 1693–1714. <https://doi.org/10.1007/s11098-021-01722-0>

Wittenborn, A. K., Hosseinichimeh, N., Rick, J. L., & Tseng, C.-F. (2020). Systems theory and methodology: Advancing the science of systemic family therapy. In K. S. Wampler, R. B. Miller, & R. B. Seedall (Eds.), *The handbook of systemic family therapy: Vol. 1. The profession of systemic family therapy*. John Wiley & Sons.

<https://doi.org/10.1002/9781119790181.ch5>

Witzel, C., Racey, C., & O'Regan, J. K. (2017). The most reasonable explanation of “the dress”: Implicit assumptions about illumination. *Journal of Vision*, 17(2), 1.

<https://doi.org/10.1167/17.2.1>

Appendix A: Background and evolution of Milan systemic therapy

A.1.1 The 1940s-1975: Early Influences and the Strategic-Systemic Approach

In the 1940s, North America found itself with increased concerns about mental health and family life following the Second World War. At the same time, several new ideas and developments were arising. Freud's idea of psychological disorders originating in childhood led to the pursuit of counseling children and their parents; marriage counseling had also been established in the 1930s (Nichols & Davis, 2020). Group dynamics, field theory, and the Gestalt school of perception were bringing a greater focus on the idea of groups (Nichols & Davis, 2020). These shifts, combined with the frustration of many therapists with lack of progress with the then-current largely psychoanalytic therapy models, and the realization by some researchers that "apparently irrational behaviour of an individual made complete sense when seen in the context of that person's close relationships", together led to the birth of family therapy (Možina, 2021; Doherty, 2020, Stratton, 2016).

The field would, over the following decades, evolve in a great many different directions, including structural family therapy (Minuchin), strategic family therapy (Haley), intergenerational family therapy (Bowen, Framo, and Boszormenyi-Nagy), experiential family therapy (Satir and Whitaker), behavioural family therapy (Jacobson and Heiman), cognitive-behavioural family therapy (Lieberman and Stuart), psychodynamic family therapy (Ackerman), psychoanalytic family therapy (Skynner and Sager), and notably, systemic family therapy (Boscolo, Cecchin, Prata, and Selvini Palazzoli) (Wittenborn et al., 2020; Bray & Stanton, 2009; Lebow & Sexton, 2016). Of course, this section will follow the evolution of the latter.

The first theoretical foundations of systemic therapy are the closely related but independent ideas of general systems theory and cybernetics (particularly homeostatic feedback loops), both developed mainly in the 1940s and 50s (Carr, 2016; von Bertalanffy, 1950; Wiener, 1948). These theories together contributed to the paradigmatic shift from an individual

framework for understanding the human world to a relational one (Carr, 2016; Hecker, Mims, & Boughner, 2015). For an explanation of GST and cybernetics and how they apply to MST, see §3.2.1 and §3.2.2.

GST and cybernetics were presented at the interdisciplinary Macy Foundation conferences in the 1940s (Goldenberg & Goldenberg, 2009). One of the attendees, anthropologist Gregory Bateson, envisioned how they both might be combined and applied to human interactions (Goldenberg & Goldenberg, 2009; Carr, 2016). In the early 1950s, Bateson formed a research team in Palo Alto, California, including communications researcher Jay Haley, anthropologist John Weakland, and psychiatrist Don Jackson, among others, in order to study paradox in communication (Wittenborn et al., 2020; Goldenberg & Goldenberg, 2009; Nichols & Davis, 2020). This group, which became known as the Palo Alto Group, synthesized cybernetic principles with Bertrand Russell's theory of logical types and Milton Erickson's work on hypnosis and paradox to develop a communicational approach to behaviour and psychopathology (Bateson et al., 1956; Ray & Simms, 2016). Central to this approach was the premise that if a family system can only be understood by accounting for both individual members and the interactional relationships between them, then communication — both verbal and nonverbal — constitutes the expression of those relationships (Wittenborn et al., 2020). The Palo Alto Group was focused mainly on researching communication with people who had schizophrenia; as Nichols & Davis note, "Families with schizophrenic members proved to be a fertile area for research because their pathological patterns of interaction were so magnified" (2020).

The Palo Alto Group's work marked a decisive departure from earlier family therapeutic models, which lacked a theoretical foundation (Doherty, 2020). New models subsequently began emerging and diverging, including those mentioned in the second paragraph of this section. One of the first ones in particular would become the basis of early Milan systemic therapy: the "strategic" approach of the Mental Research Institute (MRI), which was founded in 1959 by Don Jackson of the Palo Alto Group to extend the group's theoretical work into clinical practice (Boscolo et al., 1987; Nichols & Davis, 2020; Goldenberg & Goldenberg, 2009). The MRI team also included social worker Virginia Satir, communication specialist Jay Haley, anthropologist John Weakland, psychologist and communications theorist Paul

Watzlawick, and psychiatrist Richard Fisch (Nichols & Davis, 2020; Goldenberg & Goldenberg, 2009). The MRI's strategic approach held that problems were located not inside the individual or a pathology, but in the patterns of interaction between people, and changing the interactional pattern would change the symptom (Watzlawick, Beavin, & Jackson, 1967). The MRI team accordingly set aside the question of inner motivations and emotions to focus on observable communication and behaviour — adopting the 'black box' view that had become foundational in cybernetics (Ashby, 1956; Watzlawick et al., 1967). Symptoms were seen as paradoxical: they preserved the family's existing pattern while appearing to oppose it, and were maintained by the very attempts made to solve them — “more of the same” solutions (Watzlawick, Weakland, & Fisch, 1974). The therapist's task was therefore to interrupt the maintaining pattern through strategic interventions: prescribing the symptom, reframing its meaning, and assigning paradoxical tasks where resistance to the therapist required relinquishing the symptom (the “therapeutic paradox”; Haley, 1963). (Note that the MRI is today well associated with “brief therapy”, however this approach was crystallized later, with the first publication in 1974 (Watzlawick et al., 1974; Weakland et al., 1974). As this thesis follows the influences of Milan systemic therapy, I limit myself to describing the MRI's ideas that were available at the time that the Milan group state taking inspiration from them).

Meanwhile in Europe, Mara Selvini Palazzoli, an Italian psychoanalyst who had specialised in treating anorexia nervosa, was growing frustrated with the limitations of individual therapy for this population and began developing her own approach to working with families (Selvini Palazzoli, 1978). After learning about the ideas of Bateson and the MRI group, she formed the Center for the Study of the Family with the same strategic approach in Milan in 1971, together with psychiatrists Luigi Boscolo, Gianfranco Cecchin, and Giuliana Prata (Boscolo et al., 1987). The “Milan Group”, as they would be known as, drew particularly on Watzlawick, Beavin, and Jackson's *Pragmatics of Human Communication* in their early years following their foundation, between 1971 and 1975 (Boscolo et al., 1987). Some of the main ideas of this work, and the overall strategic approach that the Milan group adopted, were outlined in the previous paragraph (Boscolo et al., 1987; Boscolo & Bertrando, 1993; Selvini Palazzoli et al, 1978; Boscolo & Bertrando, 1996).

A.1.2 1975-1980: The Emergence of New Ideas and the Milan Systemic Approach

Already in the 1960s, new ideas in cybernetics were beginning to take root that would contribute to substantial shifts in systemic therapy after 1975. Maruyama (1963) drew attention to positive feedback — amplifying deviations rather than maintaining homeostasis — as equally important for understanding how systems change and evolve, a development he termed “the second cybernetics”. Another development emerged in the 1970s, confusingly termed “second-order cybernetics”, which included the observing system (von Foerster, 1974). It should be noted that within the family therapy literature, “second-order cybernetics” came to function as an umbrella term covering both Maruyama’s deviation-amplification and von Foerster’s observer-included epistemology (e.g., Boscolo & Bertrando, 1996; Hoffman, 1985). For a more detailed explanation of cybernetics and how it applies to MST, see §3.2.2.

Secondly, Bateson made his ideas more accessible by packaging and publishing his papers in his book *Steps to an Ecology of Mind* (Bateson, 1972). These aligned closely with second-order-cybernetics and brought a major shift particularly to the Milan group, who read this book in 1975 and noted that compared to the MRI ideas they had been using, Bateson’s original ideas seemed “both purer and more complex” (Boscolo & Bertrando, 1993, p. 85). They adapted their model beyond strategic thinking and interventions, and developed a new, pure systemic model, which became known as “The Milan Approach” (Boscolo & Bertrando, 1996, p. 8). It was at this point that they established their core tenets of hypothesizing, circularity, and neutrality, which continue to be used in systemic therapy today (Selvini Palazzoli et al., 1980; Telfener & Barazzetti, 2025).

It should be noted that the strategic approach continues to be a part of systemic therapy in the form of some of their enduring conceptual contributions, although the terminology around it may have evolved: for example, the practice of attending to nonverbal communication, and the ideas that context determines the meaning of utterances and behaviour, and that people construe the same event differently (Barbetta & Telfener, 2021; Boscolo & Bertrando, 1993; Selvini Palazzoli et al., 1978, p. 8; Watzlawick et al., 1967). However, the strategic interventions and praxis have largely been abandoned, in part due to ethical critiques of their manipulateness (Jackson & Landers, 2020). As Nichols & Davis note, the early strategic

approach involved “a fascination with strategies to outwit and provoke families into changing, with or without their cooperation” (2020, p. 79). The Milan group themselves moved away from this stance starting in the 1980s (Boscolo & Bertrando, 1993).

It was around the late 1970s that the Milan group’s work began to attract therapists outside Italy — some through visits and personal contact, others through the team’s published writings or direct collaboration. Several of these figures became major interpreters and disseminators of the systemic approach internationally, including David Campbell and Ros Draper (Campbell & Draper, 1985; Campbell, 2003) in the UK, Karl Tomm (Tomm, 1984, 1987) in Canada, and Lynn Hoffman (Hoffman, 1981, 1991) and Peggy Penn (Penn, 1982, 1985) in the United States.

A.1.3 Early 1980s: The Split and the Rise of the Observer

The 1980s brought more radical ideas. Perhaps most notable is the theory of Maturana (1980), whose ideas were introduced to systemic therapy by American psychologist Paul Dell, through his article “Family Theory and the Epistemology of Humberto Maturana” (1982) and then a deeper dive into it with the paper “Understanding Bateson and Maturana: Toward a Biological Foundation for the Social Sciences” (1985). Maturana’s main idea applied to family therapy was structural determinism, which is explained in §3.2.4.

Another key idea from Maturana’s theory, further developed by Ernst von Glasersfeld, is constructivism (Maturana & Varela, 1980; von Glasersfeld, 1984), which is explained in § 4.3.5. When this idea was brought to systemic therapy by Paul Watzlawick (1984), it triggered a shift which brought back greater emphasis to the importance of cognition and the assumptions people have about their problems. As Nichols and Davis state, “the goal of therapy changed from interrupting problematic patterns of interaction to helping clients find new perspectives on their lives” (2020).

Just as these new ideas were beginning to emerge, the Milan group split up in 1979, and each subgroup developed in a different theoretical direction (Boscolo & Bertrando, 1993). Selvini Palazzoli and Prata left the Milan center and retained the team’s earlier strategic orientation and prescriptive techniques based on first-order cybernetics. Together with Stefano Cirillo, Anna Maria Sorrentino, and Matteo Selvini, they later developed the

systemic-family-individual (SFI) approach now taught at the Scuola di Psicoterapia Mara Selvini Palazzoli in Milan (Carr, 2016; Sexton & Lebow, 2016). Meanwhile, Boscolo and Cecchin moved towards Maturana's, von Foerster's, and von Glasersfeld's ideas of second-order cybernetics, the observing system, and constructivism (Boscolo & Bertrando, 1993). As this thesis is interested in exploring the theoretical alignment between predictive processing and cybernetic ideas in therapy, I will follow the line of Boscolo and Cecchin and their successors.

They also found themselves having to revise some of their core tenets in light of these emerging ideas, resulting in the paper "Hypothesizing, Circularity, and Neutrality Revisited: An Invitation to Curiosity" (Cecchin, 1987). The main revisions made to these tenets are included in the explanations in §§3.2.2 and 3.3.2.

A.1.4 Early 1980s: Applications to Individual Therapy

The development of second-order cybernetics and constructivism caused a shift putting the observer (i.e., the individual) in the forefront. This triggered an evolution starting in the early 1980s, with several individual therapists beginning to show interest in family and couple therapy, and several systemic family therapists beginning to pay more attention to the individual members (Boscolo & Bertrando, 1996). It is worth noting that Bateson also gave importance to the inner world, the unconscious, and emotions (Bateson, 1972). However, contrary to Freud's unconscious which is made of drives (instinctual forces), Bateson's unconscious is made up of habits and individual premises, acquired through a process (Boscolo & Bertrando, 1996).

Another development furthered the greater focus on the individual: the "black box" theory, explained earlier, began to be questioned (Boscolo & Bertrando, 1996). The early family and systemic therapists had adopted the theory to eliminate the complexity inherent in personality theories, particularly psychoanalytic theory which dominated the field of psychotherapy (Nichols & Davis, 2020). However, dealing strictly with the relationships among persons while ignoring their motivations, fantasies, and emotions had "a reductionistic flavor" left many training programs with no clear guidelines for understanding and treating the individuals

that make up the systems (Boscolo & Bertrando, 1996, p. 19; Breunlin, 2016). Researchers began to criticize the “one-dimensional conception of family members exemplary types rather than [sic] complex individualities”, and the idea that “the family system could change its rules; the family members remained, more or less, the same” (Bertrando & Lini, 2021, p. 2).

Some systemic models quickly extended themselves to also include the individual and their psyche. The Milan original group maintained its clear focus on family and couple therapy, grounded in the belief that symptomatic behaviour was related to the behaviour of other family members, as well as in Bateson’s cybernetic epistemology, which held that any change in the family system would inevitably encompass all its members; dealing only with the individual was thought to restrict the therapist’s possibilities for helping (Boscolo & Bertrando, 1996). However, towards the late 1980s, the team became increasingly interested in individual therapy — motivated by curiosity, the emerging focus on the individual as observer, and the frequency with which trainees were bringing individual cases for supervision (Boscolo & Bertrando, 1996). As such, they adapted their model to allow for individual therapy as well [mention how / where?].

Today, contemporary Milan-tradition therapy works across multiple settings — family, couple, individual, and group — rather than being restricted to whole-family work (Bertrando & Lini, 2023; Telfener & Barazzetti, 2025). However, relationships “remain the basic focus of therapeutic intervention, even when the complexities of inner life are taken into consideration” (Bertrando & Lini, 2021, p. 2). For the purpose of this thesis, I will focus on examining how predictive processing may apply to systems of multiple individuals; however, it is useful to keep in mind that the individual psyche must still be considered and not ignored in favor of the overarching/overlying group dynamics.

A.1.5 Late 1980s, 1990s: Critique of Systemic Thinking and Postmodern Ideas

Building on constructivism, a similarly named and related idea called “social constructionism” would emerge in the 1980s and 1990s, which would also have a strong influence on systemic therapy (Boscolo & Bertrando, 1996). This concept is explained in §3.2.6. This was supported/butressed by the development of postmodern thinking, which radically criticized the therapist’s power and authority, fostering instead a non-expert, non-hierarchical, and “not

knowing” position for the therapist (Anderson and Goolishian, 1992). These ideas have influenced systemic therapy, and many systemic therapists found their ideas evolving more in the direction of narrative therapy, in some cases into something almost opposite to systemic therapy (Tomm, 2009; Seikkula & Olson, 2003). Some even consider the systemic model to be “outdated” and largely replaced by these new narrative models (Bertrando, 2020). However, other therapists argue that social constructionist and postmodern ideas are in fact compatible with systemic thinking, perhaps the former even rely on the latter; and today systemic therapy has incorporated elements of social constructionism and postmodernism into its practices (Bertrando & Lini, 2019; Bertrando & Lini, 2021).

Alongside these epistemological shifts, the Milan approach was challenged by two social developments starting during the mid-1980s (Campbell, 2003). The first was an increased awareness of abuse within families: child abuse and marital violence revealed that “the family could be a dangerous place to be” (Campbell, 2003, p. 19). The second was the emergence of feminist family therapy, which asserted issues such as the oppression of women in many family relationships and the gendered assumptions therapists may bring to their work (Goldner, 1988; Jones, 1993; Walters, Carter, Papp, & Silverstein, 1988). Together, these developments forced the field to reconsider the values, beliefs, and positions therapists bring into the consulting room (Campbell, 2003) — a reconsideration that was particularly difficult for the Milan approach because of its concept of *neutrality*, which was widely used but frequently misunderstood (Campbell, 2003). Feminist critics argued that neutrality inadvertently legitimised a status quo of male supremacy and the implicit endorsement of abuse (Goldner et al., 1990; Jones, 1993). The Milan response was not to abandon neutrality but to further refine and reformulate it — this is included in the explanation of the concept in §3.3.2 (Cecchin et al., 1992). Feminism thus joined social constructionism, narrative theory, and the postmodern critique of expertise as a major influence on the post-Milan tradition, and continues to be an important part of the practice, as we will see in the following two sections (Bertrando, 2000, 2007).

A.1.6 Early 2000s: the Rise of Disorder

The early 2000s saw one of the most basic assumptions of systemic thinking — the idea of *order* — placed under sustained pressure for the first time. As Bertrando and Lini (2021, p. 3) describe the situation in retrospect:

“We can see systemic understanding as based on the idea of order. If I think, like in first-order cybernetics, that the system functions according to embedded rules (Jackson, 1965), in order to maintain its homeostasis (Jackson, 1959), I am also thinking that in the system, an immanent order exists that I should discover. If I think, like in constructivism, that I am superimposing some order according to my internal coordinates or, like in social constructionism, that some order is created by the dance of social interaction, still I am seeing the world as a place needing order.”

In the early 2000s, this very notion of order was put into question. The postmodern and social-constructionist views had at first been seen favourably in therapeutic circles as liberating (Gergen, 1999), but the social conditions of the new millennium — recurrent economic and social crises — soon cast the absence of foundations in a more pessimistic light. It was experienced not only as freedom from rigid structure but also as a source of economic instability, loss of identity, existential uncertainty, and the forced “flexibility” demanded by contemporary working life (Bauman, 2000; Sennett, 1998).v

A.1.7 The Newer Models of Milan Systemic Therapy

After the original Milan team split, the Boscolo and Cecchin lineage being investigated in this thesis continued to evolve in two further contemporary branches: the systemic-dialogical model, and the institutional continuation at the Centro Milanese di Terapia della Famiglia (CMTF).

The systemic-dialogical lineage runs through psychiatrist Paolo Bertrando, Boscolo’s longtime collaborator and co-author of *The Times of Time* (Boscolo & Bertrando, 1993) and *Systemic Therapy with Individuals* (Boscolo & Bertrando, 1996). By 2003 Bertrando had left CMTF to

co-found Episteme in Turin (Bertrando & Arcelloni, 2006), and he is now director of the Scuola di Psicoterapia Sistemico-Dialogica in Bergamo (Lini & Bertrando, 2019), where he has worked with psychologist Claudia Lini to develop what they call the *systemic-dialogical model* of therapy.

By their own account, this model “originates from an evolution of Milan systemic therapy, through a reflection on Bakhtinian dialogue and on power and knowledge dynamics in therapeutic sessions” (Bertrando & Lini, 2019, p. 176), drawing its main theoretical references from Gregory Bateson, Mikhail Bakhtin, and Michel Foucault. With these new ideas, hypothesizing has been adapted into micro-hypothesizing: small, partial hypotheses generated moment-to-moment around emotional events in the session, rather than overarching interpretations of the family system. As well, the tenet of neutrality/curiosity has been replaced with positional responsibility: rather than refraining from taking a position, the therapist must take one and acknowledge it (Lini, 2022, p. 339). The therapist’s broader stance is described as *finding one’s place* — an integration of positioning with emotional awareness that acknowledges that the therapist cannot stand outside the system she works in (Lini & Bertrando, 2020). The model also drops much of the directive technical apparatus the original Milan team had built up — paradoxical prescriptions, rituals, reframings, and most rules of setting — in favour of what Bertrando and Lini describe as the “endless flow of dialogue”: a polyphonic, heteroglot exchange in which no voice (including the therapist’s) is treated as final (Bertrando & Lini, 2019).

Looking now to the other Milan lineage, the CMTF is the institutional continuation of the original Milan Centre founded in 1971 (Barbetta & Telfener, 2020), which Boscolo and Cecchin ran after the 1979 split. Its current generation includes psychologist Umberta Telfener — a former trainee of Cecchin and Boscolo who later joined the teaching faculty — together with psychotherapist Pietro Barbetta and other colleagues. The CMTF lineage has retained the three original directives of the 1980s: “Hypothesizing, circularity, and curiosity are the three points of reference of the Milan approach that remain unchanged” (Barbetta & Telfener, 2020, p. 11).

To these three directives the Centre has added two contemporary developments of note. The first is the corporeal turn, the integration of “embodied experience as preconceptual know-how from which concepts are structured” along with attention to “the connections between bodies and social issues that enact forms of knowledge and understanding” (Barbetta & Telfener, 2020, abstract). Secondly, they also revise the concept of neutrality/curiosity, defining it as a complementary stance between epistemology and ontology: the therapist works to deactivate her own presuppositions and prejudices so that clients can disclose themselves on their own terms, while also recognising that in contexts of violence, injustice, or strong moral intensity she must take a position, take a side, and remain aware of the categories she uses (Barbetta & Telfener, 2020). The CMTF’s contemporary theoretical references — Bateson, Maturana and Varela, von Foerster, Prigogine and Stengers, Deleuze and Guattari, Foucault, Viveiros de Castro, De Landa, Latour, Stiegler, and Scarry (Barbetta & Telfener, 2020) — show a broader philosophical engagement than the Bertrando/Lini lineage, while retaining the cybernetic-constructivist commitments that have defined the Milan tradition since the late 1970s.