

COMPUTATIONAL APPROACHES TO *D. RERIO* RETINAL ORGANOGENESIS

by

Michael Mattocks

A thesis submitted in conformity with the requirements  
for the degree of Doctor of Philosophy  
Graduate Department of Cell and Systems Biology  
University of Toronto

# Abstract

Computational Approaches to *D. rerio* Retinal Organogenesis

Michael Mattocks

Doctor of Philosophy

Graduate Department of Cell and Systems Biology

University of Toronto

2020

Increasing integration of sophisticated statistical and computational techniques into the analysis of cell and molecular biological data has created many opportunities for explaining organogenic phenomena by numerical analysis. This thesis first examines the suitability of the most developed of these explanations for the development of the *D. rerio* eye, and demonstrates that it is neither formally correct nor explanatory. By documenting the phenomena associated with postembryonic retinal neurogenesis, a framework of observations which cellular models of this process would be called upon to explain is generated. The desiderata of a model comparison framework suitable for evaluating the evidence supplied by confocal datasets from postembryonic fish are explicated. Nuclear dynamics in a mutant fish, *rys*, are

This work is dedicated to the memory of my brother Gareth Akerman.

## Acknowledgements

Acknowledgements text

# Contents

<b>Introductory Notes</b>	<b>1</b>
<b>I Modelling Studies</b>	<b>2</b>
<b>Précis of Modelling Studies</b>	<b>3</b>
<b>1 Canonical CMZ RPC phenomena and their explanations</b>	<b>5</b>
1.1 The Harris Stochastic Mitotic Mode Explanation (SMME) . . . . .	5
1.2 Explanations for RPC function in 2009 and the drive to unification . . . . .	6
1.3 Canonical vertebrate RPC phenomena: the RPC “morphogenetic alphabet” . . . . .	7
1.3.1 Proliferative phenomena . . . . .	8
1.3.2 Specificative phenomena . . . . .	9
1.3.3 Other morphogenetic phenomena . . . . .	11
1.4 Macromolecular mechanistic explanations for RPC phenomena . . . . .	12
1.4.1 Transcription factor networks . . . . .	12
1.4.2 Intercellular signalling networks . . . . .	13
1.4.3 Patterning mechanisms . . . . .	14
1.4.4 Chromatin dynamics . . . . .	15
1.5 A unified theory of RPC function? “Blurring” to order . . . . .	15
1.6 Explanatory Strategy and Intent of the SMME . . . . .	17
<b>2 “Stochastic mitotic mode” models do not explain zebrafish retinal progenitor lineage outcomes</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Results and Discussion . . . . .	21
2.2.1 Gomes SSM: Ancestral Model of the SMME SSMs . . . . .	22
2.2.2 He SSM: Explaining variability in zebrafish neural retina lineage size . . . . .	22
2.2.3 Boije SSM: Explaining variability in zebrafish RPC fate outcomes . . . . .	25
2.2.4 Model selection demonstrates the SMME is not the best available explanation for RPC lineage outcomes . . . . .	26
2.2.5 SMME SSMs cannot explain the post-embryonic phase of CMZ-driven zebrafish retinal formation . . . . .	29
2.3 Conclusion . . . . .	33

<b>3 Toward a computational CMZ model comparison framework</b>	<b>36</b>
3.1 SMME Postmortem: a wrong turn at Gomes . . . . .	36
3.2 Implications of the SMME's failure for modelling CMZ RPCs . . . . .	38
3.3 Desiridata for spatial CMZ models in a putative model comparison framework . . . . .	40
3.3.1 Spatial dimension of the models: the "slice model" . . . . .	40
3.3.2 Temporal resolution of the models . . . . .	41
3.4 Bayesian decisionmaking for structuring models under uncertainty . . . . .	42
<b>4 Bayesian periodization of postembryonic CMZ activity</b>	<b>44</b>
4.1 Preliminaries: Calculation of Bayesian evidence militates for independent log-Normal modelling of CMZ parameters . . . . .	44
4.2 Survey of CMZ population and gross retinal contribution . . . . .	48
4.3 Periodization of postembryonic CMZ activity by Galilean Monte Carlo Nested Sampling .	50
4.4 Characterisation of asymmetrical CMZ population and retinal contribution dynamics .	51
4.5 Investigating CMZ RPC lineage outcomes . . . . .	53
4.6 Microglial apoptotic fate of <i>D. rerio</i> retinal neurons & functional significance of CMZ activity . . . . .	53
4.7 Toward cell-based computational models of the CMZ . . . . .	56
<b>5 Mutant npat results in nucleosome positioning defects in <i>D. rerio</i> CMZ progenitors, preventing specification but not proliferation</b>	<b>57</b>
5.1 Introduction . . . . .	57
5.2 Results . . . . .	58
5.2.1 The microphthalmic zebrafish line rys is an npat mutant . . . . .	58
5.2.2 rys CMZ cells display altered nuclear morphology and progenitor marker expression .	59
5.2.3 Proliferating cells in rys embryos display altered cell cycle parameters; those in the CMZ have a dramatically truncated S-phase . . . . .	59
5.2.4 rys embryos express higher levels of core histone transcripts, particularly polyadenylated transcripts, and histone proteins . . . . .	60
5.2.5 rys embryos display altered nucleosome positioning and chromatin organisation .	60
5.2.6 The rys CMZ does not contribute to the differentiated neural retina; the central retina of rys is disorganised . . . . .	63
5.2.7 The fate of rys CMZ cells is caspase-dependent apoptosis and subsequent phagocytosis by microglia . . . . .	63
5.3 Discussion . . . . .	64
<b>6 Inferring and modelling nuclear dynamics in retinal progenitors</b>	<b>67</b>
6.1 Intro . . . . .	67
6.2 Rys . . . . .	67
6.3 Histones and DNA sequence . . . . .	67
6.4 Inferential logic . . . . .	67
6.5 Technical caveats . . . . .	67
6.6 Separability of cell cycle and specification . . . . .	67
6.7 Frontiers of nuclear dynamics . . . . .	68

6.8	Integrating nuclear dynamics into models of CMZ- abstraction and biosemiotics . . . . .	68
<b>II</b>	<b>Software Technical Reports</b>	<b>69</b>
<b>7</b>	<b>GMC_NS.jl</b>	<b>70</b>
7.1	Implementation notes . . . . .	70
7.2	Ensemble, Model, and Model Record interfaces . . . . .	71
7.3	Usage notes . . . . .	71
<b>8</b>	<b>BioBackgroundModels.jl</b>	<b>72</b>
<b>9</b>	<b>BioMotifInference.jl: Independent component analysis motif inference by nested sampling for Julia</b>	<b>73</b>
9.1	Introduction . . . . .	73
9.2	Implementation of the nested sampling algorithm . . . . .	73
9.3	Usage notes . . . . .	74
9.4	Recovery of spiked motifs . . . . .	74
9.5	Recovery of promoter motifs . . . . .	74
<b>III</b>	<b>Supplementary Materials</b>	<b>75</b>
<b>10</b>	<b>Supplementary materials for Chapter 2</b>	<b>76</b>
10.1	Materials and methods . . . . .	76
10.1.1	Zebrafish husbandry . . . . .	76
10.1.2	Proliferative RPC Histochemistry . . . . .	76
10.1.2.1	Anti-PCNA histochemistry . . . . .	76
10.1.2.2	Cumulative thymidine analogue labelling for estimation of CMZ RPC cell cycle length . . . . .	77
10.1.2.3	Whole retina thymidine analogue labelling of post-embryonic CMZ contributions . . . . .	77
10.1.3	Confocal microscopy and image analysis . . . . .	77
10.1.4	Estimation of CMZ annular population size . . . . .	78
10.1.5	Modelling lens growth . . . . .	78
10.1.6	Estimation of 3dpf CMZ cell cycle length . . . . .	78
10.1.7	CHASTE Simulations . . . . .	79
10.1.7.1	Project code . . . . .	79
10.1.7.2	SPSA optimisation of models . . . . .	79
10.2	Supporting information . . . . .	81
10.2.1	S1 File. . . . .	81
<b>11</b>	<b>Supplementary materials for Chapter 4</b>	<b>86</b>
11.1	Description of the computational cluster used in the work . . . . .	86
11.2	Developmental progression of naso-temporal population asymmetry in the CMZ . . . . .	86
11.3	Methods . . . . .	86

11.3.1 Statistical Analyses . . . . .	86
<b>12 Supplementary material for Chapter 5</b>	<b>88</b>
<b>13 Theoretical Appendix A: Statistical methods and model theory</b>	<b>89</b>
13.1 Statistical Theory . . . . .	89
13.1.1 Bayesian parameter estimation . . . . .	89
13.1.1.1 Problems with frequentist inference using normal models of sample data .	89
13.1.1.2 The Bayesian approach to normal models of unknown mean and variance	90
13.1.2 Bayesian Epistemological View on Model Comparison . . . . .	90
13.2 Model Theory . . . . .	90
13.2.1 Model optimization . . . . .	90
13.2.1.1 MLE . . . . .	90
13.2.2 Overfitting, Underfitting . . . . .	90
13.2.3 Monte Carlo simulation . . . . .	90
13.2.4 The Akaike Information Criterion For Model Selection . . . . .	90
13.2.5 Simple Stochastic Models . . . . .	90
13.3 Traditional roots of the SMM: Till and McCulloch and the simple stochastic model (SSM)	92
<b>14 Theoretical Appendix B: Metaphysical Arguments</b>	<b>94</b>
14.0.1 Metaphysics, Epistemology . . . . .	96
14.0.2 Bayesian Epistemological View on Statistical Generalisation . . . . .	96
14.1 Study Structure in the Light of Tradition . . . . .	97
14.1.1 Feyerabend's Scientific Traditions . . . . .	99
14.1.2 Classical and Molecular Genetics: Shared Traditions, Plural Methodologies .	101
14.1.3 Sense of "Tradition" in this study . . . . .	102
14.2 EHJMEx as a Subtraditional Metascientific Unit of Molecular Biological Theory .	104
14.2.1 Law-based and Causal Mechanistic Explanations . . . . .	104
14.2.1.1 Law-based MEx . . . . .	105
14.2.1.2 Causal MEx (CauMEx) . . . . .	105
14.2.2 (Extended) Heterogenous Joint Mechanistic Explanations ((E)HJMEx) .	109
14.2.2.1 Joint MEx (JMEx) . . . . .	110
14.2.2.2 Heterogenous JMEx (HJMEx) . . . . .	112
14.2.2.3 Extended HJMEx (EHJMEx) . . . . .	118
14.3 The Feyerabendian Modeller . . . . .	119
14.3.1 Laws, Rules, Models, and Games . . . . .	120
14.3.2 NetLogo Agent Model Simulation . . . . .	121
14.3.3 A Global Approach to Model Selection . . . . .	122
14.3.4 The Global Semiotic Agent Model . . . . .	124
14.4 The Systems Biology Encounter - The Molecular Biological Tradition Under Impact .	127
14.4.1 Historical sketch of the biological encounter with complexity . . . . .	127
14.4.2 Explanatory Models in the Stem Cell Biology Tradition . . . . .	127
14.5 "Metaphysical ingredients" encountered in MEx . . . . .	127
14.5.1 Information theory . . . . .	127

14.5.2	RPC Fate Specification as a “Stochastic” Process . . . . .	128
14.5.2.1	Chance versus Randomness . . . . .	129
14.5.2.2	Chance in molecular mechanisms . . . . .	130
14.5.2.3	Quantum indeterminacy - Is relevant to the ISE? . . . . .	132
14.5.2.4	Randomness in RPC fate specification . . . . .	132
14.5.2.5	Summary: “Stochastic” or “variable”? . . . . .	133
14.5.3	Biological Variability . . . . .	134
14.5.3.1	Noise . . . . .	135
14.5.4	Time . . . . .	135
14.5.5	Semiosis, Meaning . . . . .	135
14.5.6	Emergence, self-organisation . . . . .	135
14.5.7	Scale Hierarchies, Specification Hierarchies . . . . .	135
14.5.8	Waddington’s topological model of development . . . . .	135
14.5.9	Morphogenetic fields . . . . .	136
14.5.10	Physical explanations . . . . .	136
14.6	Implications of Limits in Theory and Practice For Model Selection . . . . .	136
14.6.1	Nicholas Rescher’s Account of Scientific Progress . . . . .	136
14.6.2	Implications of Theoretical Limits of Science . . . . .	136
14.6.3	Implications of Thermo-economic Limits of Science . . . . .	136
14.6.4	Practical computational limits . . . . .	136

## Bibliography

**137**

# List of Tables

2.1	AIC values for models assessed against training and test datasets . . . . .	29
4.1	Likelihood ratio comparison between normal and log-normal models of retinal population parameters . . . . .	45
4.2	Evidence favours log-normal models of retinal population parameters . . . . .	45
4.3	Evidence favours uncorrelated linear models of CMZ-population and retinal volume over time . . . . .	47
4.4	Evidence favours whole-CMZ linear cycle models over separate D/V models .	51

# List of Figures

1.1	Histogenetic birth order of retinal neurons . . . . .	10
2.1	Structure of Gomes SSM . . . . .	23
2.2	Structure of He SSM . . . . .	24
2.3	Structure of Boije SSM . . . . .	25
2.4	Model comparison: the SPSA-optimised He SSM and a deterministic alternative . . . . .	28
2.5	Most zebrafish retinal neurons are contributed by the CMZ between one and three months of age . . . . .	30
2.6	Per-lineage probabilities of mitoses, He et al. observations compared to model output . . . . .	31
2.7	CMZ population of proliferating RPCs: estimates from observations and simulated Wan-type CMZs . . . . .	33
4.1	CMZ population and retinal volume estimates are uncorrelated at 3dpf . . . . .	47
4.2	Population and activity of the CMZ over the first year of <i>D.rerio</i> life . . . . .	49
4.3	Developmental progression of dorso-ventral population asymmetry in the CMZ. . . . .	52
4.4	Representative 23dpf lineage marker confocal micrographs . . . . .	54
4.5	CMZ contributions to the neural retina over time by layer and lineage marker . . . . .	55
5.1	rys chromosomes are differentially enriched and depleted of nucleosome position density and occupancy. . . . .	61
5.2	Novel nucleosome positions in rys occur in similar numbers to those lost from sibs. . . . .	62
11.1	Developmental progression of naso-temporal population asymmetry in the CMZ. . . . .	87
13.1	Simple stochastic stem cell model, representing probabilities of cell division events, excerpted from Fagan 2013 pg. 61. Black circles denote proliferative cells, while white and grey circles denote different types of postmitotic offspring. “Number of progeny in P” is the number of mitotic offspring produced by each type of division. The probability of each division type must sum to 1, as all possibilities are represented, granting that the division types are defined by the postdivisional mitotic history of the offspring. . . . .	91
14.1	The consensus view of mechanisms, excerpted from [Fag13, p.96] (a) Two levels: component x’s and overall mechanism S. (b) Downward expansion: three mechanistic levels. . . . .	106

14.2 Levels of causal relations in MEx, excerpted from [Fag13, p.102]. (a) System-level: mechanism S $\Psi$ -ing produces phenomenon P. (b) Component-level: component x's causally interact ( $\phi$ -ing) to effect one another. (c) Inter-level: component $\phi$ -ing x's make a difference to S's outcome (P). (d) Constitutive MEx: component x's interact ( $\phi$ -ing) to constitute mechanism S's $\Psi$ -ing . . . . .	108
14.3 Cellular systems model-construction, excerpted from [Fag15b, p.7]. The system of equations and subsequent steps are based on a 2-element wiring diagram. . . . .	113
14.4 Relationship between processes and biological structures, excerpted from [Lar92]. The six cell behaviors and two tissue strategies from which all morphology results. . . . .	125
14.5 The complex system of interactions underlying the epigenetic landscape, excerpted from [Wad57]. The pegs in the ground represent genes; the strings leading from them the chemical tendencies which the genes produce. The modelling of the epigenetic landscape, which slopes down from above one's head towards the distance, is controlled by the pull of these numerous guy-ropes which are ultimately anchored to the genes. . . . .	136

## Introductory Notes

### Thesis Guide

This document is split into three parts. Part I contains results and discussion pertaining to empirical modelling studies that will be of interest to committee members and developmental biologists. Part II contains technical reports pertaining to novel software that was written in order to perform the analyses presented in Part I. Part III contains a variety of supplementary materials arising from Parts I and II. These include detailed descriptions of the methods used in Part I, less-technical explanations of relevant statistical and model theory, the source code of all software and analyses, and the bibliography. The source code has been omitted from the print document.

Readers of the .pdf document will find some text unobtrusively highlighted in plum throughout the thesis. Section indices highlighted in this way link to the specified section. Technical terms have also been highlighted when pertinent material is available in Part III to explain them for the reader who may be unfamiliar.

### Terminology and Style

Throughout [Chapter 1](#), [Chapter 2](#), and [Chapter 3](#), I have used the appellation "Harris" when referring to the output of William Harris' research group. This is a large body of research spanning several decades, and involves many co-authors. My use of "Harris" here is a convenience, as Harris is the only common author across the period in question, and presumably the agent carrying these ideas forward from project to project. It is not intended to slight or minimize the contributions of any of the other members of Harris' group (many of whom are now senior scientists in their own right). I have used "Raff" in an identical sense in referring to the Raff group's work in [Chapter 3](#).

I have preferred the terms "specification", "determination", and their derivatives, to refer to cells assuming a particular lineage fate. "Differentiation" is well-understood, but ambiguous, and often understood to relate to the mitotic event itself, which I generally do not intend. "A cell has specified" means it has assumed a stable macromolecular identity or "fate". This term is intended to correspond exactly with the appearance of stable markers of cell type.

There are a number of specialised philosophical terms that appear in quoted material in the Part III. I have made occasional use of some use of them to save space. These are as follows:

iff - "if and only if"

explanadum - the fact to be explained

explanans - the explanation itself

Unless otherwise noted, formatting of quoted material is preserved, so that italic emphases appear in the original. An exception to this is citations in original material, which have been removed wherever irrelevant or uninformative (ie. numbered references). I have indicated my own editorial comments and alterations to quoted material with [square braces].

# **Part I**

# **Modelling Studies**

## Précis of Modelling Studies

The primary concern of this thesis is developing and assessing computational models of retinal progenitor cell (RPC) phenomena. While many different, well substantiated explanations for various aspects of eye development exist, these explanations have usually been supported by assessing the frequentist significance of observed effects, not by building formally testable models of the phenomena themselves. Of the explanations which could be subject to model comparison, virtually all of these are derivatives of the Simple Stochastic Model (SSM), dating to the earliest work, performed by Till and McCulloch [TMS64].

The explosion of molecular biological information has massively increased the number of parameters that might be included in RPC models. The question of how RPC activity might be successfully predicted and controlled *in vivo* may revolve around finding adequate models that explain retinogenetic processes. The development and assessment of such models is, therefore, of fundamental basic and applied interest. This is particularly so, in light of increasing interest in entraining RPCs for the purpose of retinal regenerative medicine. Our group is particularly interested in the possibility that RPCs located in the peripheral retinal annulus of the circumferential marginal zone (CMZ), which are present in zebrafish and mammals alike, could be harnessed for retinal repair.

It is surprising that, despite the use of SSMs, the literature on neural stem cells contains virtually no systematic statistical approaches to the construction, optimization, or comparison of models. This thesis is an attempt to address this problem. It proceeds in three basic strokes. Firstly, Chapters 1 and 2 evaluate the existing explanations of RPC behaviour, and find that those with formal model components are deficient, and cannot explain the activities of RPCs in the CMZ. Secondly, Chapters 3 and 4 propose a general CMZ modelling framework under which the Bayesian evidence for different model-hypotheses might be assessed, and, by testing a variety of population-level hypotheses, provide guidance and develop the methods required to test cell-based hypotheses. Lastly, Chapters 6, 5, and ?? introduce the zebrafish CMZ mutant *rys*, and apply some of the methods developed in the second stroke to explain aspects of the aberrant morphology and behaviour of *rys* RPCs.

Chapter 1 begins by summarizing the range of explanations that have been offered for RPC activities, and introduces the primary set of formal models that have recently been used to favour an explanation centered around a supposed stochastic mitotic mode of RPCs. Chapter 2 elucidates the structure and development of the stochastic mitotic mode explanation (SMME) models, finding fundamental logical errors in their interpretation of "stochasticity". Moreover, by pursuing a basic information theoretical model selection approach, it is demonstrated that the SMME is not even the best available explanation relying on this flawed notion of stochasticity.

Chapter 3 discusses the implications of Chapter 2 for models of RPCs generally, and in the CMZ particularly; it also recommends a better model selection approach from Bayesian theory, nested sampling. Chapter 4 surveys the activity of RPCs in the post-embryonic zebrafish CMZ, developing Bayesian methods for doing so. These methods include the use of nested sampling to solve the long-standing problem of estimating cell cycle parameters of subpopulations of cells from cumulative thymidine labelling measurements of the entire super-population. It concludes with recommendations for future modelling approaches to the CMZ, structured by the foregoing Bayesian analysis.

Chapter 5 identifies the causative mutation in *rys* and shows that the zebrafish CMZ mutant *rys* CMZ RPCs have disorganized chromatin, likely blocking differentiation but not proliferation. Nested sampling is used to demonstrate that the nuclear phenotype is likely the consequence of a shift in

nucleosome histone composition in *rys* progenitors. ?? discusses some of the theoretical issues raised by the *rys* analysis, and their implications for modelling retinogenesis more broadly, especially in integrating nuclear dynamics into models of RPCs.

# Chapter 1

## Canonical CMZ RPC phenomena and their explanations

### 1.1 The Harris Stochastic Mitotic Mode Explanation (SMME)

Chapter 2 comprises a critical examination of the best-developed theory of *D. rerio* retinal progenitor cell (RPC) function, and a broader, global general modelling approach motivated by the examination’s findings. The theory in question originates with preëminent retinal biologist William Harris’ research group, referred to hereafter Stochastic Mitotic Mode Explanation (SMME). This theory purports to explain the function of zebrafish RPCs in terms of the stochastic effects of two particular transcription factors.

The SMME for RPC function is of fundamental theoretical and practical interest. It arises from a concerted effort by the Harris group to make sense of the difficult problem of explaining how a complex tissue like a retina can arise from a field of similar proliferating cells. In 2009, Harris coauthored a detailed review chapter, documenting the bewildering array of macromolecules and cellular processes thought to be involved in RPC function [AH09]. This enumeration contains many caveats and notes that the effects of particular macromolecules routinely differ between developmental stages, cell types, organisms, and so forth. At this time, no clear, detailed, comprehensive models of RPC function had been advanced, and the review is typified by statements like “It is difficult to reconcile all the studies on the initiation and spread of neurogenesis in a single model.”<sup>1</sup> It is therefore remarkable that, over the next nine years, the Harris group would go on to promulgate a simple stochastic model of zebrafish RPC function invoking only two named macromolecules.

Harris thus seems to be making a bold attempt to cut the Gordian knot of conflicting evidential threads and advance a simple, comprehensible, “mind-sized” model of RPC function. In effect, Harris’ explanation for zebrafish RPC function is a microcosm of the broader promise of “Systems Biology” to make sense of the contemporary welter of conflicting datasets. By using sophisticated mathematical methods drawn from information and complexity theories, the apparent confusion will be clarified and underlying molecular mechanisms will be revealed. Given Harris’ track record and preeminence in the

---

<sup>1</sup>This was in no sense a problem with Harris’ understanding. A similarly high-level review coauthored by Pam Raymond [AR08] concluded, with regard to models of photoreceptor fate specification: “The data reviewed in the preceding sections indicate that a ‘one-size-fits-all’ model is not possible...”

field, we have good reason to take seriously the possibility that he has succeeded. Examining whether this is the case is our first priority. In order to appreciate the scope of his theoretical maneuver, and possible alternatives to it, it is necessary to begin by summarising the state of the art at the time of his 2009 review (as well as relevant subsequent additions).

## 1.2 Explanations for RPC function in 2009 and the drive to unification

In many ways, molecular biologists have been attempting to explain the same remarkable features of retinal progenitor cells for decades. Animal retinas, having relatively well-understood functions and highly stereotypical structures, seem like highly tractable tissues for typical molecular biological explanations. With well defined cell types present in tightly regulated, neurotopologically limited proportions and organisations, it is no surprise that both theoretically-inclined molecular biologists and clinically-inclined regenerative medical practitioners have been keenly interested in the retina<sup>2</sup>. The high level of regular, easily detected order in eyes seemed to suggest a similar level of order and regularity in the macromolecular processes which underlay the formation of the tissue. Retinal biologists have long offered explanations suggesting that RPCs are more-or-less identical and go through highly stereotypical macromolecular processes. It is, however, the persistently observed departures from this (by now, obviously naïve) conception that have occupied most of our attention.

We may crudely gather the RPC-related phenomena studied by biologists under the headings of proliferation, specification, and organisation. A molecular biologist guided by a preference for cognitive simplicity (that is, Occam’s Razor) will reasonably suppose that the simplest explanation for the regularity of retinal development is that any given RPC is executing the same strict “developmental program” as its neighbours. In other words, any one RPC lineages’ proliferative and specificative outcomes are the same as any other RPC lineage. If every progenitor produces a similar number of cells, and the progeny are specified in similar proportions, well-understood principles of cellular adhesion could understandably give rise to the characteristic organisation we observe in animal retinas. However, by the 1980s, vertebrate lineage tracing experiments were routinely revealing a surprising degree of inter-lineage variability in many neural progenitor systems, not least of which was the retina. In their seminal 1987 paper, David Turner and Connie Cepko, using retroviral lineage labelling techniques, demonstrated that individual RPC lineages in rats had a wide range of proliferative and specificative outcomes [TC87]; Harris’ group confirmed this result in *Xenopus* the subsequent year [HBEH88], suggesting this variability was a common feature of vertebrate RPC function.

Indeed, at this point, we find fairly clear accounts of what retinal biologists took their theoretical options to be in explaining this variability. As Harris’ 1988 report states:

Changes in cell character associated with cell type diversification may be controlled in an autonomous way, reflecting either a temporal program inside the cell (Temple and Raff, 1986), the asymmetrical segregation of cytoplasmic determinants (Strome and Wood, 1983; Sulston and Horvitz, 1977), stochastic events inside the cell (Suda et al., 1984), or some combination of these processes. Alternatively, cell type may be controlled in a nonautonomous way, as in cases in which the extracellular environment (Doupe et al., 1985) or cellular interactions (Ready et al., 1976) elicit or limit cell fate. With its multiplicity of cell types, the vertebrate

---

<sup>2</sup>Indeed, if central neuroregenerative medicine is to become a clinical possibility, it seems likely that the theoretical and practical issues are most likely to be resolved in eyes before other areas of the CNS.

nervous system would seem to require the ultimate sophistication in its means of cellular determination. [HBEH88]

It is striking, then, that Harris' review of the literature two decades later describes the situation similarly:

Once differentiation is initiated, regulatory mechanisms within the retina ensure that progenitors retain the capacity to undergo more divisions, in parallel with churning out differentiated cells, and that progenitors cease dividing at variable times. There is still debate about the extent of early programming that allows progenitors to step through a series of stereotypical divisions and the extent of regulation from within the whole retina. The production of differentiated cells alters the retinal environment with time...

Moreover, cells from the same clone do not all differentiate at the same time, suggesting three possibilities: a stochastic mechanism for the decision to differentiate, exposure of the two daughters to different environments, or asymmetric inheritance of determinants. [AH09]

In the same paper, he states that the “simple structure and accessibility of the retina make it a useful model to study cell division and differentiation, and as a result most aspects of this have been studied, from lineage tracing of progenitors, to the morphological aspects of division, to the molecular mechanisms involved.” Thus, by Harris' own account, some twenty years of additional research into almost every variety of macromolecular explanation for a huge range of RPC-related phenomena had not provided any means to narrow down the possibilities he had already laid out in 1988. We still have Raff's temporal program (“early programming … step[ping] through a series of stereotypical divisions”), asymmetric segregation of cytoplasmic inheritants during mitotic events, “stochastic” events internal to the cells, and possible “environmental” extracellular determinants. While the number of particular macromolecules functionally implicated in proliferative, specificative, and organisational RPC phenomena had greatly expanded, this had not provided any means to differentiate between these theoretical options. This is only one example of a general phenomenon experienced by molecular biologists, in which enumerationist research programs, directed at producing more and more facts about macromolecular involvement in cellular phenomena, have failed to generate additional theoretical understanding [Kan06]. Harris' SMME therefore represents an example of a “Systems” biological explanation, in which biologists apply the analytical and interpretative methods of the physical and mathematical sciences in an effort to resolve the problems posed by biological complexity [Mor09].

Before proceeding to the SMME itself, let us briefly summarise the diversity of phenomena implicated in RPC function by 2009, as well as the panoply of mechanisms offered as explanations. In doing so, it will become clear what has been elided in the SMME, and what may need to be restored in any alternative modelling approach.

### 1.3 Canonical vertebrate RPC phenomena: the RPC “molecular genetic alphabet”

The majority of our knowledge of RPC behaviour stems from histological observation employing a limited number of techniques. Simple observations of mitotic figures in a variety of animals had, by the 1950s, revealed the surprising diversity of RPC proliferative phenomena across vertebrate clades. However, it was the advent of lineage tracing techniques, particularly those marking single clonal lineages in whole

retinae, and the extensive use of these techniques in the 1980s-90s, that formed most of the basic body of observations that any macromolecular explanation is now called upon to account for.

Since the majority of vertebrate retinas of biomedical interest are mammalian, and these retinae are fully formed in an early developmental period, RPC behaviour has been best-studied in an embryonic and early developmental context. Here, vertebrate RPCs are derived from the eye field population of the early neural plate and later neural tube. This population is separated into left and right eye primordia, which in turn pouch outwards toward the ectoderm, and, in conjunction with the lens placode (itself induced from the ectoderm), form the optic vesicle. The primitive eye is formed when this vesicle completes a complex morphological folding process, resulting in the cup-shaped structure of the retina [Cav18]. During this process, the cells of the neural retina are differentiated from the overlying retinal pigmented epithelium (RPE). Sometime after the formation of the retinal cup, RPCs begin to exit the cell cycle and are specified as retinal neurons. Studies of this early period revealed numerous difficult-to-explain features of RPC behaviour. Following Larsen's observation that tissue form is attributable to only six behaviours [Lar92], which she describes as the "morphogenetic alphabet", I have categorised RPC phenomena as relating to proliferation, specification, migration, growth, death, and extracellular matrix formation<sup>3</sup>. Since the vast majority of reported phenomena fall under the first two categories, those belonging to the last four are described collectively.

### 1.3.1 Proliferative phenomena

Clonal lineage tracing experiments have reliably found that vertebrate RPCs give rise to highly variable numbers of offspring over the collective proliferative "lifetime" of their descendants. The most dramatic of these findings found that rat RPC lineage sizes vary across two orders of magnitude *in vivo*, from 1 to over 200 [TSC90]<sup>4</sup>. The physical organisation of these clones is complex; as detailed below, RPC progeny may appear in any of the 3 retinal layers in a wide variety of specified fate combinations, and may engage in short-range migrations to appropriate positions for their specified fates. Most clonal lineages are "extinguished"; that is, after some time, all of its members have become postmitotic. However, it has long been noted that not all cells produced by RPCs are strictly postmitotic neurons; specified Müller glia retain the ability to reenter the cell cycle in response to stimuli (normally, retinal damage) [DC00, FR03], and peripheral CMZ RPCs remain proliferative in those vertebrates whose eyes grow beyond early development (notably in frogs and fish, while the chick retina has a CMZ of more limited output [FR00]). Therefore, some clonal lineages may be organised into clumps associated with Müller responses, while others in frogs and fish may continue to be "plated out" in a more-or-less linear manner at the retinal periphery for as long as the lineage "lives" [CHW11]. This ongoing RPC contribution to peripheral neurogenesis has long been recognised, so that by 1954 we find the following remarkable statement casually introducing a study of unusual mitoses in the retina of a deepsea fish:

It is conventional<sup>5</sup> to hold that the growth of the vertebrate retina is only possible due to the presence, in this tissue, of a peripheral germinal zone. In this region, young elements

---

<sup>3</sup>I have renamed Larsen's categories to clarify them for the retinal context, but retained her conceptual scheme, which is discussed in more detail in the [Theoretical Appendix](#).

<sup>4</sup>While at least some of this variability must be related to differential integration of lineage markers into "older" (giving fewer offspring) and "newer" (giving more) RPCs, it is generally accepted that vertebrate RPC lineages tend to vary in size by at least one order of magnitude, from less than ten to tens of neurons.

<sup>5</sup>Unfortunately, I am unable to locate the source of this convention, likely due to the poor preservation of many of these older reports. That this required no citation in 1954 suggests the original observations of CMZ proliferation may be in the early 20th century.

actively multiply, and, by subsequent differentiation, give rise to the diverse nervous and sensory constituents of the retina. [VL54]

[author's translation from the French]

Despite this, the proliferating RPCs in this “peripheral germinal zone” (also known as the “ciliary marginal zone”, or CMZ, for its proximity to the retinal ciliary body) have not received the same level of attention as those associated with the central retina. As a consequence, these RPCs have generally, and in particular by Harris, been treated as though they are a type of “frozen” progenitor population, recapitulating spatially, along the peripheral-central axis, the process which RPCs in the central retina undergo in a time-dependent fashion<sup>6</sup> [HP98].

The length of the RPC cell cycle has been of considerable interest, since the evolution of this parameter in time, in conjunction with the RPC population size (the number of cells specified in the eye field), determines the eventual size of differentiated retinal neural population, and therefore the retina. RPC cell cycle length has generally been inferred from clonal lineage size, although it has also been occasionally assayed directly in cumulative thymidine analogue labelling experiments. Vertebrate RPCs have generally been found to undergo a period of relative quiescence, in which the cell cycle lengthens, before the neural retina begins to be specified (in zebrafish, this period is 16-24 hpf). The cell cycle shortens as RPCs begin to exit the cell cycle [HH91, LHO<sup>+</sup>00]. After the central retina is specified, the RPC cell cycle once again lengthens, and is presumed to continue to slow until RPCs have completed differentiation<sup>7</sup>.

Finally, the orientation of the RPC division plane in mitosis has also been implicated in retinal organisation. The orientation of divisions has been associated both with the proliferative and differentiative fate of RPC progeny. For instance, interfering with spindle orientation in the developing rat retina, such that more RPC divisions occur parallel to the neuroepithelial plane (rather than along the apico-basal axis) results in more proliferative and fewer postmitotic, specified progeny[ŽCC<sup>+</sup>05]. That said, it seems that whatever effects are attributed to mitotic orientation are likely species-specific, as zebrafish RPCs display a different pattern of axis orientation, dividing mainly in the epithelial plane[DPCCH03].

### 1.3.2 Specificative phenomena

Irrespective of their location in the retina, vertebrate RPC lineages have offspring which may enter any of the three cellular layers of the retina. Moreover, single lineages can include any possible combination of cell fates, so that RPCs cannot be said to be of different “types” on the basis of lineage fate outcomes [HBEH88, TSC90, WF88]. While some specified progenitors have propensities to give rise to similar cell types, these relations appear to be species-specific, and do not seem to define separate progenitor pools [AR08]. In general, then, RPCs are taken to be totipotent with respect to the neural retina- all of the cell types<sup>8</sup> of the differentiated retina are derived from similar RPC lineages. Very little about this picture has changed since its initial development, using a variety of lineage tracers (including retroviruses, thymidine analogues, and injectable dyes) and histochemical markers to supplement morphological identification of specified neurons. In particular, sophisticated modern live imaging experiments in zebrafish

<sup>6</sup>Since both early central and later peripheral proliferation and specification are similarly, and non-trivially, organised both in space and time, this idea will be addressed in somewhat more detail below.

<sup>7</sup>This presumption is demonstrably incorrect in the /textit{D. rerio} eye, see Figure 2.7

<sup>8</sup>The “cell type” concept is unusually well-defined in the retina, as there are an abundance of distinct morphological and molecular features which differentiate numerous subtypes of the seven general types of retinal neuron.

(many pioneered by Harris), have broadly confirmed the findings of the 80s and 90s in mammalian fixed specimens, explants, and the like [BRD<sup>+</sup>15].

Of particular note here are the observations of the Raff group [WR90, CBR03], who demonstrated that dissociated rat RPCs, cultured at clonal density, took on morphological and histochemical features associated with different specified neural types in similar numbers and proportions, and on a similar schedule, to same-aged RPCs cultured in retinal explants. These results dramatically suggested that both the proliferative and specificative behaviour of RPCs depended less on intercellular contact, and the complex signalling environment of the developing retina, than on factors intrinsic to the RPCs themselves. These studies contain the essential germ of Harris' eventual commitment to SSM explanations, purporting as they do to “test the relative importance of cell-intrinsic mechanisms and extracellular signals in cell fate choice” and providing convincing evidence for the preponderant importance of the former.

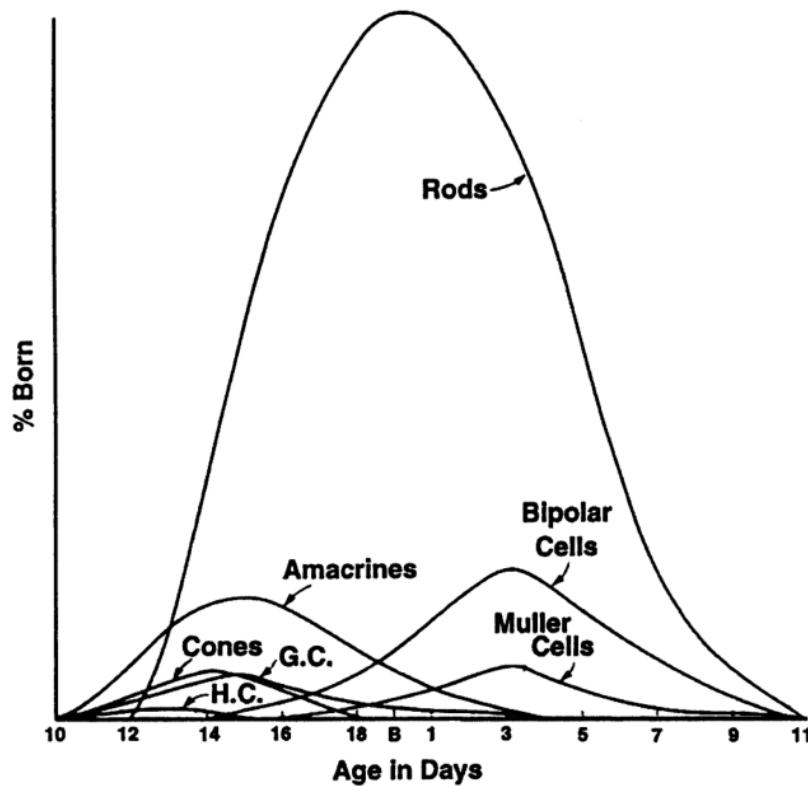


Figure 1.1: Histogenetic birth order of retinal neurons

Adapted from [You85] by [CAY<sup>+</sup>96]. G.C., Ganglion Cells; H.C., Horizontal Cells. Data from embryonic and perinatal mouse eyes.

In spite of the apparent ability of RPCs to produce offspring specified to any of the possible neural cell fates, at any time during their lineage history<sup>9</sup>, vertebrate retinal development displays a temporal ordering, such that in any particular location, retinal ganglion cells (RGCs) tend to be produced first,

<sup>9</sup>Even in papers arguing for a strict, linear sequence of specificative outcomes in all RPC lineages, the actual data show RPCs occasionally giving rise to “late-born” photoreceptors subsequent to their first division[WR09], giving lie to the notion that the process is particularly strict.

followed by the other cell types in what has been described as an overlapping “histogenetic order”, pictured in Figure 1.1. As noted above, RPCs also exit the cell cycle in a spatiotemporally defined order (from central to peripheral over time), and specification follows the same pattern<sup>10</sup>. This naturally gave rise to some question about the origin of the “overlap” observed in the sequential production of various cell types; conceptually, this overlap could be produced by identical RPCs executing identical rigid specifiable programs if they begin to execute it along the spatiotemporal gradient noted above, as originally suggested by Cepko et al. [CAY<sup>+</sup>96]. However, it is impossible to reconcile this notion with the recent results of Harris’ *in vivo* zebrafish lineage tracing studies [DPCH03, HZA<sup>+</sup>12, BRD<sup>+</sup>15], which confirm mammalian cell culture work in demonstrating that vertebrate RPC lineages do not execute identical (or even vaguely similar) specifiable programs.

### 1.3.3 Other morphogenetic phenomena

The outcomes of RPC proliferation and lineage commitment have generally been taken to be sufficient to explain the formation of the neural retina. Indeed, after the eye cup has been formed (prior to the specification of any neurons), RPCs are, in effect, already “in place”. Therefore, there are few documented RPC phenomena outside of the “proliferation” and “specification” categories of the morphogenetic alphabet, which enumerates the cellular processes contributing to tissue form and structure. RPCs do not seem inclined to migrate long distances, they seem not to generate much in the way of extracellular material<sup>11</sup>, and, in non-pathological conditions, they are rarely seen to die. Still, there are a number of phenomena which appear to be important for proper retinal organisation that fall into these other “alphabet” categories.

The most notable of these is interkinetic nuclear migration (INM), in which RPC nuclei move back and forth between the apical and basal surfaces of the retina. Common in many neural tissues, INM affects both proliferative and differentiative outcomes, but is dissociable from them- both cell cycle and differentiation proceed (albeit in a precocious manner) when INM is disrupted[MZLF02]. The environment provided by the apical retinal surface appears to be required for RPC mitosis to proceed<sup>12</sup>, and INM seems to consist of a directed, probably actomyosin-mediated movement to the apical surface, followed by an undirected “random walk” away from the apical surface and hence toward the basal retina [NYLH09]. This “random walk” may arise from displacements due to the directed INM of neighbouring progenitors [AHW<sup>+</sup>20]. More committed RPCs also appear to actively migrate to positions appropriate for the specified cell type [CAR<sup>+</sup>15, IKRN16], although it is unclear to what extent this short-range migration is related to INM undergone by more actively proliferating progenitors. It seems likely that these short-range migrations of more specified cells are especially significant in the early retina, before the neural plexiform layers (consisting of axons and other neural processes) have begun to divide the cellular layers into bounded compartments.

Notable lacunae in the study of the RPC morphogenetic alphabet include the regulation of cell size

---

<sup>10</sup>In many studies, cell cycle exit is conflated with specification such that evidence of the former is taken as evidence of the latter. There are, however, reasons to believe cell cycle exit and specification are not the same process, discussed below.

<sup>11</sup>The formation of a laminin-rich basement membrane seems to be necessary for optic vesicle formation [ICW13], but it is not clear that RPCs produce this. ECM function in eye formation remains under-studied.

<sup>12</sup>Nonapical divisions do occur, notably in specified, but proliferative cells [GWC<sup>+</sup>07]. The extent to which RPCs depend on the apical surface may depend on how “RPC” is defined. In general, RPCs are no longer considered as such when they acquire characters associated with differentiated neurons (cell type markers, morphological traits, etc.), although it is clear that the acquisition of these characters does not necessarily imply that the cell is postmitotic. Any complete explanation of how RPCs give rise to the structure of the eye must also consider these nonapical divisions.

and growth, and potential roles for cell death, both of which have hardly been studied at all. While cell size and growth are known to be tightly linked to proliferative behaviour in yeast [YDH<sup>+</sup>11], any related effects in RPCs have not been elucidated. This may be a significant oversight, given the requirement for RPCs to continuously grow during their proliferative lifespan. Cell death does not seem to play the same “pruning” role for RPCs that it often does in other neural tissues, and observed rates of cell death in normal RPC populations are very low, so few studies have been conducted.

## 1.4 Macromolecular mechanistic explanations for RPC phenomena

Having surveyed the cellular phenomena pertaining to RPC function in retinal development, let us examine a selection of the many macromolecular mechanistic explanations (MEx) that have been offered to explain aspects of RPC behaviour. Unsurprisingly, given the field’s focus on the early events of eye development, these MEx are intended mainly to explain the phenomena associated with this period. Thus, the majority of MEx offered have been concerned to explain tissue-level phenomena like the initial “wave” of cell cycle exit and specification, without necessarily seeking a global explanation for RPC behaviour irrespective of context, so that it is unknown how many of these might pertain to ongoing peripheral neurogenesis or other, adult neurogenic phenomena like those exhibited by Müller glia. That said, we now turn to examine some of the best-developed of these explanations.

### 1.4.1 Transcription factor networks

Perhaps the most notable MEx offered to explain RPC specification and development is the eye field transcription factor network, or EFTFN. The roots of this MEx are found in the Pax6 “master gene” explanation popularised in the 1990s [Geh96]. This explanation revolved around the observation of the apparently universal involvement of Pax6 gene products in eye formation in model organisms, and the promiscuous inter-species effects of Pax6 (with mouse RNA able to induce ectopic formation of eye structures in *Drosophila* imaginal discs, for instance [HCG95]), so that it appeared to be a highly conserved genetic “switch” for eye development.

Importantly, this explanation purported to resolve what Darwin regarded as a serious problem for his theory, the apparent implausibility of the gradual evolution of eyes (and other “organs of extreme perfection”) from some primitive ancestral structure [Dar88, p.143-4]. In particular, Pax6 suggested to some theorists an alternative to the surprising suggestion of Mayr and Salvini-Plawen, that differences in eye structure and function across clades suggested the independent appearance of eyes in more than 40 clades [vM77]. Pax6 was thus taken to provide a molecular pointer to a potential common ancestor for all animal eyes [EHML09].

Subsequent investigations revealed that vertebrate Pax6 is a conserved member of a complex network of cross-activating and inhibiting transcription factors, including Pax6, Rx1, Six3, Six6, Lhx2, ET, and Tll [Zub03]. Members of this network tend to promote proliferation and suppress markers of differentiated neurons, and their loss commonly results in the failure to form the eye field at all [AH09]. The expansion of this explanation to include other TFs in a network revealed significant differences between species [Wag07]- while the role of Pax6 seems to be conserved between *Drosophila* and vertebrates, the roles of other members of the EFTFN are not. Moreover, the universality of Pax6 was only apparent, and

not real, as there are bilaterian eyes whose development is Pax6 independent (including in *Platynereis*, *Branchiostoma*, and planarians) [Koz08]. This highlighted the great difficulty in connecting morphological characters such as those observed by Mayr with a genetic basis- it is simply not clear that Pax6 conservation points to a common ancestor for all eyes, or even all photoreceptive neurons<sup>13</sup>. Moreover, the expansion of the moncausal “master gene” explanation, to include a network of TFs with broad gene regulatory effects, clearly highlighted the problems of complexity in offering MEx for RPC function. The components of this network interact in complex, context-dependent ways, and while the EFTFN as a whole is taken to promote RPC proliferation and to delay specification<sup>14</sup>, its individual components have also been held responsible for the specification of particular classes of differentiated neurons, and remain expressed in those postmitotic cells. Notably, Pax6 is implicated in the expression of bHLH TFs required to specify multiple classes of retinal neuron [MAA<sup>+</sup>01], and is known to directly activate Ath5, necessary for RGC specification [WSP<sup>+</sup>09]. The EFTFN has thus been taken as an explanation for the maintenance of the multipotent, proliferative RPC state, but how this network is disassembled, and its components repurposed to promote specification, remains obscure.

The EFTFN is not the only transcription factor network offered as a MEx for RPC function. Another well-developed MEx involves Chx10 (aka vsx2), a transcription factor which was found to be important for normal proliferation of RPCs, its loss causing microphthalmia in the mouse [BNL<sup>+</sup>96]. Chx10 was subsequently found to repress Mitf, which is involved in RPE specification, and hence to promote neural retinal fates over pigmented epithelial ones [Hor04]; in the absence of Chx10 the early eye cup does not stratify properly between apical pigmented cells and the neural retina. Much like the multifunctional EFTFN components, Chx10/vsx2 has also been implicated in the specification of particular neural fates, notably bipolar neurons [BNL<sup>+</sup>96] and the regulation of Vsx1 (a parologue of Chx10), Foxn4, and Ath5, associated with specification of subpopulations of bipolar cells, horizontal and amacrine cells, and RGCs and PRs, respectively [CYV<sup>+</sup>08, VJM<sup>+</sup>09].

Clear hypotheses advocating for particular relationships between different TF MEx are rarely stated. It is tempting simply to arrange them in some kind of “developmental order”, perhaps with the Chx10-Mitf network “downstream” of the EFTFN. That this would be facile is evident from the changing roles of these transcription factors depending on developmental and cellular context. To date, no unifying framework has been applied. Obvious candidates include the “developmental gene regulatory network” concept [LD09], a type of cybernetic explanation which assembles genes into feedback networks. Given the popularity of this type of explanation, it is worth noting that no one has yet had any success in offering one for RPC function.

These transcription factor network MEx frequently incorporate extracellular signals (often as an explanation for the appearance or “set-up” of the TF network), and it is generally recognised that these signals have a profound influence on these networks, and on RPC behaviour generally. We therefore turn to explanations invoking these signalling mechanisms.

---

<sup>13</sup>Indeed, the relevant “unit” of homology for evolutionary explanations for eyes remains contested, with some arguing for the cell itself over any particular set of gene sequences [EHML09]

<sup>14</sup>This is sometimes referred to as “promoting RPC fate”, since RPCs are taken to be those cells which proliferate but do not yet display markers of specification. Since it is, by now, widely recognised that cells that appear to be well-specified may remain in cell cycle [GWC<sup>+</sup>07, ESY<sup>+</sup>17] this terminology should probably be jettisoned.

### 1.4.2 Intercellular signalling networks

Virtually every developmentally significant class of signal has been implicated in RPC function, so that Harris, by 2009, simply glosses over a majority of these pathways by briefly summarising them in tabular form and not otherwise mentioning them [AH09]. These include BMP, CNTF, FGF, Glucagon, Hedgehog, IGF, Notch, TGF $\alpha$ , TGF $\beta$ , VEGF, wnt, and a host of neurotransmitters. This diversity of signalling pathways has proved to be a formidable problem for integrated explanations, since almost all of these pathways converge on the same two cellular outcomes in RPCs, that is, proliferation and specification. Thus, most signalling MEx for RPC function elide the majority of other signals which are known, or thought, to affect the same processes. That said, let us explore a few of the more detailed signalling explanations.

In developmental terms, the first phenomenon requiring explanation is the appearance of the eye field to begin with- what is it that accounts for the differentiation of RPCs from the rest of the anterior neural plate and tube? Wnt signalling MEx have been offered to explain the appearance of the *Xenopus* eye field. Fz3 signalling seems to promote expression of eye field transcription factors (see below)[RDT $^{+}01$ ], while an unspecified non-canonical interaction between Wnt11 and Fz5 inhibits canonical  $\beta$ -catenin signalling through Wnt8b/Fz8a, which would otherwise promote prospective anterior forebrain fates [CCY $^{+}05$ ]. Inhibition of FGFR2 signalling, and activation of ephrinB1 signalling have also been implicated in early *Xenopus* eye field cell movements [MMDM04]. Subsequent experiments determined that the xenopus ADP signalling through the P2Y1 receptor directly activates the eye field transcription factor network (EFTFN), another well developed MEx described below [MBE $^{+}07$ ]. More recent experiments in zebrafish suggest that precocious acquisition of neuroepithelial apicobasal polarity, probably driven by interactions with a Laminin1 basement membrane, distinguishes the early eye field [ICW13].

Developmentally subsequent to the appearance of eye field RPCs and their rearrangement into the optic cup, the apparent central-to-peripheral “wave” of RPC exit from cell cycle and specification of early RGCs [HE99], has had detailed MEx advanced to explain it. In both zebrafish and chick retina, FGF3 and FGF8, originating from the optic stalk, initiate this early cell cycle exit and specification [MDBN $^{+}05$ ], while inhibiting FGF signalling prevents this from occurring, and ectopic expression of FGF can cause it to occur inappropriately. The progression of this “wave” of cell cycle exit and specification has been separately explained, by Sonic Hedgehog (Shh) signalling from the newly specified RPCs inducing cell cycle exit and specification in adjacent cells [Neu00]. This process is dependent on, and downstream of, the above-mentioned FGF induction [MDBN $^{+}05$ ]. The role of Hh signalling has been challenged on the basis that Hh inhibition in subsequent experiments did not display the same effect size [SF03], and that its effects on Ath5 expression, required for RGC specification, are ambiguous [AH09]. More recent MEx advanced by Harris have suggested that Hh signals may decrease the length of the cell cycle, resulting in increased proliferation and earlier cell cycle exit and specification [LAA $^{+}06$ , ALHP07].

A well-developed “local” signalling MEx (mainly advanced by Harris) invokes the classic Notch/Delta lateral inhibition model, with small fluctuations in Notch/Delta activity giving rise to a positive feedback response that differentiates neighbouring cells. Cells which have high Delta expression tend to be specified as retinal neurons, while those with high Notch tend to remain proliferative, either as RPCs or Müller glia [DRH95, DCRH97]. Such a mechanism could regulate the activity of both early, central RPCs, as well as peripheral RPCs, and may contribute to inter-RPC variability. These differences between RPCs located in different parts of the developing retina have been of significant interest, and it is worth briefly examining patterning MEx that may also explain spatial differentiation between RPCs.

### 1.4.3 Patterning mechanisms

Among the most interesting features of RPCs is that they reliably give rise to specified neurons, in particular RGCs, that seem to “know where they are” in the retina, enabling them to wire their axons in correct retinotopic order in the superior colliculus (SC) or optic tectum (OT). The most robust MEx explaining this refer to gradients of EphA and EphB receptors expressed in RGCs, and their respective ephrin ligands expressed in the SC or OT. In the retinal RGC population, an increasing nasotemporal gradient of EphA is paired with an increasing dorsoventral gradient of EphB. A corresponding increasing rostrocaudal gradient of ephrin-A is paired with an increasing lateromedial gradient of ephrin B in the SC/OT. This allows for a two-axis encoding of an RGCs’ position in the retina [TK06]. As the RGCs’ axon pathfinding depends on repulsive effects mediated by Eph receptors, this code is sufficient to allow correct wiring of even single RGCs [GNB08]. This code seems to be established, in part by the action of Gdf6a, in RPCs themselves, prior to specification [FEF<sup>+</sup>09].

Indeed, there are numerous similar observations of expression gradients that create spatial differences between RPCs themselves. Most relevant to the proliferation dynamics highlighted in this chapter is the observation that, in *Xenopus* eyes, a decreasing dorsoventral gradient of type III deiodinase renders the cells of the dorsal CMZ refractory to thyroid hormone (as the deiodinase inactivates TH) [MHR<sup>+</sup>99]. The effect of this is to set up a differential response to TH in post-metamorphic RPCs, so that the ventral population selectively expands in response to TH [BJ79].

These patterning mechanisms are of particular interest here, in large part because they clearly establish that the RPC population is heterogenous, both with respect to proliferative and specificative behaviours, and perhaps others as well. This is of critical importance for any modelling effort, as virtually all mathematical models used by stem cell biologists (and those used to justify Harris’ SMME) assume, at least initially, homogenous populations of stem or progenitor cells. Since RPCs do not meet this condition, special care is needed to use these models.

### 1.4.4 Chromatin dynamics

In recent years, the great importance of chromatin conformation in RPC proliferation and specification has become more clear. Indeed, chromatin dynamics are now widely invoked in explaining stem and progenitor cell behaviour, and suggested as a target for cell reprogramming [Kon06, TR14]. In RPCs, detailed accounts of three-dimensional chromatin dynamics have yet to appear. However, a number of studies point to the importance of chromatin state in informing the overall cellular state. In particular, histone deacetylation seems to be important for RPC specification, as the loss of histone deacetylase 1 (HDAC1) in zebrafish results in overproliferation and decreased specification, and correlated increases in Wnt and Notch activity [Yam05]. In mouse retinal explants, pharmacological inhibition of HDAC results in decreased proliferation and specification [CC07]. Additionally, the chromatin remodelling complex SWI/SNF has repeatedly been implicated in RPC function. Notably, one particular component of this complex seems to be particularly associated with vertebrate RPCs (BAF60c, an accessory subunit) [LHKR08]. A switch to other subunits seems to be necessary for specification [LWR<sup>+</sup>07]. Details regarding the subunits involved in specification and their downstream effects are complex and context dependent, much like the signalling pathways mentioned above.

## 1.5 A unified theory of RPC function? “Blurring” to order

From the foregoing discussion, we can clearly see Harris’ theoretical conundrum. Macromolecular explanations for RPC behaviours, like those throughout the molecular biological tradition, have generally been built outwards from particular transcription factors, receptors, etc. The result is an archipelago of MEx, at best connected by tenuous speculation, and in most cases, without any known means to form an integrated model. Furthermore, the degree of complexity and context-dependence evident from the literature might seem to preclude such a model. As we have seen, Harris found that the evidence did not allow for clear discrimination between logically distinct types of mechanisms for producing the observed variability in RPC lineage outcomes.

In this situation, Harris effectively had two theoretical options. The first is simply to “crank the handle”- to generate more and more facts describing the difference particular molecules make to RPC outcomes in dozens of relevant contexts, piling up exceptions and idiosyncrasies, in the hope that doing so will eventually bridge the explanatory “islands” of the MEx archipelago. I have referred to this as the Enumerationist approach and explained why it has failed, and continues to fail, in the Theoretical Appendix.

The second option, the one actually chosen by Harris, is more theoretically sophisticated. As Nicholas Rescher has noted regarding the *in-principle limits to scientific knowledge*, the phenomenal universe has infinite descriptive complexity- one can always add more detail to a description of some phenomenon, and no such description is ever complete [Res00, p.22-9]. Moreover, “even as the introduction of greater detail can dissolve order, so the neglect of detail can generate it.” [Res00, p.62] As Rescher goes on to comment:

[W]e realize that in making the shift to greater detail we may well lose information that was, in its own way, adequate enough ... information at the grosser level may well be lost when we shift to the more sophisticated level of greater fine-grained detail. The ‘advance’ achieved in the wake of ‘superior’ knowledge can be - and often is - purchased only at a substantial cognitive loss.

...

It is tempting on first thought to accept the idea that we secure more - and indeed more useful and more reliable - information by examining matters in greater precision and detail. And this is often so. But the reality is that this is not necessarily the case. It is entirely possible that the sort of information we need or want is available at our ‘natural’ level of operation but comes to be dissolved in the wake of greater sophistication. [Res00, p.65-6]

Rescher’s greater point is that “blurring” detail, at levels below the phenomenal one under consideration (for RPCs, generally, the cell or lineage), may actually be necessary to produce an ordered explanation that is useful with respect to some objective. Given the number of apparently contradictory MEx for RPC behaviour, we have exactly the kind of situation Rescher is describing- more sophistication, and more detail, has dissolved order, not revealed it<sup>15</sup>. The inability to assemble an unified explanation for RPC function has left us without a good fundamental understanding of how highly ordered neural tissues like eyes can be generated from composites of units with highly variable, temporally and spatially ordered outcomes like RPC lineages. As this is a common feature of vertebrate neurogenesis

---

<sup>15</sup> At least part of this problem is likely related to the fact that the majority of biomedical findings cannot be replicated [Ioa05]. The finding, mentioned above, that Shh effect sizes on RPC function were not as large as initially reported when subsequently investigated, is typical and symptomatic of this replication problem. “Blurring” may therefore be necessary not only because of fundamental epistemic limits, but also because it is often difficult to distinguish bona fide results and explanations from spurious ones.

more generally, the theoretical problem here leaves us without the ability to produce complete models of neurodevelopmental processes in many species. Moreover, in the absence of a clear framework for comparing the explanatory power of the diverse array of MEx so far advanced, practical contributions of clinical relevance have been scanty and tentative, with RPC transplantation, and more recently, gene therapies taking little note of complex MEx for RPC function[CAI<sup>+</sup>04, GS07, YQW<sup>+</sup>18].

In a situation of this kind, it may well be that this type of “blurring” is required, and it seems that is what Harris is attempting in advancing his SMME. Harris is asking, in some sense, whether most of the MEx offered for RPC behaviour are extraneous to an adequate understanding of how RPCs work. By cutting down to the simplest possible explanation, Harris hopes to bring into view order that was previously obscured by detail. There is, of course, a significant danger here: how does one decide what is “blurred out” and what remains? We can easily understand how a practitioner’s biases could lead to a sort of relativism, where the “blurring” makes apparent a spurious order that conforms to these biases rather than to reality as such. With this in mind, let us survey the general thrust of Harris’ SMME, before proceeding to examine it in detail, in Chapter 2.

## 1.6 Explanatory Strategy and Intent of the SMME

As we have seen in Section 1.2, Harris’ long-held understanding of the explanatory options for RPC function divides them into four broad categories, which I summarise as follows:

1. A linear algorithmic “program” of proliferation and specification
2. Asymmetric segregation of specificative determinants
3. “Stochastic processes” internal to the cells
4. Influences of extracellular factors

Harris’ sophisticated discussions of RPC MEx rarely treat these categories as exclusive, and concede that good explanations for RPC behaviour may involve phenomena from more than one of them. Indeed, the SMME necessarily contains elements that Harris concedes are “linear” and “deterministic” [HZA<sup>+</sup>12]. Still, his overall strategy for the SMME is, first, to substantiate the predominant influence of one of these categories of phenomena (that is, category 3, internal stochastic processes or effects), and subsequently to specify an actual macromolecular system that could plausibly be such an “internal stochastic process”. These two theoretical maneuvers, while tightly linked, serve different purposes within Harris’ overall explanatory framework, which must be examined separately.

The SMME for zebrafish RPC function has been developed across three separate papers [HZA<sup>+</sup>12, BRD<sup>+</sup>15, WAR<sup>+</sup>16]. Each builds on the earlier publications, collectively purporting to explain the behaviour of RPCs wherever, and whenever, they may be found in the zebrafish eye. The underlying model is originally derived from an earlier paper pertaining to rat RPCs [GZC<sup>+</sup>11]. He et al. [HZA<sup>+</sup>12] and Wan et al. [WAR<sup>+</sup>16] use essentially the same model and make up the substance of the first of these two maneuvers. Boije et al. [BRD<sup>+</sup>15] substantially modifies this model, specifying the activity of two known transcription factors (Ath5 and Ptf1a) as the model’s biological referents. This paper constitutes the second theoretical thrust.

The first maneuver intends to provide support for the contention that zebrafish RPCs are a group of equipotent cells which give rise to variable outcomes that depend on independent “stochastic” processes

within each of these cells. This support is to be provided by demonstrating that a suitably configured **Simple Stochastic Model** (SSM) of an RPC, numerically simulated many times by **Monte Carlo** methods to represent a population of RPCs, produces similar outcomes to populations of RPCs *in vivo*. This explanatory strategy is common in the stem cell literature, the original example having been published in 1964 by Till, McCulloch, and Siminovitch[TMS64]. The SMME therefore represents an example of a traditional scientific logic- an explanatory pattern deployed by stem cell biologists in diverse contexts, and widely accepted because of its ongoing use in the literature, although with varying interpretations.

The success of this first maneuver thus depends on two outcomes. Firstly, the output of the SSM should accurately reflect the observed proliferative and specificative outcomes of zebrafish RPC lineages, giving weight to Harris' claim that it "provides a complete quantitative description of the generation of a CNS structure in a vertebrate *in vivo*" [HZA<sup>+</sup>12]<sup>16</sup> Secondly, the internal structure of the SSM should provide good reason to believe that one of the Noise explanations is a better explanatory option for RPC lineage outcomes than those identified by the other three categories of theoretical options ennumerated above.

The second theoretical maneuver is the specification of particular biological referents for entities in the model. This offers an opportunity to move beyond a purely conceptual argument about the kind of process that might produce variable RPC lineage outcomes, and to begin the work of explaining how the behaviour of a particular macromolecular system constitutes such a process, so that empirically verifiable hypotheses may be generated. The success of this maneuver depends on the biological plausibility of the identification between model structure and the biological function of transcription factors, *Ath5* and *Ptf1a*, that the model names. A good SSM-based explanation would point the way for further research by identifying *how* so-called "stochastic processes" in RPCs might function, and make some predictions about this. With that said, let us turn to the SMME and determine how well these manuevers have succeeded.

---

<sup>16</sup>This claim is somewhat extravagant; the SSM, by definition, includes no spatial information, so it is unclear how one could be a "complete description" of any spatially organised structure. Still, it can be complete with regard to cell population numbers.

# Chapter 2

## “Stochastic mitotic mode” models do not explain zebrafish retinal progenitor lineage outcomes

*The text of this chapter has been adapted from a manuscript originally prepared for submission to PLOS One and is formatted accordingly; the methods and supplementary materials are available in ??*

### 2.1 Introduction

Mechanistic explanations (MEx) derive their utility from the resemblance of the conceptual mechanism’s output to empirically observed outcomes. As maps to biological territories, biological MEx are naturally identified with the living systems they represent. Most well-developed MEx take the form of a model, whose internal structure is taken to reflect the underlying causal structure of a biological phenomenon. The nature of the causal relationship between a mechanistic model and the phenomenon it purports to explain remains a topic of active dispute in the philosophy of biology[Fag15b]. Biologists, nonetheless, usually accept that a model which explains empirical observations well (usually measured by statistical or information theoretic methods), and reliably predicts the results of interventions, bears a meaningful structural resemblance to the actual causal process giving rise to the modelled phenomenon.

Biological phenomena are notable for exhibiting both complex order and unpredictable variability. A significant challenge for MEx in multicellular systems is to explain how complex, highly ordered tissues, like those produced by neural progenitors, can arise from cellular behaviours with unpredictably variable outcomes. Stem cell biologists have traditionally resorted to Simple Stochastic Models (SSMs) in order to explain the observed unpredictable variability in clonal outcomes of putative stem cells[TMS64, Fag13]. Because SSMs are susceptible to Monte Carlo numerical analysis as Galton-Watson branching processes, they have been convenient explanatory devices, appearing in the literature for more than half a century. By specifying the probability distributions of symmetric proliferative (PP), symmetric postmitotic (DD), and asymmetric proliferative/postmitotic (PD) mitotic modes, SSMs allow cell lineage outcomes to be simulated.

SSMs are “stochastic” insofar as they incorporate random variables with defined probability distribu-

tions. As Jaynes has noted, “[b]elief … that the property of being ‘stochastic’ rather than ‘deterministic’ is a real physical property of a process, that exists independently of human information, is [an] example of the mind projection fallacy: attributing one’s own ignorance to Nature instead.”[JBE03] Despite this, macromolecular processes are often described as “stochastic” in the stem cell literature. Generally speaking, the behaviour of an SSM’s random variable is taken to represent sequences of outcomes that are produced by multiple, causally independent events. Recently, the influence of “transcriptional noise” on progenitor specification has been identified as a candidate macromolecular process that may produce unpredictable variability in cellular fate specification. Therefore, one explanatory strategy for stem and progenitor cell function compares SSM model output to observed lineage outcomes, in order to argue that “noisy”, causally independent events give rise to the proliferative and speciative outcomes of progenitor lineages.

In this report, we evaluate one of the best-developed of these explanations, proffered by William Harris’ retinal biology group. We have dubbed this the ”Stochastic Mitotic Mode Explanation” (SMME) for zebrafish retinal progenitor cell (RPC) function. The SMME is noteworthy because it claims: (1) to ”provid[e] a complete quantitative description of the generation of a CNS structure in a vertebrate in vivo”[HZA<sup>+</sup>12]; (2) to have established the functional equivalency of embryonic RPCs and their descendants in the postembryonic circumferential marginal zone (CMZ)[WAR<sup>+</sup>16]; and (3) to have established the involvement of causally independent transcription factor signals in the production of unpredictably variable RPC lineage outcomes[BRD<sup>+</sup>15]. Moreover, the SMME is taken to explain histogenetic ordering (the specification of some retinal neural types before others, notably the early appearance of retinal ganglion cells (RGCs)) in vertebrates, without reference to the classical explanation of a temporal succession of competency states[TR86]. These would be significant achievements with important consequences for both our fundamental understanding of CNS tissue morphogenesis and for retinal regenerative medicine. However, these models were not subjected to typical model selection procedures, nor has their output been examined using modern information theoretic methods, as advocated by mainstream model selection theorists[BA02].

In order to facilitate the critical evaluation of the two SSMs which form the SMME’s MEx for RPC function, we have re-expressed the models as cellular agent simulations conducted using the open source C++-based CHASTE cell simulation framework[MAB<sup>+</sup>13]. We explored the structure of the SSMs, dubbed the He and Boije SSM respectively, compared to their explanatory forebear, dubbed the Gomes SSM[GZC<sup>+</sup>11]. This analysis strongly suggested that the introduction of an unexplained progression of temporal “phases” was largely responsible for the SMME model fits to data. In order to investigate this possibility, we built an alternative model with a deterministic mitotic mode and compared its output to the He SSM. We found that the deterministic alternative model was a better explanation for the observations, demonstrating that SMME fails when compared to alternatives. We therefore suggest that an explanatory approach based on the use of SSMs is incapable of distinguishing between theoretical alternatives for the causal structure of RPC lineage behaviours. Furthermore, by comparing the output of the models with novel postembryonic measurements of proliferative activity, we find that the SMME explanation cannot account for quantitative majority of retinal growth in the zebrafish, driven by the CMZ. Finally, we discuss the place of SSMs, the concept of “mitotic mode”, and the role of “noise” in explaining RPC behaviour, and suggest ways to avoid the modelling pitfalls exemplified by the SMME.

## 2.2 Results and Discussion

The SMME for zebrafish RPC function has been advanced using two SSMs. One first appears in He et al., and again, unmodified, in Wan et al.[HZA<sup>+</sup>12, WAR<sup>+</sup>16]; it is concerned primarily to explain lineage population statistics and time-dependent rates of the three generically construed mitotic modes (that is, PP, PD, DD). We have called this the He SSM. The second appears in Boije et al.[BRD<sup>+</sup>15]; its intent is both to introduce the role of specified macromolecules into the mitotic mode process, and to explain neuronal fate specification in terms of the process. We have called this the Boije SSM. The He model is directly descended an SSM advanced to investigate causally independent fate specification in late embryonic rat RPCs, formulated in Gomes et al.[GZC<sup>+</sup>11]. The Boije SSM differs substantially from the He and Gomes models, but inherits its general structure from the He SSM.

The metascientific analysis of biological explanations developing in time remains understudied. Perhaps most refined tool for global evaluation of biological theories (Schaffner's "Extended Theories"), treats biological explanations as hierarchically organised logical structures, after the fashion of Imre Lakatos, and proposes the use of Bayesian logic to distinguish between them[Sch93]. However, biologists rarely offer explanations in this form; we rather prefer MEx, expressed in diagrammatic form or as mathematical model-objects.

In this report, we accept Fagan's view that MEx for the behaviour of stem and progenitor cells consist of assemblages ("mechanisms") of explanatory components which are understood to be causally organised by virtue of their intermeshing properties[Fag15b]. While Fagan treats SSMs separately from macromolecular MEx, and we find that the Gomes SSM was not deployed in this role, the He and Boije SSMs were used as explanations for the behaviour of RPCs. As Feyerabend famously observed, scientists often operate as epistemological anarchists; the development of our explanatory logic is not bound by an identifiable set of rules, but rather arises organically from our scientific objectives, extrascientific context, and so on[Fey93].

We have therefore chosen to examine the structure of the Gomes, He, and Boije SSMs arranged in chronological order, to highlight how the explanatory logic of zebrafish SMME SSMs differs from their immediate ancestor, and from the traditional uses of SSMs in stem cell biology. We have diagrammatically presented these SSMs as MEx, consisting of components describing the proliferative and fate specification behaviour of cellular agents. The proliferative and specificative components of the MEx are causally organised by their Faganian intermeshing property, mitotic mode. We have used abbreviations to denote important classes of model components and inputs, informed by the emphasis of Feyerabend on the persuasive role of metaphysical ingredients and auxiliary scientific material; these are as follows:

- MI - Model ingredient, making reference to some conceptual or metaphysical construct
- AS - Auxiliary scientific content
- RV - Random variable
- PM - Parameter measurement, model parameter set by measurements, independently of model output considerations
- PF - Parameter fit, model parameter set without reference to measurement, in order to produce model output agreement with observations

We draw the reader’s attention to the expansion of the “mitotic mode” intermeshing property in later SSMs; this explanatory component comes to dominate the later models, which makes its interpretation critical to the success of the mechanistic explanation in meaningfully representing a real biological process.

Numerous theoretical options to explain observed variability in RPC lineage outcomes have historically been considered. Harris has previously argued that these belong to three cell-autonomous categories of process, in addition to extracellular influences: (1) a linear temporal progression of competency states, (2) asymmetric segregation of determinants during mitoses, and (3) intracellular “stochastic events”[HBEH88, AH09]. All of the SSMs are used to argue for the predominant influence of the third type of process in RPC lineage outcomes, essentially by demonstrating that the output of the SSM resembles observations.

### 2.2.1 Gomes SSM: Ancestral Model of the SMME SSMs

The Gomes SSM is presented in Fig 2.1. The model’s structure is straightforward; there are three independent random variables, drawing from empirically-derived probability distributions for the time each cell takes to divide, the mitotic mode of the division, and the specified neural fate of any postmitotic progeny. The random mitotic mode variable functions as the “intermeshing property” linking cycle behaviour to fate specification. The model thus represents a scenario where the processes governing proliferation, leading to cell cycle exit, and governing cell fate commitment are, at every stage, causally independent of each other and of their foregoing history of outcomes. The objective of the Gomes et al. study is to compare the lineage outcomes of dozens of individual E20 rat RPCs in clonal-density dissociated culture with the model output, developing earlier work in this system[CBR03]. Although the model incorporates conventional proper time (clock-time), none of the RVs reference it to determine their values. The abstract “cells” represented by this model do not have any timer or any source of information about their relative lineage position. Fate specification is construed in terms of conventional histochemical markers of stable cell fates; only the neural types generated late in the retinal histogenetic order are represented, as these are the only neurons specified by E20 rat RPCs.

This SSM is explicitly built as a null hypothesis, or a model of background noise. It represents an extreme case in which all of the specificative and proliferative behaviours of an RPC are totally independent of all other RPCs and events in its clonal lineage. The stated purpose of this model is “to calibrate the data”, serving “as a benchmark”[GZC<sup>+</sup>11], a purely hypothetical apparatus to produce sequences of causally unrelated lineage outcomes. Substantial deviations of the observed data from the fully independent events of the SSM may then be interpreted as causal dependencies between RPC outcome and their relative lineage relationships, as might be observed in a developmental “program” or algorithmic process. Finding that, with some exceptions, observed proliferative and specificative outcomes generally fall within the plausible range of Gomes SSM output, Gomes et al. conclude that RPC lineage outcomes in late embryonic rat RPCs seem to be dominated by causally independent events, suggesting that causally isolated “noisy” macromolecular processes may give rise to the unpredictable variability in the behaviour of these cells.

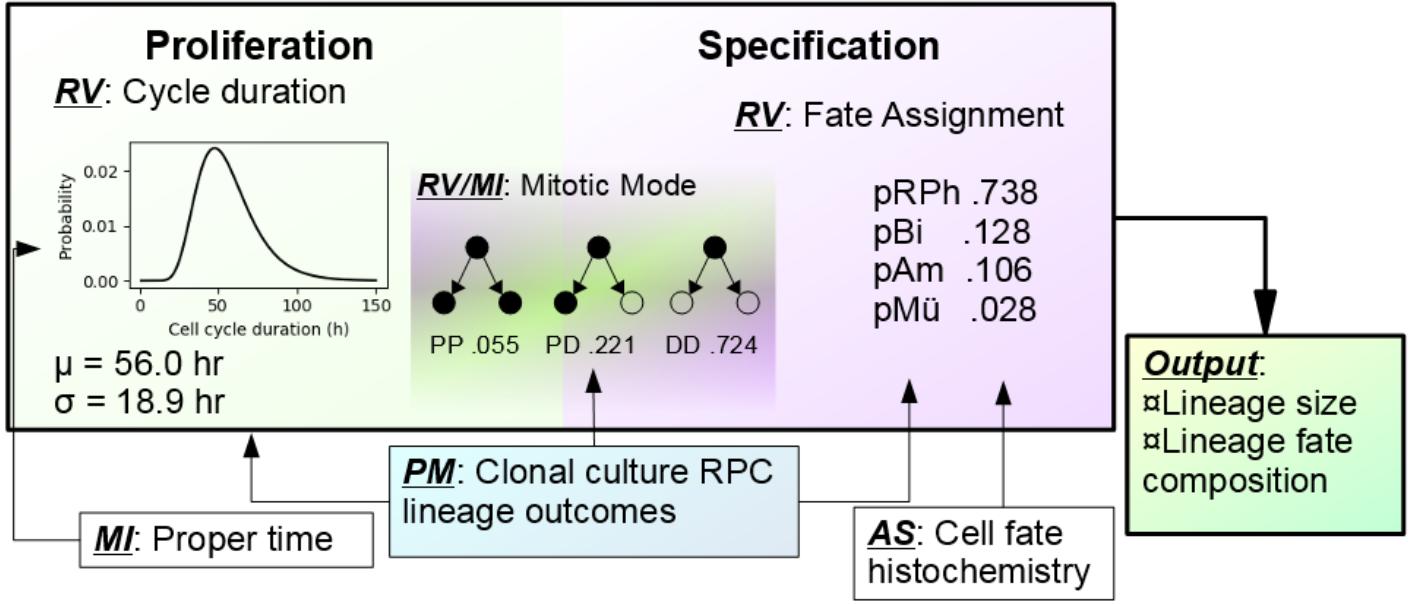


Figure 2.1: Structure of Gomes SSM

Structure of the Gomes SSM. pRPh, probability of rod photoreceptor specification. pBi, probability of bipolar cell specification. pAm, probability of amacrine cell specification. pMü, probability of Müller glia specification.

### 2.2.2 He SSM: Explaining variability in zebrafish neural retina lineage size

The He SSM, shown in Fig 2.2, is deployed in a explanatory role rhetorically identical to the Gomes SSM. The extent to which model output “captures.. aspects of the data” is taken to obviate the need for explicit “causative hypothes[es]”. On this account, only residual error between model output and observations may be ascribed to non-stochastic processes like “histogene[tic ordering] of cell types or a signature of early fate specification”. That is, if the model can be fit to observations, this is taken to exclude the presence of any cell-autonomous temporal program, asymmetric segregation of fate determinants, extracellular influences, and the like.

There are fewer direct empirical inputs to the He model’s parameters than the Gomes SSM, limited to the duration of the cell cycle. The close synchrony of zebrafish RPC divisions is modelled by assigning sister RPCs the same cycle length, shifted by a normal distribution of one hour width, in contrast to Gomes RPCs, which are treated as fully independent. The lineage outcomes the He SSM is called on to explain differ significantly from those referred to by the Gomes SSM. Most notably, the Gomes SSM does not account for the early appearance of RGCs, which are not produced by the late E20 progenitors examined in that study. The zebrafish RPC lineages studied by He et al. produce all of the retinal neural types, including RGCs, which are typically produced by PD-type divisions. The He SSM does not model particular cell fates, supposing that mitotic mode is “decoupled” from fate specification. The mitotic mode RV linking cell cycle to fate specification in the Gomes SSM has thus subsumed the specification outcomes of RPC lineages entirely, and the model is concerned only to explain the sizes of lineages (marked by an inducible genetic marker at various times), and the observed progression of mitotic modes in these early RPCs.

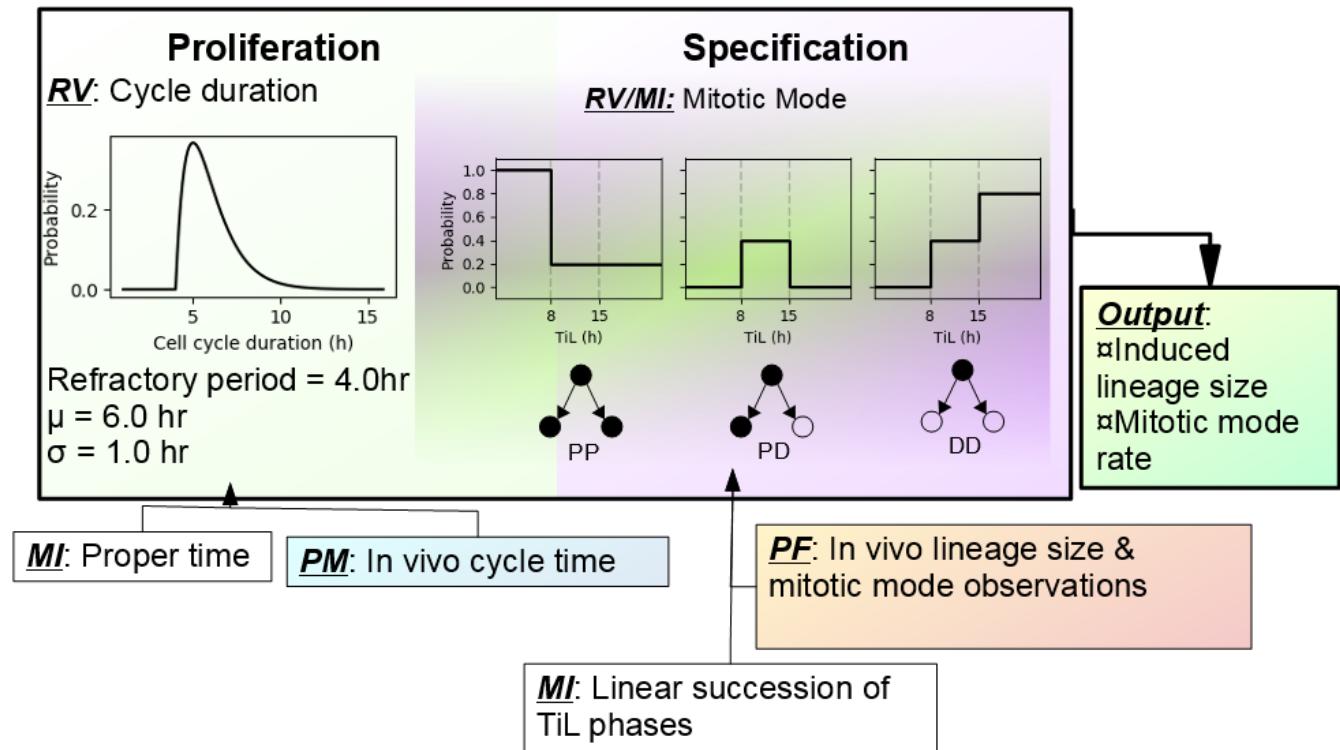


Figure 2.2: **Structure of He SSM**  
Structure of the He SSM. TiL, Time in Lineage.

The He model is, therefore, called on to explain the temporal structure of the proliferative and speciative behaviour of early zebrafish RPCs that dissociated late rat RPCs do not exhibit. A Gomes-type SSM, in which the RVs determining RPC behaviour are independent of any measure of time, cannot account for this temporal progression. In order to address this, He et al., assume a linear progression of three phases which cells in each lineage pass through, the timing of these phases being determined relative to the first division of the RPC lineage, called here “Time in Lineage” or TiL. The values the mitotic mode RV may take on is determined by these TiL phases. The temporal structure of the phases and their effect on the mitotic mode RV are selected to produce a model fit. While He et al. acknowledge that the model therefore represents a “combination of stochastic and programmatic decisions taken by a population of equipotent RPCs,” no effort is made to determine the relative contribution of the model’s stochastic vs. linear programmatic elements. Instead, the purportedly stochastic nature of mitotic mode

determination is emphasized throughout the report.

### 2.2.3 Boije SSM: Explaining variability in zebrafish RPC fate outcomes

The second SMME model advanced to explain zebrafish RPC behaviour, the Boije SSM, is displayed in Fig 2.3. This model is primarily concerned with the lineage fate outcomes that the He SSM does not treat, while abandoning the explicit proper time of the He and Gomes SSMs in favour of abstract generation-counting. The mitotic mode model-ingredient now subsumes all RPC behaviours. Boije et al. make a laudable effort to specify the particular macromolecules ostensibly involved in determining mitotic mode, nominating the transcription factors (TFs) Atoh7, known to be involved in RGC specification, and Ptfla, known to be involved in the specification of amacrine and horizontal cells, as primary candidates. The contribution of vsx2 is taken to determine the balance between PP and DD divisions late in the lineage, with the latter resulting in the specification of bipolar or photoreceptor cells. The binary presence or absence of these signals is determined by independent RVs structured by the phase structure present in the He SSM, translated into generational time from proper time.

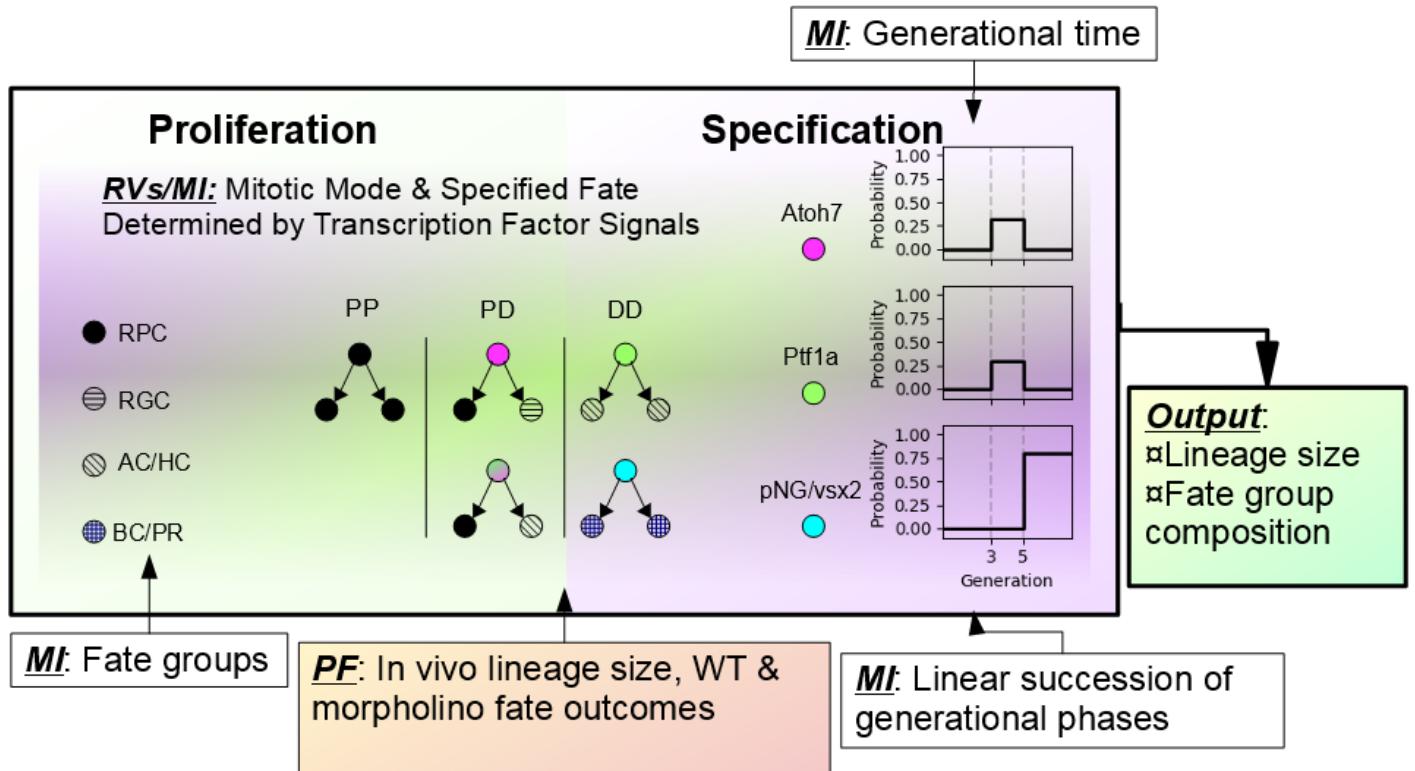


Figure 2.3: Structure of Boije SSM

Structure of the Boije SSM. RPC, retinal progenitor cell. RGC, retinal ganglion cell. AC, amacrine cell. HC, horizontal cell. BC, bipolar cell. PR, photoreceptor. pNG- parameter representing contributions of Vsx1 and Vsx2 to specification of BC & PR fates in absence of Atoh7 or Ptfla signals.

By this point, the SMME has become a very different type of explanation from its Gomes SSM forebear. We are no longer dealing with RVs that model causally and temporally independent processes

for different aspects of RPC behaviour. There is, rather, one temporally dependent process, the determination of mitotic mode, which is explained by unpredictable subsets of each lineage generation expressing particular TF signals. Where the Gomes SSM takes its parameters directly from empirical measurements of lineage outcomes, asking whether it is sufficient to assume that these are independently determined, the Boije SSM’s parameters are derived solely from model fit considerations. In spite of these considerable differences, the explanatory role of the SSM is effectively the same: the model’s fit to observations is taken as evidence of the predominant influence of “stochastic processes” determining mitotic mode on RPC behaviour. While Boije et al. acknowledge that whether some process is called “deterministic” or “stochastic” is “a matter of the level of description”[BRD<sup>+</sup>15] (i.e. is a property of the model-description and not of the physical process), the explanatory role of stochasticity for RPC lineage outcomes is, once again, emphasized throughout.

#### 2.2.4 Model selection demonstrates the SMME is not the best available explanation for RPC lineage outcomes

From a modeller’s perspective, it is notable that neither of the reports which use the He SSM[HZA<sup>+</sup>12, WAR<sup>+</sup>16], nor that using the Boije SSM[BRD<sup>+</sup>15] report in any detail their fitting procedures, nor do any of the above report any statistical measures of goodness-of-fit. Additionally, no models representing the alternative “theoretical options” available to explain variability in RPC lineage outcomes are compared to those advanced as evidence for “stochastic processes”. Given that the He SSM and Boije SSM depart from the Gomes SSM by the addition of an unexplained temporal structure to the mitotic mode model ingredient, it is striking that the overall argument remains similar to Gomes et al.’s, despite the persuasive force of the latter deriving from the lack of such structures. While He et al. and Boije et al. acknowledge that their models involve both stochastic and linear programmatic elements, no effort is made to quantify their relative influence on model output, and macromolecular explanation is only applied to the stochastic elements. Moreover, the emphasis on this mitotic mode model construct increases with each successive model, to the extent that in the Boije model there are no other elements that are used to explain RPC behaviours. All cellular behaviours are, in effect, progressively collapsed into this stochastic mitotic mode concept. Finally, the He and Boije SSMs contain more parameters, which are determined by fewer empirical measurements than the Gomes SSM. Clearly, then, the SMME SSMs are at significant risk of representing trivial overfits to their data, the modeller’s equivalent of the “just-so” story, not representing any actually-existing macromolecular or cellular process.

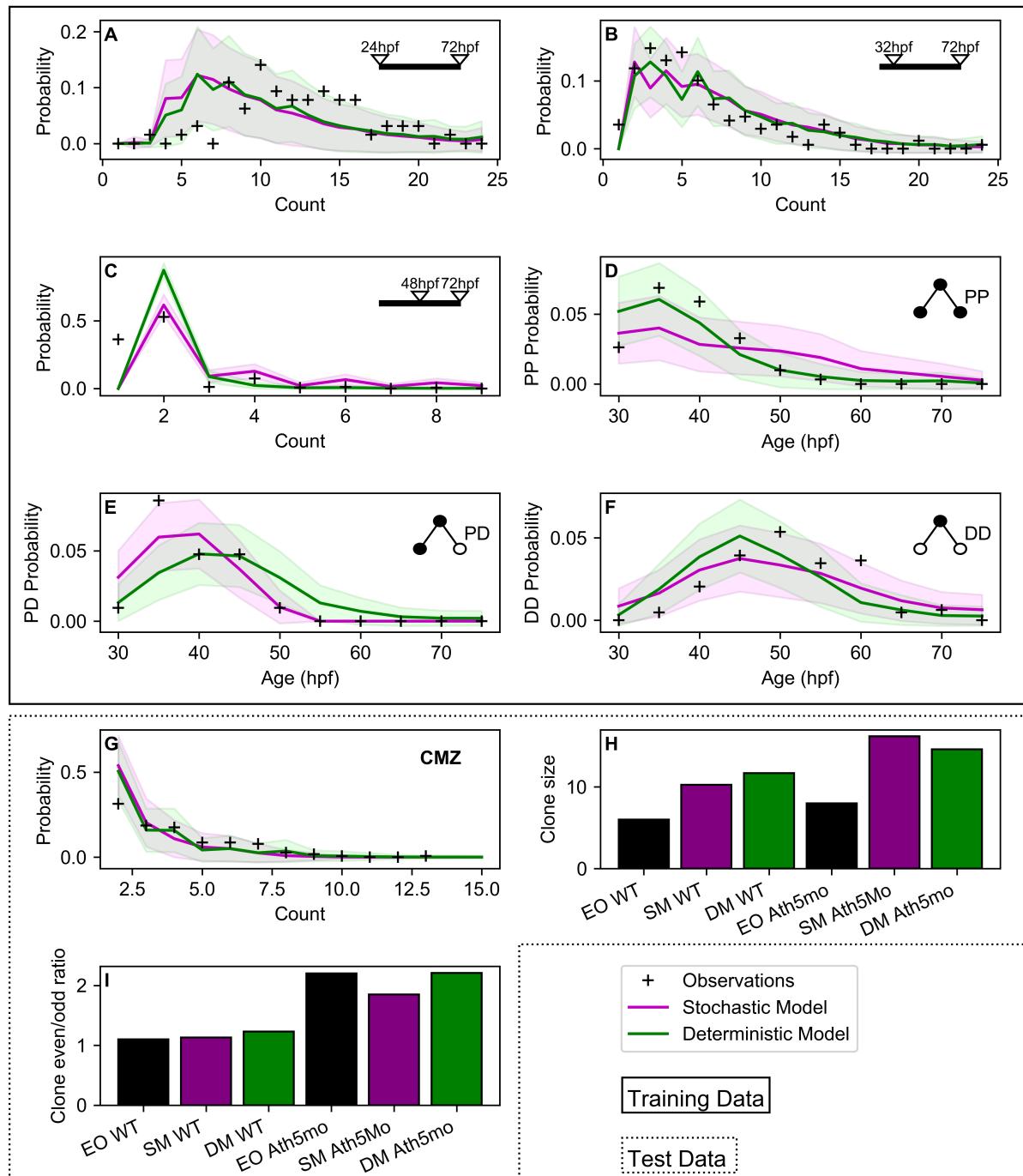
Fortunately, we may employ standard statistical model optimisation and selection techniques to adjudicate this possibility. The most straightforward way to do so is to reexamine the He SSM alongside a competing alternative model. The data to which the He SSM has been applied is particularly amenable to a scheme in which the models are fit to a “training” dataset, followed by a “test” dataset. In this case, the training data are the induced lineage size and mitotic mode rate data from He et al., and the test dataset are the Atoh7 morpholino observations from He et al., and the CMZ lineage size data from Wan et al. This allows us to test how well the He SSM, and an alternative, hold up under novel experimental conditions, without relying on a trivial ordering of goodness-of-fit to training data. Since the Boije SSM draws straightforwardly on the He SSM for its temporal phase-parameterisation and stochastic mitotic mode model ingredient, this analysis of the He SSM also bears directly on the validity of the Boije SSM.

As an alternative to the SMME He SSM, we constructed a model that differs only in its construal of the process governing the RPC mitotic mode. Rather than variability arising from a stochastic-

process mitotic mode changing across phases of fixed length, we simply supposed that mitotic mode is deterministic in each phase (guaranteed PP mitoses in the first phase, PD in the second, and DD in the third), but the phase lengths are variable between lineages and shift slightly between sister cells. That is, we represented this linear progression of deterministic mitotic mode phases using the same type of statistical construct the He SSM applies to model cell cycle length, to avoid introducing any novel or contentious elements into the model comparison. More specifically, each lineage has a first PP phase length drawn from a shifted gamma distribution, followed by a second phase length drawn from a standard gamma distribution. Upon mitosis, these phase lengths are shifted in sister cells by a normally distributed time period, exactly like cell cycle lengths in the He SSM.

We take this to be a reasonable representation of the classic suggestion that RPCs step through linear succession of competency phases, given the conceptual tools that the He SSM uses to model mitotic phenomena. If we suppose that this temporal program is governed by RPC lineages passing through a stereotypical series of chromatin configurations which allow for PP, then PD, then DD mitoses in turn, along with the associated competence to produce the particular cell fates associated with PD and DD mitoses, it seems entirely plausible to suggest that lineages differ in the lengths of time they occupy each state. This is particularly true if we concede that an SSM, forego any spatial modelling whatsoever, must necessarily abstract extracellular and spatial influences on these processes. Moreover, since these chromatin configurations must be broken down and rebuilt with each mitosis, the re-use of the “sister shift” model ingredient from the He SSM’s cell cycle RV is congenial, representing the same sort of cell-to-cell variability that results in the small differences between sister cells in cycle timing.

While the code used to implement the SSMs mentioned above has not been published, the relevant reports provide enough detail to reconstruct these models in full, which we did using the CHASTE cell-based simulation framework, in order to provide transparent and reproducible implementations of the SSMs. Because the values selected for the He SSMs’ parameters seem to have been selected as a series of rough estimates, with only the value for the probability of PD-type mitoses in the second model phase being varied to produce the fit, we suspected that the fit would not be at or near the local minimum for any reasonable loss function. That is, a model fit produced in this manner is likely to be located in a region of the parameter space that is highly sensitive to small perturbations, and therefore may depend strongly on implementation-specific idiosyncrasies. He et al. report that they experienced difficulty in obtaining a good fit to their 32 hour induction data, with changes in the phase two PD probability producing large differences in fit quality. Unsurprisingly, when we rebuilt the He SSM with the original fit parameterisation, we substantially reproduced the original fit, except for the 32 hour data, where the model output diverges substantially from that reported in He et al. (see [S1 Fig](#)). Since this is clearly not the best fit available for the He SSM, and we wish to directly compare the best fits (i.e. the parameterisations at minima of some loss function) for the He SSM and our putative alternative model, we used the simultaneous perturbation stochastic approximation (SPSA) algorithm[[Spa98](#)] to optimise both model fits. SPSA is particularly convenient for complex, multi-phase models like the He SSM, because no knowledge of the relationship between the model’s parameters and the loss function is required. We used Akaike’s information criterion (AIC) as the loss function to be minimised, in order to provide a rigorous comparison between the two differently-parameterised models (the deterministic alternative has two fewer parameters than the He SSM).



**Figure 2.4: Model comparison: the SPSA-optimised He SSM and a deterministic alternative**

Empirical observations (black crosses and bars) and SPSA-optimised model output (magenta, stochastic mitotic mode model; green, deterministic mitotic mode model). Model output in panels A-G is displayed as mean  $\pm$  95% CI. Panels A-F: Training dataset, to which models were fit. Panels A-C: Probability of observing lineages of a particular size ("count") after inducing single RPC lineage founders at (A) 24, (B) 32, and (C) 48 hours with an indelible genetic marker, tallying their size at 72hpf. Panels D-F: Probability density of mitotic events of modes (D) PP, (E) PD, and (F) DD over the period of retinal development observed by He et al. Panel G: Probability of observing lineages of a particular size ("count") originating from the CMZ; RPCs are taken to be resident in the CMZ for 17 hours before being forced to differentiate, lineage founders are assumed to be evenly distributed in age across this 17 hour time period. Panel H: Average clonal lineage size of wild type and Ath5 morpholino-treated RPCs. Ath5mo treatment is taken to convert 80% of PD divisions, which occur in the second phase of the He model, to PP divisions, resulting in larger lineages. Panel I: Ratio of even to odd sized clones in wild type and Ath5 morpholino-treated RPCs.

After (re)-fitting to the training dataset, we calculated AIC for the He SSM (hereafter SM, for stochastic model) and our deterministic mitotic mode alternative (hereafter DM), for both training and test datasets. The combined results are displayed in Fig 2.4, with the output of the two models being presented separately in S2 Fig and S3 Fig. Remarkably, the DM closely recapitulates the output of the SM for both datasets. Moreover, while the more highly-parameterised SM permits a better fit to the training data, the DM proves to be a better fit to the test dataset, as summarised in Table 2.1. This is, therefore, a classic case of model overfitting: a higher-parameter model fits some training dataset better than a simpler model, but fails upon challenge with a new dataset. Given this model selection scheme, it is plain that we should choose the DM over the SM as a superior explanatory model, both on the basis of explanatory power and of Occam’s Razor.

Table 2.1: **AIC values for models assessed against training and test datasets**

Model	Training AIC	Test AIC
Stochastic (He fit)	-93.26	255.64
Stochastic (SPSA fit)	<b>-93.80</b>	92.24
Deterministic (SPSA fit)	-69.33	<b>86.60</b>

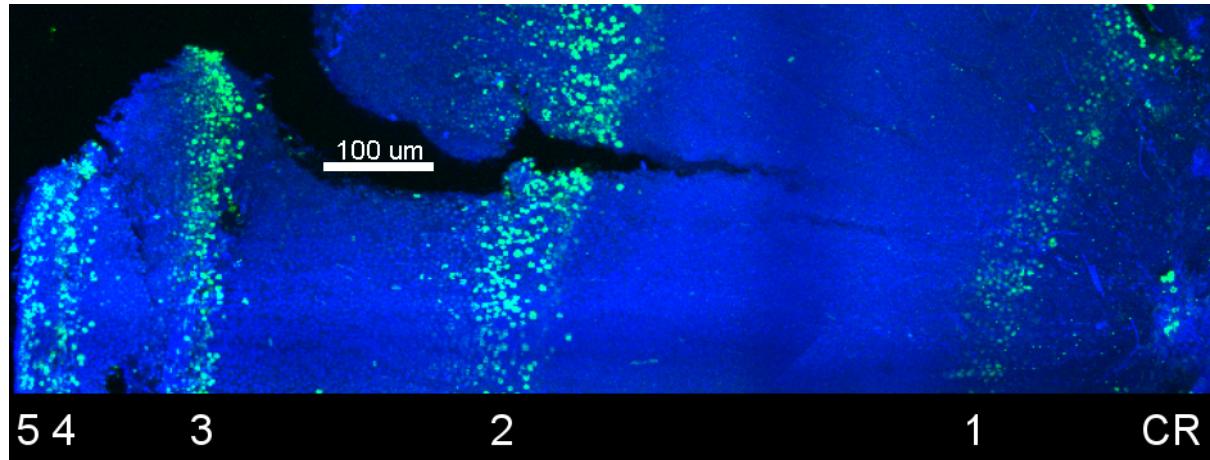
AIC: Akaike’s information criterion. Lower values reflect better model explanations of the noted dataset. Lowest values are noted in bold.

We therefore conclude that, had the Harris group employed standard model selection procedures, comparing plausible alternatives to their favoured explanation, they would have been forced to conclude that the SMME is not the best available explanation for variability in zebrafish RPC lineage outcomes. It is clear from this analysis that the assumed, but unexplained, linear succession of mitotic mode phases is what provides the overall structure of the model output. Variability in lineage outcomes may be supplied by entirely different model ingredients without any loss of explanatory power (indeed, with some improvement, in our case). The rhetorical emphasis on a stochastic process governing mitotic mode, ostensibly ruling out other types of explanation, is unjustified. It is likely that any number of different types of SSMs, representing other sorts of processes (such as asymmetric segregation of fate determinants, or differential spatial exposure to extracellular signals), can produce identical model output. Given this, we conclude that the SSM model type is simply inadequate for the task of locating the source of variability in RPC lineage outcomes.

## 2.2.5 SMME SSMs cannot explain the post-embryonic phase of CMZ-driven zebrafish retinal formation

The zebrafish retina, like other fish retinas, and unlike the mammalian retina, continues to grow long after the early developmental period, indeed, well past the organism’s sexual maturity. This may be unsurprising, given that zebrafish increase in length almost ten-fold over the first year of life[PEM<sup>+09</sup>], necessitating a continuously growing retina during this period. In fact, the quantitative majority of zebrafish retinal growth occurs post-metamorphosis, outside of the early developmental period. This growth occurs due to the persistence of a population of proliferative RPCs present in an annulus at the periphery of the retina, called the ciliary or circumferential marginal zone (CMZ), which plate out the retina in annular cohorts. A typical “tree-ring” analysis from our studies, marking the DNA of cohorts of cells contributed to the retina at particular times with indelible thymidine analogues, is

shown in Fig 2.5. It is immediately obvious from such experiments that the structure of the zebrafish retina is quantitatively dominated by contributions from the CMZ in the period between one and three months of age. Why peripheral RPCs in zebrafish remain proliferative, while those in mammals are quiescent[Tro00], and whether and how their behaviour might differ from embryonic RPCs, remain unresolved. Answers to these questions may have significant fundamental and therapeutic implications, especially given e.g. the possibility of harnessing endogenous, quiescent, peripheral RPCs in humans for regenerative retinal medicine.



**Figure 2.5: Most zebrafish retinal neurons are contributed by the CMZ between one and three months of age**

Maximum intensity projection derived from confocal micrographs of a zebrafish whole retina dissected at 5 months post fertilisation (mpf). The animal was treated with BrdU at 1, 2, 3, 4, and 5 mpf for 24hr. Anti-BrdU staining of the whole retina reveals the extent to which the CMZ (which is responsible for retinal growth after approximately 72hpf) contributes new neurons in tree-ring fashion, extending out from the center of the retina (CR), which is formed before 72hpf. Monthly cohorts are labelled appropriately. Scale bar, 100  $\mu$ m.

Indeed, the SMME SSMs were originally brought to our attention when the He SSM was deployed in Wan et al.[WAR<sup>+</sup>16], with the claim that the He SSM explains the behaviour of CMZ RPCs. Wan et al. argue that a slowly mitosing population of bona fide stem cells, at the utmost retinal periphery, divides asymmetrically to populate the CMZ with He-SSM-governed RPCs. In other words, the usual suggestion that CMZ RPCs undergo a somewhat different process than embryonic RPCs, perhaps recapitulating across the peripheral-central axis some progression of states or lineage phases that embryonic RPCs pass through in time[HP98], is repudiated in favour of one model which describes the behaviour of all RPCs throughout the life of the organism, with the addition of a small population of stem cells to keep the CMZ stocked with RPCs. If so, this might suggest that the problem of activating quiescent stem cells in the retina is simply that- one need only sort out how to throw the proliferative switch in these cells, since the proliferative and fate specification behaviours of the resultant RPCs will reliably be the same as those observed in development.

While we determined that the SMME is not the best available explanation for RPC lineage outcomes, we still felt that the SSMs associated with this explanation might be used to elucidate this point. In particular, if it is the case that the He SSM provides good estimates of lineage size and proliferative

dynamics in the early zebrafish retina, it should be possible, using this model, and the estimates of putative stem cell proliferative behaviour provided by Wan et al., to simulate the population dynamics of the CMZ, at least through the first few weeks of the organism's life.

In pursuing this point, we noted a peculiar feature of the He SSM not documented by any of the SMME reports: the cell cycle model overstates the *per-lineage* rate of mitoses by as much as a factor of 3. That is, the mitotic mode rate data presented in He et al., recapitulated here in Fig 2.4, panels D, E, and F, and used to optimise both the SM and DM, are probability density functions that are not standardised on a per-lineage basis. These data simply indicate the distribution of mitotic events of a particular type. When we take all of the mitotic events documented by He et al. and calculate the probability of any such event occurring per lineage, per hour, we obtain the values presented in Fig 2.6.

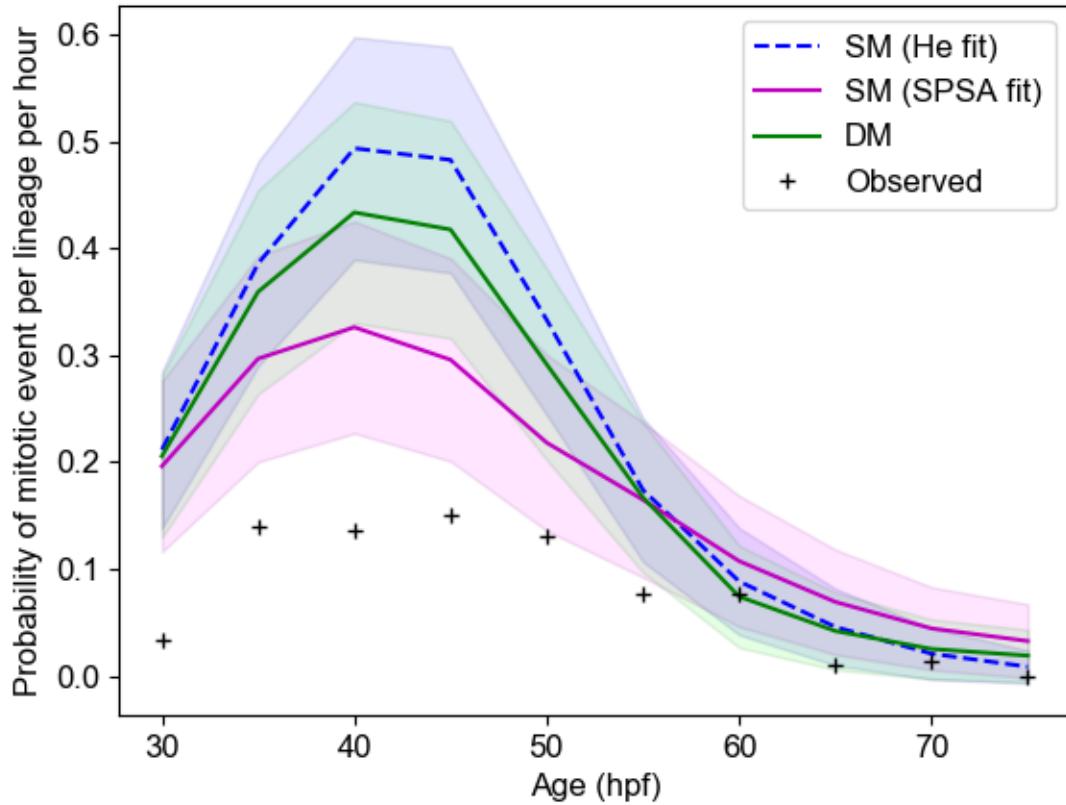


Figure 2.6: **Per-lineage probabilities of mitoses, He et al. observations compared to model output**

Probability of observing a mitotic event over the period studied by He et al., per hour, per lineage. Model output is presented as mean  $\pm$  95% CI.

Similarly, when we performed cumulative thymidine analogue labelling of the 3dpf CMZ, (using the assumptions of Nowakowski et al. [NLM89], which treat the proliferating population as a homogenous, and dividing asymmetrically; these assumptions are flawed but adequate for a rough estimate) we obtain an average cell cycle length of approximately 15 hours, more than twice as long as the He SSM's mean

cycle length. These data are displayed in 10.2.1. Therefore, the He SSM (in both its original and refit parameterisation, as well as the deterministic alternative, since they all rely on the same proliferative model elements) substantially overstates the proliferative potential of both embryonic and early CMZ RPCs. Still, because we observed a massive build-up of proliferating RPCs between two the first few weeks of post-embryonic CMZ activity involve a massive build-up of proliferating RPCs between two and four weeks post-fertilisation, we thought this might actually suit this later context better.

In order to produce estimates of total annular CMZ population in zebrafish retinas over time, we counted proliferating RPCs present in central coronal sections of zebrafish retinas throughout the first year of life, treating these as samples of the annulus, and calculated the total number of cells that would be present given the diameter of the spherical lens measured at these times. Our simulated CMZ populations were constructed at 3dpf by drawing an initial population of RPCs governed by the He SSM (using the original fit parameters, which further exaggerate the proliferative potential of these lineages) from the observed distribution. An additional number of immortal stem cells amounting to one tenth of this total was added. This is likely an overestimate, given that these putative stem cells are thought to be those in the very peripheral ring of cells around the lens, of which typically two to four may be observed in our central sections with an average of over one hundred proliferating RPCs. Moreover, these simulated stem cells were given a mean cycle time of 30 hours, proliferating about twice as quickly as Wan et al. suggest. Finally, to reflect the fact that these stem cells contribute to the retina in linear cohorts[CAH<sup>+</sup>14], and more of them are therefore required as the retina grows, the stem cells were permitted to divide symmetrically when necessary to maintain the same density of stem cells around the annulus of the lens. Two hundred such CMZ populations were simulated across one year of retinal growth.

The results of these simulations are displayed in Fig 2.7, overlaid over CMZ population estimates derived from observations. Given the generous parameters of the population model, consistently overestimating the proliferative potential of embryonic and early CMZ RPCs, it is surprising that the He SSM proves completely unable to keep up with the growth of the CMZ in the first two months of life. Indeed, the unrealistically active stem cell population the simulated CMZs are provided with is unable to prevent a near-term collapse in RPC numbers, only catching up after the months later as the number of stem cells increases with the growth of the lens. Thus, even given permissive model parameters, the He SSM is not able to recapitulate the quantitatively most important period of retinal growth in the zebrafish.

This analysis strongly suggests that observations of RPCs in embryogenesis and early larval development are unlikely to provide a good quantitative model of the development of the zebrafish retina, even abstracting away spatial and extracellular factors as SSMs necessarily do. Our data point to a second, quantitatively more important phase of retinal development, between approximately one and four months of age, in which RPCs are far more proliferatively active than in early development. It is likely that models that closely associate proliferative behaviour with fate specification, like those of the SMME, will necessarily be unable to explain this period. Recent evidence suggests that mitotic and fate specification behaviours in RPCs may be substantially uncoupled[ESY<sup>+</sup>17]. This would permit the CMZ population to scale appropriately with the growing retina. Alternatively, it is also possible that RPCs are substantially heterogenous with respect to their proliferative behaviour, and that this heterogeneity is mainly apparent later in development.

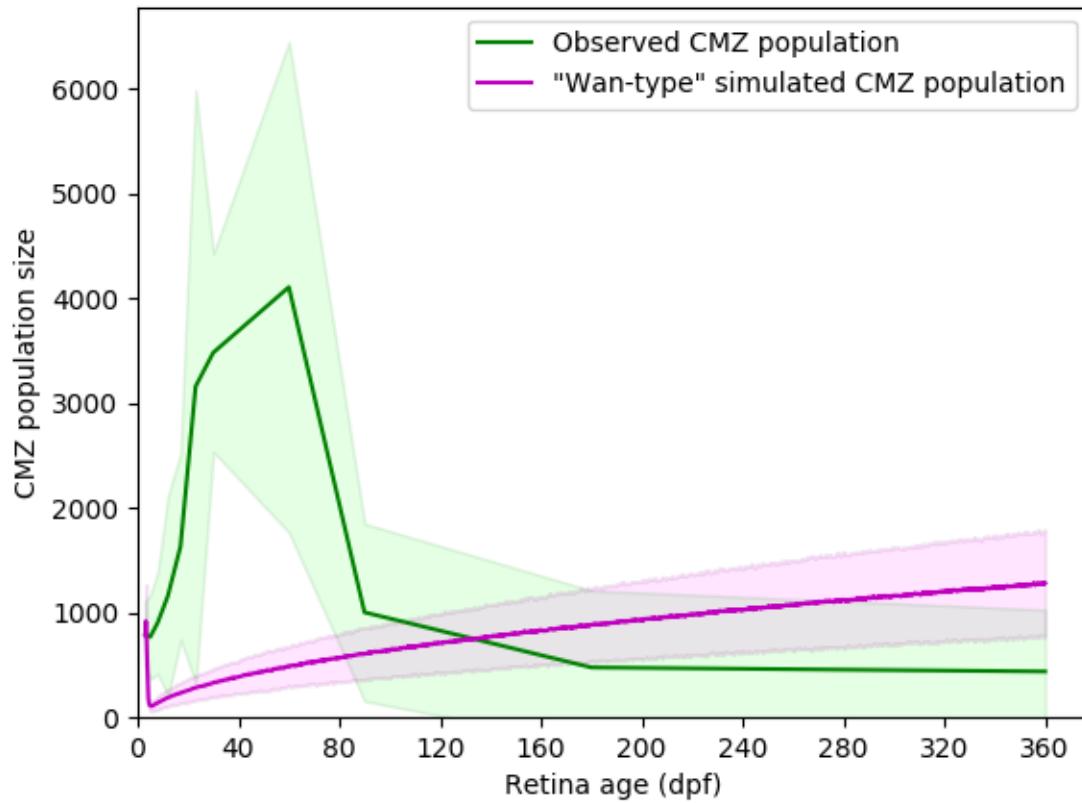


Figure 2.7: CMZ population of proliferating RPCs: estimates from observations and simulated Wan-type CMZs

Annular CMZ population was estimated from empirical observations of central coronal cryosections stained with anti-PCNA, a marker of proliferating cells, standardising by the diameter of the spherical lens observed in these animals. "Wan-type" CMZs were initialised with a number of RPCs drawn from the observed 3dpf population distribution, and given an additional 1/10<sup>th</sup> of this number of immortal, asymmetrically dividing stem cells, as described in the text, and simulated for the first year of life. All data is presented as mean  $\pm$  95% CI.

## 2.3 Conclusion

Simple stochastic models are familiar tools for stem cell biologists. Introduced by Till, McCulloch, and Siminovitch in 1964, they proved their utility in describing variability in clonal lineage outcomes of putative stem cells, originating from macromolecular processes beyond the scope of cellular models (and beyond the reach of the molecular techniques of the time). More prosaically, their simplicity afforded computational tractability in an era when processing time was relatively scarce, allowing early access to Monte Carlo simulation techniques. That said, their abstract nature necessarily emphasizes an aspatial, lineage-centric view of tissue development, which cannot account for the generation of structurally complex tissues like retinas beyond cell numbers, and, perhaps, fate composition. As a result of this relatively loose relationship to the complex morphogenetic environment, there are few

constraints on the model configurations that may produce similar outputs. As we hope has been made plain by the analyses herein, this can result in modellers being led astray by their apparently good fits to observations. This is one reason why it has long been unacceptable in other biological modelling communities (particularly, among ecologists) to present the fit of one highly parameterised model as evidence for some theory; it is very rare that there is not a model representing an alternative theory that cannot be made to produce a reasonable model fit. Indeed, as we demonstrate here, the use of standard model selection techniques may make plain that a completely contradictory theory is a better explanation for the data.

To some extent, this can be ameliorated by careful attention to the particular macromolecular or cellular referents that particular model constructs represent. The “mitotic mode” construct present in all SSMs is a particularly ambiguous and problematic one in this regard. “Mitotic mode” is not, itself, a property of a mitotic event, but is rather a retrospective classification of the event after an experimenter observes whether progeny resulting from the event continue to proliferate. Its original appearance in SSMs was simply to allow calculation of clonal population sizes; it was never intended to represent a particular type of process or “decision” made by cells at the time of mitosis to continue proliferating or not. Any number of pre- or post-mitotic signals and processes may result in a particular mitosis being classified as PP, PD, or DD, without anything about the mitotic event itself determining this. The use of such a retrospective classification, rather than the identification of some physical property of the mitotic event (such as the asymmetric inheritance of fate determinants), is straightforwardly a concession that the actual macromolecular determinants of cellular fate are outside the scope of the model.

The SMME represents an attempt to connect this abstract, retrospective model construct to observations of noisy gene transcription[Rv08]. Motivated by the observation that momentary mRNA transcript expression in RPCs is highly variable from cell-to-cell[TSC08], the suggestion is that this transcriptional variability may be responsible for the observed variability in RPC lineage outcomes. Since the extent of transcription noise may be “tuned” to a degree by e.g. promoter sequences[RO04], parameters related to this noise could plausibly be under selective pressure. There are two significant problems with identifying the SMME’s stochastic mitotic mode with this type of process. The first we have demonstrated by constructing an alternative model with deterministic mitotic mode but variable phase lengths: the source of variability may be located elsewhere without compromising the explanatory power of the model. It is therefore impossible to determine what sort of process might give rise to variability in RPC lineage outcomes by using SSMs in this fashion. The second, more fundamental problem is that a causal explanation of the presence of the signal, for which noise is a property, is elided entirely in favour of emphasis on “stochasticity”. This is most apparent in the Boije SSM. In this model, Atoh7 and Ptfla TFs are available to provide their noisy signal in the 4th and 5th lineage generations, and at no other time. We do not dispute that a noisy signal may contribute to variability in that signals’ effects; rather, we suggest that it is the temporal structure of such a signal (assuming this structure can be empirically demonstrated, rather than assumed) that calls for causal explanation. The SMME reports explicitly disclaim the necessity for “causative hypotheses” in the case that a stochastic model provides a good fit to observations. As Jaynes remarked in his classic text on probability theory, “[stochasticity] is always presented in verbiage that implies one is describing an objectively true property of a real physical process. To one who believes such a thing literally, there could be no motivation to investigate the causes more deeply ... and so the real processes at work might never be discovered.”[JBE03]

In identifying the model construct with the physical processes determining RPC outcomes, the SMME

obscures what seems to us to be the primary lesson to be drawn from the Harris groups' beautiful *in vivo* studies of zebrafish RPCs. That is, compared to late rat RPCs in dissociated clonal culture, RPCs in intact zebrafish retinas produce far more orderly outcomes. To the extent that these outcomes are variable, the source of variability remains unidentified. We suggest that, in order to produce good explanations for both the order and unpredictable variability exhibited in RPC behaviours, more sophisticated models and more rigorous modelling practices are required. Resort to "stochasticity" as an explanatory element should not be made in the absence of model comparisons that rule out alternatives with well-defined causal structures, lest we fall into the trap Jaynes warned us about.

## Chapter 3

# Toward a computational CMZ model comparison framework

### 3.1 SMME Postmortem: a wrong turn at Gomes

Let us return to the question posed in [Section 1.6](#): has Harris succeeded in finding order by blurring out the chaotic welter of mechanistic explanations for RPC function in retinogenesis? [Chapter 2](#) argues that it has compounded several modelling failures to obscure the basic reality of the matter: the SMME models do not support an explanation based on the noise of some macromolecular process, they rather support explanations based on the original idea of a linear, deterministic series of stages through which RPCs progress, originally put forward by Cepko et al. [[CAY<sup>+</sup>96](#)]. This is the fundamental model ingredient that produces the structure resembling the data, not the various random variables associated with mitotic mode, which we have proven by demonstrating that a deterministic progression of mitotic mode models the data better than its stochastic counterpart. Only by the process of counterinduction, the comparison of models with different propositional structures about reality, does this become clear.

The critical failing of the SMME models is where they depart from the original Gomes SSM: the assumption of the linear structure of temporal phases. This has to be done in order to add the early contribution of RGCs in the modelled zebrafish neurons, and is an essential explanandum for any explanation purporting to relate to retinogenesis; to the extent that it is left without some biological rationale, retinogenesis remains unexplained. It follows that the SMME does not achieve its lofty aim of being a complete description of the aspects of retinogenesis an [SSM](#)[SSM] is capable of explaining, and this is confirmed by our observation that the Wan model [[WAR<sup>+</sup>16](#)] has no explanatory power outside the first few days of life, which makes up a tiny portion of the total retinal contribution to the zebrafish retina.

We must also assess that Harris' first theoretical maneuver, explaining the data with a model whose structure supports a stochastic resolution of mitotic mode "decisions", fails in the face of an alternative which locates stochasticity elsewhere. Simply by introducing variability into the lengths of the linear temporal program already assumed by the SMME models, we can remove variability from the mitotic mode and produce a superior model. The underlying data used to inform these models speak better to any entirely different selection from the array of theoretical options outlined in [Section 1.2](#), the linear progression of competencies. Because the SSM model form does not usually have spatial dimensions, we did not test models involving variability in extracellular signals, but it is a good bet that such a model

could be made to fit about as well as either of the ones tested in the previous chapter. In effect, if we are to follow Harris' inferential logic, variability in RPC outcomes can be explained by variability in virtually any parameter which affects fate outcomes. The explanation collapses into itself, which is why Boije et al. is forced to defend their idiosyncratic definition of "stochasticity" at length: variability is being used to explain variability. If "randomness" or "stochasticity" is accepted as a property of existents, rather than representations of our uncertainty about them, all biological variability is trivially explained by referring to it. Because of the seriousness of this problem for biological inferences<sup>1</sup>, an explanation of the Bayesian epistemological view of probability has been provided in Section 14.0.2. Moreover, thorough explanation of the sense in which there is no good argument for "randomness" being a property of real existents is provided in Section 14.5.2. It suffices here to conclude that variability in RPC outcomes is not well explained by the SMME models, in part because their interpretation depends entirely on an incorrect understanding of the concept of stochasticity, and in part because they were never tested against alternatives.

The second theoretical maneuver was to nominate a particular macromolecular system as the physical locus of the stochastic process, in this case represented by *Atoh7* and *Ptf1a* as labels for abstract Bernoulli distributed processes. It remains obscure in what sense the model variables relate to their namesake transcription factors (transcription of the TF itself? activation of other genes?). Plainly, these factors are involved in relevant RPC behaviours, but it is unclear why they have been nominated as causally upstream of the "mitotic mode" selection. Since the Boije model inherits the assumption of a linear progression of stages, it is very likely that a model with similar explanatory power could be built using the same strategy outlined above, locating random variability outside mitotic mode, although this would be superfluous.

On the basis of the above, we can conclude that the sole formally testable model of RPC function in the zebrafish retina is not adequate for our purposes. But how did we get here? As noted in Section 1.3.2, the notion that the most significant proliferative and speciative phenomena associated with RPCs are produced by mechanisms intrinsic to the cells derives much of its empirical support from the Raff group's work. This story began developing with the observation that co-culturing E15 rat RPCs in dissociated pellet cultures with P1 cells did not accelerate the appearance of the first rods derived from the E15 progenitors (which occurred at a similar time as *in vivo*), suggesting that the speciative "schedule" of these cells is at least partially intrinsic<sup>2</sup> [WR90]. The scope of these observations were dramatically expanded by Raff's subsequent work, intended to address the relative significance of intrinsic versus extrinsic processes in RPC function by comparing clonal RPC lineages in fully dissociated clonal-density cell culture to those in intact explants [CBR03]. The remarkable finding of this study was that dissociated E16-17 rat RPC lineages produce very similar numbers and types of retinal neurons, albeit without morphological or molecular markers of mature neurons. This observation provided strong evidence for the predominant importance of RPC-intrinsic processes in determining both the proliferative and speciative outcomes for late RPC lineages, since the complex, spatially organised context of intact explanted tissue seemed to only be required for the maturation of neurons, and was not required to regulate proliferation or the initial commitment to an appropriate distribution of lineage outcomes. As these are the two most obvious and critical parameters that must

---

<sup>1</sup>Having experienced well-respected stem cell biologists confidently asserting that Harris' work proves that RPC mitotic outcomes are 'spontaneous', ie. causeless and without explanation, I believe these studies have been both influential and confusing for many.

<sup>2</sup>Co-culturing with P1 cells, did, however, significantly increase the proportion of E15 RPCs specified as rods, resulting in Raff's suggestion here that both intrinsic and extrinsic factors are important.

be achieved for RPCs to produce a functional retina of the appropriate size, the suggestion that both are largely determined by intrinsic processes seemed a striking confirmation of Williams and Goldwitz's much earlier suggestion [WG92], against the prevailing view of the day, that lineage had a greater role to play than cellular microenvironment in RPC contributions. Interestingly, Raff's interpretation of their 2003 data was that RPCs were most likely stepping through a linear, programmed developmental sequence rather than undergoing shifts in the probabilities of variable outcomes over time. It is notable that this interpretation arises not from the data collected in their study, but from considerations of a single unusual clone reported by [TSC90], and by analogy with drosophila neuroblasts. In retrospect, these arguments for linear sequences of deterministic RPC outcomes do not seem particularly strong, and it is perhaps unsurprising that these studies are remembered mainly for highlighting the importance of RPC-intrinsic processes.

The Gomes study, with its detailed study of particular rat late-embryonic RPC lineages, thus seemed to solidify the notion that unpredictable RPC-intrinsic processes dominate lineage outcomes [GZC<sup>+</sup>11]. But this study, like its forebears originating in Raff's work, is concerned with a population of RPCs that is too old to produce RGCs. This is the essential point: the SMME does not even try to explain the early appearance of RGCs in terms of some macromolecular mechanism, although this is the single most significant difference between the zebrafish RPC lineages studied by Harris and the rat lineages studied by Raff and Cayouette [CBR03, GZC<sup>+</sup>11]. It is only by assuming the unexplained linear temporal structure of RPC specification that Harris is able to continue the rhetorical focus on the stochastic elements of the model, and so is lead down the garden path of spurious model fits and logically unsound interpretation of model structure.

## 3.2 Implications of the SMME's failure for modelling CMZ RPCs

Having taken seriously the possibility that an adequate model of zebrafish retinogenesis already exists, and concluded in the negative, we must proceed to a plan to rectify this state of affairs.

Firstly, the obvious lessons in statistical acumen must be learned. The statistical practices used in the SMME reports are not acceptable by any contemporary standard, and would not have been accepted in another field where practitioners are more familiar with modern statistical techniques. We must be dissuaded of the notion that a single, hand fitted model has any explanatory power at all; it may be an exercise in modelling prowess, but it is not a legitimate part of the quantitative scientific discourse unless some statistical property of the model is actually computed<sup>3</sup>. Moreover, to have any confidence about our interpretation of the model, we must test alternatives that make fundamentally different assumptions about the causal structure of the phenomenon being explained by the models.

Can we conclude that the approach used in Chapter 2 is adequate to our task? There are two broad elements to consider here: the appropriateness of the overall modelling approach represented by the SSM in relation to the hypotheses we wish to test, as well as the soundness of the statistical procedures.

Taking up the SSM, by far its most attractive features are its computational efficiency and the ease of producing a custom model to fit some particular case- for all their failings, the SMME models include some elements like the correlation of cell cycle lengths in mitotic sisters that greatly improve the temporal modelling of RPC lineage outcomes, for instance.

---

<sup>3</sup>The commonly maligned Fischerian t-test procedure [HKB08, pp.181] is, in this sense, better than simply plotting some model output over observations.

The observations arising from the SMME strongly suggest that we would like to test hypotheses about RGC specification in order to find better models of RPC function. Carefully reconsidering the hypothesis of Neumann et al. [Neu00], that Shh from nearby RGCs induces cell cycle exit and RGC specification, would seem to be a high priority, for instance. Can such a scenario be represented in an SSM? One could model the probability of being within Shh induction range of an RGC in the early part of a lineage's life, for instance. Doubtless, a model that incorporated variable RPC specification outcomes determined on this basis could be made to fit data about as well as the variable mitotic mode or variable phase length models tested above. That said, it is not very clear that the aspatial data to which SSMs are fitted is capable of constraining the likelihood of model alternatives that include spatial components. While we determined that our deterministic mitotic mode model fit test data somewhat better than the stochastic model, the modest size of the calculated AIC differences suggests that comparing SSMs is not a particularly good way to distinguish even alternative hypotheses about the structure of processes the SSM models explicitly, like cell cycle length and mitotic mode.

We can therefore predict that we will be unable to test important models, involving documented macromolecular explanations of RGC specification, without the ability to represent some spatial information. Still, the computational benefits of the SSM are too appealing to leave out of the toolkit entirely. Taken together, this suggests that we need a very general method of testing models of potentially very different structures against one another.

Turning then to our statistical procedures, are these adequate to our goals? We can broadly characterize the approach taken as estimating the local optimum of a loss function for model output, given the dataset; specifically, minimizing Akaike's information criterion by simultaneous perturbation stochastic approximation (SPSA). This procedure has some advantages. AIC is a well-understood measure, well-grounded in information theory, which penalizes superfluous model complexity in a consistent and rigorous way. SPSA is a widely used, well understood algorithm that can be applied to any reasonable euclidean parameter spaces; cases where parameter spaces are bounded (typical of biological models where negative parameter values are usually nonsensical) are explicitly accounted for, and so on. This procedure is better than many that are available.

Still, after the experience of the SMME model comparison, a certain uneasiness is warranted. The AIC calculation is, after all, based on a single estimate for the locally optimal parameterisation of the model. Relative AIC rankings, then, are strongly influenced by the "well depth" of the AIC surface at the local minimum in parameter space. We can easily imagine a case where a model with a very narrow range of parameter space that fits data very well appears to be better, by the AIC metric, than one that fits the data slightly more loosely, but over a much broader range of the parameter space. The first model is a "fragile" fit, probably depending heavily on the particular structure of the training dataset, while the second would likely be much more robust across a range of observations; still, in this case, AIC would lead us to select the fragile model over its robust counterpart. Indeed, blind interpretation of AIC rankings has lead to its use in ecology being described as a "cult" [BB20]. It is easy to imagine practitioners being reduced to "AIC hacking" in the same manner that "p hacking" occurs in order to achieve some arbitrary value for a hypothesis.

Moreover, while SPSA is an eminently practical algorithm, useful in a wide variety of contexts, its statistical guarantee is only that it will almost-surely find the local minimum of the loss function. While this will be good enough in some applications, for highly parameterised models with complex loss function surfaces, there will be many local minima for SPSA to get "stuck" in that are nowhere near the

global optimum<sup>4</sup>. If we are evaluating hypotheses in order to make decisions about potentially years-long research projects, more certainty about the reliability of the method is necessary. Finally, while SPSA is requires much less computational effort than more global Monte Carlo parameter estimation techniques like simulated annealing or Hamiltonian Monte Carlo, it requires about as much application-specific tuning.

Fortunately, a complete system of Bayesian inference which addresses all of the problems mentioned above has been promulgated over the last 15 years: nested sampling. Originally introduced by John Skilling [Ski06], this use of this system has become widespread in cosmology [Tro08], where its generality and ability to cope with complex, high dimensional parameter spaces has been well proven.

### 3.3 Desiridata for spatial CMZ models in a putative model comparison framework

The simulations presented in ?? consisted solely of abstract collections of lineages. The members of these model colonies have no activities beyond proliferation and no functional attributes beyond their proliferative status<sup>5</sup>. Despite the underlying code consisting of relatively high performance C++, these simple models nonetheless occupied the local component of the cluster used in this work for some weeks. We can therefore see how a simple approach to ranking basic models against a smallish dataset approaches the reasonable limits of what most labs are likely to be able to achieve with any given machine in a month or so. Because the introduction of spatial simulation implies, perhaps, an order of magnitude more computational time, we may plausibly be limited to one inference based on model comparison per year with local resources. The problem of computational limits to model selection is discussed in the relevant section of the Technical Appendix, so it will suffice here to state that this constraint dominates all other considerations in the selection of the overall boundaries within which we intend to build models of the CMZ. In any environment where funding constraints limit the cloud-export of computational burden from the confines of the research institution, it is reasonable to proceed upwards in model computational expense from the cheaper-but-inadequate in search of the adequate-but-still-affordable. In this case, we move from the aspatial SSM into the realm of explicit spatial modelling, seeking the minimum model complexity required to compare hypotheses of interest to us.

#### 3.3.1 Spatial dimension of the models: the "slice model"

Although decomposing the RPC population of the CMZ into a collection of unordered, independently-proliferating SSMs prevents us from assessing many interesting hypotheses, a computational model of the entire CMZ or retina may not be required to compare many interesting hypotheses. Because numerous observations suggest that individual RPC lineages contribute to the retina in linear cohorts of neurons, it follows that any particular centrally-oriented slice of the CMZ annulus will be responsible for the generation of the neurons central to it. Conceptually, then, the retina can be thought of as a series of these "slice units", lined up radially like slices of pie. A complete "slice unit" would include the central-most larval remnant, contributed by embryonic retinogenesis, surrounded by the CMZ's more ordered neural contribution from the postembryonic period, and, peripherally, the CMZ itself. Depending on

---

<sup>4</sup>This is part of what is meant by "the curse of dimensionality", when speaking of the difficulty of sampling the loss function in high dimensional parameter spaces.

<sup>5</sup>I.e. the specified identity of post-proliferative cells in these models has no function within the model.

the hypotheses to be tested, the differentiated central retina may be mostly irrelevant, so the modelled slice may consist mainly of the CMZ and its interface with these central neurons. It is important to note that these slices are conceptually different from the linear cohorts generated from particular lineages, the so-called ‘ArCCoS’, which justify the slices. It is not necessarily the case that a slice model would consider only one RPC lineage; the slices are better thought of as spatial boxes that sample the CMZ and adjacent central neurons, the conceptual equivalent of the histological section through the eye<sup>6</sup>.

Slice models are especially attractive because the observed proliferative status, position, specified fate, etc. of simulated cells can be easily related to the summary statistics used to describe fixed sections of retinal tissue in this thesis and elsewhere. If our histological sections are of an appropriate width, the slice model can be selected to produce output that is directly comparable to observed histological results. This is especially important for studying the postembryonic CMZ, which rapidly becomes optically inaccessible due to the increasing thickness and pigmentation of overlying tissue, so that live imaging cannot be used <sup>7</sup>.

If the retina can be usefully thought of as a series of slice models, and the activity of the CMZ is basically homogenous around its circumference, it follows that the activity of the CMZ can be usefully represented with a single such model. Moreover, the slice need only include one portion of the retinal periphery, the orientation of which is irrelevant (i.e. the model parameterisation for the dorsal portion of a coronal slice is identical to the ventral). It may be erroneous to abstract such a slice from its context, because the modelled cells are in contact with adjacent slices. However, if the CMZ homogeneity premise is valid, this can easily be incorporated into the model by providing a layer of “ghost” cells on either side of slice whose parameters are determined by the slice itself <sup>8</sup>. That is, given the premises, a slice model which interacts with copies of itself is a complete model of CMZ-driven retinogenesis.

All this said, the zebrafish retina is a manifestly asymmetrical structure, with an optic nerve positioned ventro-temporally relative to the center of the optic cup. The CMZ itself is generally understood to be an asymmetric structure, with a larger dorsal than ventral population. This calls into question the assumption of CMZ homogeneity outlined above. It is nevertheless plausible that the appearance and maintenance of these structural features of the zebrafish retina could be explained with a set of slice models, for example, one for each of the dorsal, ventral, nasal, and temporal extrema. Ultimately, these considerations only relevant if our objective is to compare model explanations for retinogenesis as a whole. If we restrict ourselves to the more modest goal of, say, ranking causal influences on the proliferative and specificative behaviours of RPCs in the dorsal extremity of the CMZ, this could provide significant insight into the regulation of peripheral stem cells in other species. In this sense, a single slice model could still be valuable even if it fails to explain aspects of tissue-level zebrafish retinogenesis.

### 3.3.2 Temporal resolution of the models

Another important consideration for any CMZ model comparison framework is the time-scale of any phenomena which are to be admitted as possible causal contributors to retinogenesis. While it is reasonable to think that proliferative events may be well-described with a model that operates on a scale of days and fractions thereof, and that such a model would be well suited to describing the full sweep

---

<sup>6</sup>The case of only one simulated lineage may still arise given thin enough sample boxes or old enough animals, but it is a limiting case and probably would not be the norm.

<sup>7</sup>It is worth noting that the spatial parameters of such models would pertain to fixed and not live retinas; it is possible that some scaling relation compensating for fixative shrinkage could allow the inclusion of live imaging data.

<sup>8</sup>This is implemented in CHASTE as “ghost nodes”.

of CMZ activity across the life of the organism, very few of the relevant macromolecular processes are likely to be well-described with a resolution more coarse than seconds or minutes. The implied difference in the number of calculations required to simulate any given time period is thus several orders of magnitude. A simplifying assumption of the temporal homogeneity of CMZ activity would allow us to abstract long developmental time frames; an explanation that pertains to a few hours can perhaps be extended without recalculation.

If an assumption of temporal homogeneity proves inadequate, the functional structure of the simulation must be structured by this consideration. To illustrate this, consider a case where we wish to incorporate measurements of eye pressure, or membrane tension across the retina, as inputs into the proliferative activity of the CMZ, by way of progenitor cortical tension [Win15]. This would permit assessing whether modelling tissue-mechanical inputs to cellular activities allows us to extract additional information from our observations. If such physical model elements are to be included, the functions which relate physical parametric data to the proliferative behaviour of CMZ progenitors must not operate on a time scale too short to reasonably cover the period of interest. Let us suppose we are mainly interested in explaining the assembly of functional units of the mature, specified retina by the CMZ, from the first division of the presumed distal stem cell responsible for the unit to the determination of the last neuron of its functional column. This process certainly takes days; the cumulative thymidine analogue labelling conducted for this thesis never revealed labelled cells in the structured, determined retinal layers within 24 hours of the end of the analogue pulse. Generally, one must wait at least 3 days before reliably finding most of a labelled cohort in the mature retinal laminae. In this case, we might prefer simplifying assumptions about the relationship of measured retinal surface tension to cortical tension's presumed effects on proliferative activity, over a detailed finite element simulation of cortical tension itself with a resolution of minutes.

### 3.4 Bayesian decisionmaking for structuring models under uncertainty

From the foregoing discussion, we can see that the extent to which model simplifications are justified, and the types of phenomena that could be explained with these models, depend heavily on considerations that can only be informed by observations. As has been demonstrated in ??, the growth of the CMZ population cannot be explained by models fitted to embryonic RPC activity. However, it remains unclear what sort of alternative structure is justified by observations, given our uncertainty about inferred parameters of the populations being measured.

Bayesian statistical methods provide a convenient way to approach this problem. They allow the direct estimation of hypotheses' credibility given data, expressed as a probability. They also permit as well as direct comparison of the evidence for linear regression models, taking into account uncertainty on the model weights. These calculations have straightforward interpretations which allow us to answer questions about the likelihood of the assumptions discussed above actually obtaining at particular times and places in the CMZ.

All of the measurements presented here were obtained from groups of fish of a particular age. Because the measurements are population counts and tissue dimensions in different individual fish, it is assumed that they are the outcome of many independent causal processes. The central limit theorem, therefore, justifies the assumption that the measurements are normally distributed in the population of fish of any

given age.

Given this normal model of measurement distribution in the cohort, uncertainty on the mean and variance of the model is represented with a normal-gamma distribution over those parameters. We then calculate the marginal posterior distribution of the mean, given our uncertainty about the model parameters. These marginal posterior distributions are subsequently sampled by Monte Carlo in order both to calculate derived quantities with appropriate credible intervals, as well as to empirically estimate the probability of various mean comparison hypotheses. Separately, the linear regression technique often known as "Empirical Bayes" was used to perform evidence-based selection on linear models of CMZ proliferation.

The techniques used here are not complex, but they may be unfamiliar to some readers. In addition to the brief summary above, detailed methods can be found in [Section 11.3](#) of the Supplementary Materials, while more extensive theoretical background is available in [Section 13.1.1](#) of the Theoretical Appendix.

## Chapter 4

# Bayesian periodization of postembryonic CMZ activity

### 4.1 Preliminaries: Calculation of Bayesian evidence militates for independent log-Normal modelling of CMZ parameters

To take up the question of how CMZ RPC activity evolves over time, we have collected observations of a variety of CMZ and retinal parameters derived from histological studies of cryosections of zebrafish eyes harvested over the first year of the animal's life. We wish to extract as much information as possible about the structure and time-evolution of the CMZ population as a whole, in order to know how to best model its constituent RPCs. The initial approach here will be to estimate parameters of the whole-eye CMZ from sample cryosections, and to use these estimates as the dataset which will inform our selection of appropriate models.

While it remains common to assume that population data are normally distributed, and so simply to calculate means and standard deviations as the statistical representation of the underlying population, it has been known for decades [Hea67] that log-normal distributions are usually better models of the outcomes produced by additive processes with small, variable steps (like population sizes or income distributions). We therefore start with the selection of appropriate models for the CMZ- and retina-level population measurements critical to the inferences that follow.

As a first pass at this question, we calculate the likelihood ratio for the hypothesis that the parameter measurements, and the calculated quantities derived from them, are log-normally distributed against the one that they are simply normally distributed. These results are displayed in Table 4.1

These results suggest that the organism-level population distribution of eye-level CMZ population counts, as assayed by PCNA immunostaining, is better modelled log-normally than normally. This is true whether we test the primary per-section count measurements, or the estimated whole-annulus population, calculated as described in Section 10.1.4, even though this quantity is calculated using the normally-distributed lens diameter<sup>1</sup>. The most likely log-Normal representations of the CMZ populations are about two orders of magnitude more likely than the Normal alternative. Additionally, although normal models

---

<sup>1</sup>The apparent superiority of the normal model for lens diameter may be due to the paucity of data at later time points; as described in INSERT METHODS AUTOREF, lenses are difficult to retain in this histological context.

Table 4.1: Likelihood ratio comparison between normal and log-normal models of retinal population parameters

Parameter	$\mathcal{N}$ logLH	Log- $\mathcal{N}$ logLH	logLR
Sectional PCNA+ve	-285.611	<b>-283.214</b>	2.397
Lens diameter	<b>-203.102</b>	-203.854	-0.752
CMZ annular pop.(†)	-484.768	<b>-482.733</b>	2.035
RPE length	<b>-368.232</b>	-368.609	-0.377
CR thickness	<b>-240.858</b>	-241.032	-0.174
CR volume (†)	-712.256	<b>-711.766</b>	0.491

$\mathcal{N}$ : Normal distribution. logLH: logarithm of  $p(D|M)$ , the likelihood of the data given the model. logLR: logarithm of the likelihood ratio; positive ratios in favour of the log- $\mathcal{N}$  model. Superior likelihoods are bolded. †: Calculated quantities. Sectional PCNA+ve: population of PCNA-positive CMZ RPCs per  $14\mu\text{m}$  cryosection. CMZ annular pop: population of annular CMZ. RPE: retinal pigmented epithelium. CR: cellular retina.

are slightly favoured for describing the population-level distributions of RPE length and cellular retina thickness, the derived cellular retina volume quantity is better modeled by a log-Normal distribution. As the two calculated quantities will be the primary ones used in our inferences about population-level CMZ dynamics, we are most concerned with these measures; at this point, the log-Normal distribution looks to be more informative for both.

We have emphasized the danger of relying overmuch on the parameterisation of single most-likely model fits in Chapter 3, which is what the simple likelihood ratios above represent: the joint likelihood of the maximum a posteriori distribution fitted to the measurements taken from each age cohort. Since the choice of model describing the estimated annular population and retinal volume is critical to the success of later inferences, we used Galilean Monte Carlo-Nested Sampling (GMC-NS) to estimate the Bayesian evidence (the marginal probability of the data over all model parameterisations) for these hypotheses. Although it is very unlikely that GMC-NS will result in conclusions from the likelihood ratio, this simple test serves to prove the function of the `GMC_NS.jl` package, and demonstrate the inferential logic. Evidence estimates and ratios are presented in Table 4.2.

Table 4.2: Evidence favours log-normal models of retinal population parameters

Parameter	$\mathcal{N}$ logZ	Log- $\mathcal{N}$ logZ	logZR	$\sigma$ significance
CMZ annular pop.	$-744.649 \pm 0.012$	<b><math>-629.957 \pm 0.013</math></b>	$114.691 \pm 0.018$	6518.5
CR Volume	$-1287.372 \pm 0.018$	<b><math>-959.578 \pm 0.012</math></b>	$327.794 \pm 0.021$	15421.1

$\mathcal{N}$ : Normal distribution. logZ: logarithm of  $p(D)$ , the marginal likelihood of the data, or model evidence. logZR: evidence ratio; positive ratios in favour of the log- $\mathcal{N}$  model. Largest evidence values bolded. CR: cellular retina.

Unsurprisingly, full estimation of the Bayesian evidence for the Normal vs. log-Normal hypotheses for our calculated parameters produces the same basic story as the rough calculation of likelihood ratio from the single-fit MAP models. However, the numerical results reveal the sense in which model comparison based on these simpler methods can be misleading. There are approximately 115 orders of magnitude more evidence for the log-Normal model of interindividual variation in estimated CMZ annulus RPC

population over time. This result has greater than 6500 standard deviations of significance<sup>2</sup>. However, there are 327 orders of magnitude more evidence for the log-Normal model for the variability in the cellular retinal volume estimate, with over 15400 standard deviations of significance. Given the likelihood ratios alone, we might have suspected that the log-Normal model was more justified for modelling the population data than for the volume data. In fact, with the full estimation of the evidence, it becomes clear that the converse is true. In any case, we need not have any compunction about modelling population outcomes for these parameters with log-Normal distributions, as the evidence supplied by our observations plainly supports it.

If between-individual variation in retinal CMZ population and retinal volume are both well-modelled log-normally, the question of their independence immediately arises. It seems plausible that the size of the CMZ population would be roughly proportional to the overall volume of the retina, upon central specification, and establishment of the peripheral CMZ remnant. Moreover, since the growth of retinal volume over the life of the organism is driven primarily by the CMZ<sup>3</sup>, it seems likely that either the retinal growth rate is well correlated with the size of the CMZ. Since both of these questions bear upon the manner in which we model the CMZ, we performed Bayesian model selection by the so-called Empirical Bayes method for linear regression, which provides for direct estimation of the evidence for models consisting of linear equations of variables [Bis06].

Interestingly, we find that CMZ population and retinal volume estimates are better described by uncorrelated models at all in ages, with approximately 1-8 orders of magnitude of evidence in favour of the uncorrelated model at each age. These data are displayed in Table 4.3. It is interesting to note that the evidence in favour of non-correlation is weakest at 30dpf, which will be discussed below. Of particular importance are the data for 3dpf embryos, as we intend to seed model retinae with CMZ populations and volumes drawn from these distributions. The data for these animals are plotted in Figure 4.1. The general lack of correlation between CMZ population and retinal volume estimates may be due to the loss of information involved in the estimation calculations; on the other hand, it is more plausible that CMZ population should be associated with the rate of retinal volume growth rather than the volume of the retina itself. Because growth rate data are unavailable for single individuals, we cannot make this inference directly.

From these analyses, we conclude that organismal variability in the CMZ population and retinal volume estimates are best described by independent log-Normal distributions. Because log-Normal distributions are simply transformed Normal Gaussian distributions, we may model our uncertainty about their parameters with Normal-Gamma distributions over the mean and variance of the underlying Normal distribution of the log-Normal population model ("the underlying"). That is, our prior and posterior belief about the relative likelihood of values of the mean of the underlying may be modelled with a Normal distribution, and our beliefs about its variance with a Gamma, such that our joint uncertainty is the product of the two distributions. This is explained in more detail in ???. A useful analytic feature of the Normal-Gamma prior is that the marginal posterior distribution of the mean,

---

<sup>2</sup>The typical standard for a "discovery" in particle physics is  $>5\sigma$  [Lyo13]. Assuming Normally distributed error, the probability that the models were, in fact, equally good at 3 standard deviations of significance would be about a tenth of a percent (i.e. .001). At 10 standard deviations the figure is 7.62e-24. Assuming the sample is representative, we have nigh certainty about these results and no reason to pursue the matter further.

<sup>3</sup>One observes occasional proliferative clusters in the central retina throughout the life of the fish; these are typically ascribed to Müller glial repair processes. I am unaware of any estimate as to the relative contribution of these clusters vs. the CMZ. As we shall see, there is probably more turnover in the specified retina than previously believed. As a result, the relative contribution of these central clusters should probably be subject to statistical estimation; they may be more significant than mere lesion-repair sites.

Table 4.3: Evidence favours uncorrelated linear models of CMZ-population and retinal volume over time

Age (dpf)	Uncorrelated logZ	Correlated logZ	logZR
3.0	<b>-60.089</b>	-65.078	4.989
5.0	<b>-61.349</b>	-64.873	3.524
8.0	<b>-60.181</b>	-67.639	7.458
12.0	<b>-60.768</b>	-68.041	7.273
17.0	<b>-63.503</b>	-67.344	3.841
23.0	<b>-69.24</b>	-76.044	6.804
30.0	<b>-68.858</b>	-70.278	1.42
60.0	<b>-75.769</b>	-79.67	3.901
90.0	<b>-74.973</b>	-83.883	8.91
180.0	<b>-76.244</b>	-83.948	7.704
360.0	<b>-79.822</b>	-87.628	7.805

logZ: logarithm of  $p(D)$ , the marginal likelihood of the data, or model evidence. logZR: evidence ratio; positive ratios in favour of the uncorrelated model. Largest evidence values bolded.

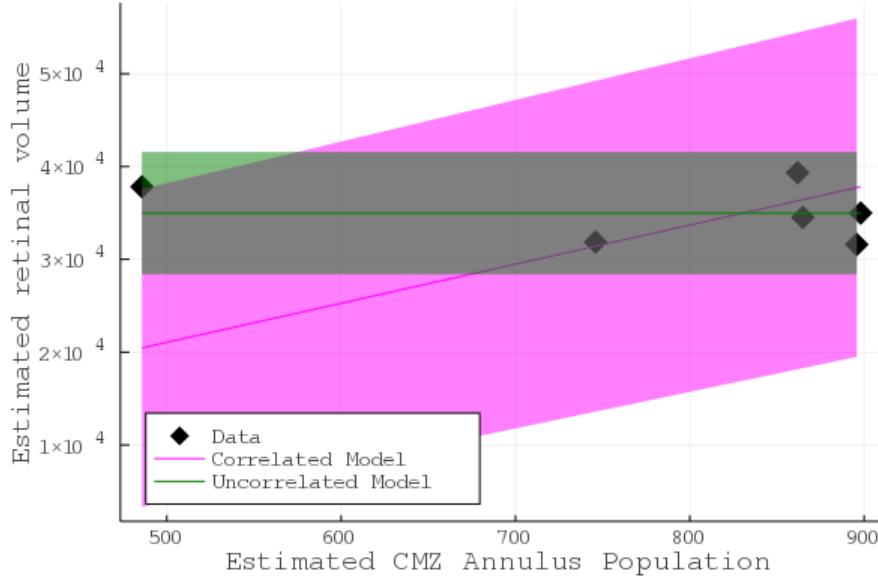


Figure 4.1: CMZ population and retinal volume estimates are uncorrelated at 3dpf  
Individual CMZ population estimate vs retinal volume estimate for 3dpf animals. Uncorrelated and correlated linear models of these variables are plotted as the mean $\pm$ 95% CI of the predictive distribution of the fitted model.

assuming an uninformative (ignorance) prior, is a location-scaled T distribution; this is so because the weighted sum of an infinite series of Normal distributions (i.e. the likelihood-weighted sum of all the Normal distributions that could underly the log-Normal models), is a T distribution. T distributions are notably more resistent to outlier distortion than Normal distributions themselves are, and may be estimated with as few as two observations, making them highly flexible. Most of the descriptive statistics in the next section, therefore, calculate the credible interval for the posterior mean of the underlying by T distributions (with the correct change of variables by exponential transformation to produce the

features of the correct log-Normal distribution). Unfortunately, differences of T distributions are not, themselves, necessarily T-distributed, so we have relied on Monte Carlo estimation of rates of change of the these posterior means over time. With our log-Normal models selected, and the descriptive tool of the posterior mean T distribution in hand, we turn now to a survey of the zebrafish CMZ in the first year of life.

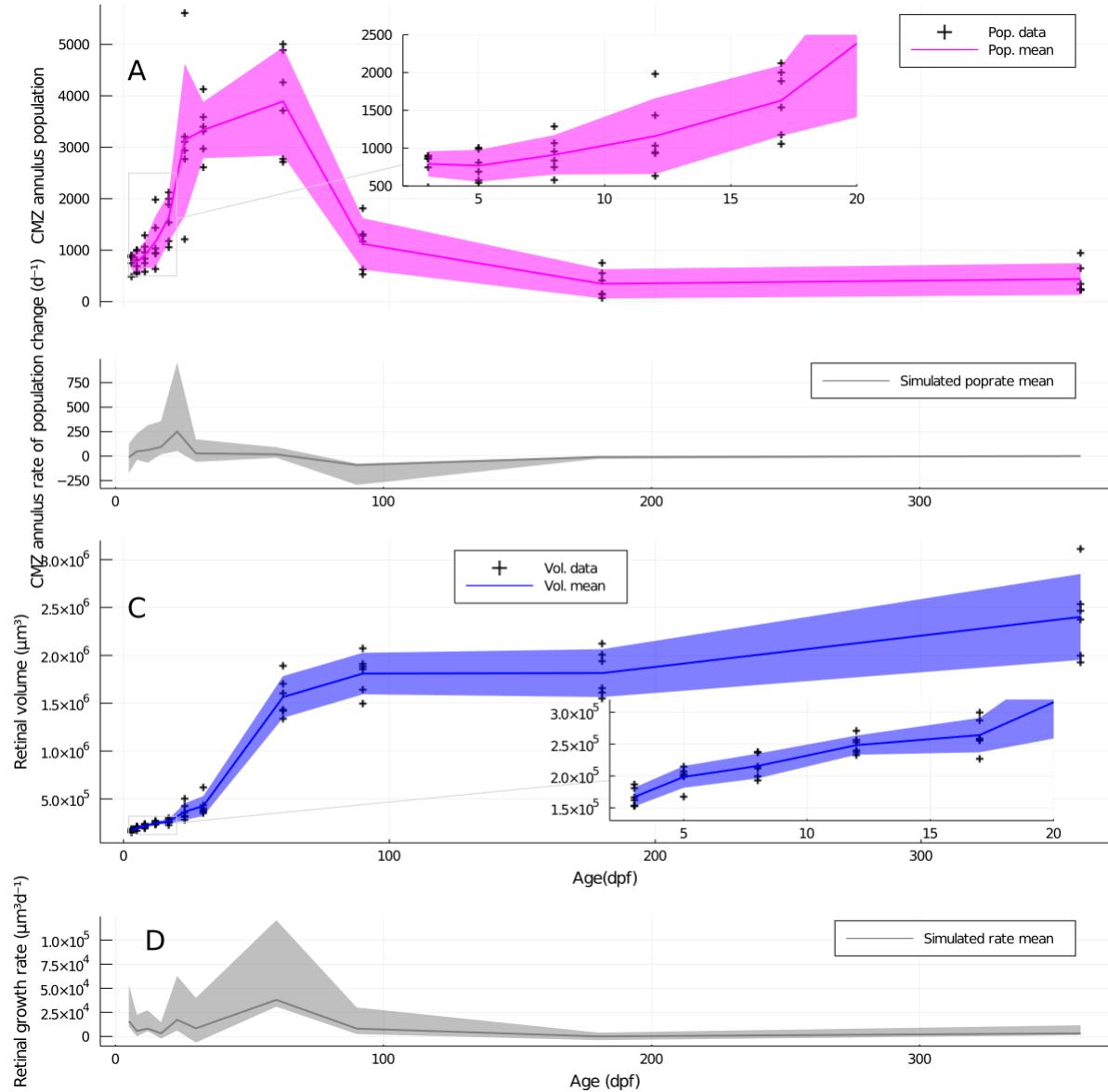
## 4.2 Survey of CMZ population and gross retinal contribution

If it is true that the majority of zebrafish retinogenesis occurs postembryonically, and that models trained on embryonic data do not describe this period well, what characterises this CMZ-driven phase of retinogenesis? We begin by presenting our estimates of individual CMZ annulus population and retinal volume over the first year of life, in Figure 4.2, panels A and C.

An initial, relatively quiescent period in the population history of the CMZ can be inferred from the lack of growth observed in the first week of life, as well as the observations presented in ??, which demonstrate that CMZ RPCs are proliferating too slowly to be labelled by a day's pulse of a thymidine analogue at 5dpf in the wild-type and heterozygote siblings of npat mutants. Interestingly, we have good confidence that retinal volume continues to grow over this time; 99.56% of the marginal posterior mass of the mean 5dpf estimate is above the 3dpf mean, suggesting that this proliferative pause is too short to appear in the retinal volume data.

In any case, the first two to three weeks of life (magnified in lens insets in panels A and C) appear to mark a relatively slow build in both estimated CMZ annulus population and retinal volume compared to the explosion which follows immediately thereafter. While zebrafish are better staged by size than age [PEM<sup>+</sup>09], the onset of this explosive growth seems to come somewhat earlier than the typical metamorphic transition from larval to juvenile stages at 45dpf [SH14]. This raises the question of the manner in which the CMZ contributes to the retina during the critical period of exponential growth of the organism, between about 45 and 90dpf. It is plausible, for instance, that the growth of the CMZ population mainly reflects the increased output of stem cells, with RPCs specifying at some steady rate, or that the CMZ builds itself up for a wave of specification somewhat later, similarly to the sequence of events in the embryonic central retina.

In order to get a better sense of this timing, we simulated the mean daily rate of CMZ annulus population and retinal volume change, by performing Monte Carlo difference operations between samples from the marginal posterior means of subsequent timepoints. These simulated mean rates and their associated confidence intervals are plotted in Figure 4.2 panels B and D. These show the basic time-structure of the phenomenon; the large increase in simulated daily CMZ population growth rate occurs before the large increase in retinal volume, but the CMZ population estimate does not begin to drop off until 60dpf, by which time the majority of volumetric growth is complete. This seems to substantiate some combination of the scenarios outlined above: there is an early buildup of CMZ population around 18-30dpf, before CMZ RPCs begin to make their primary contribution to the retina between 30-60dpf, which is characterised by much slower population growth and more rapid volumetric growth, implying steady and elevated specification of RPCs. The greater uncertainty associated with our population estimates, relative to their magnitude, results in correspondingly less certainty about the size of the population growth rate spike. For instance, we calculate a >99% probability that the mean retinal volumetric rate growth peak at 60 dpf is at least 6-fold greater than the mean at 30dpf. By contrast,



**Figure 4.2: Population and activity of the CMZ over the first year of *D. rerio* life**  
 Panel A: Marginal posterior mean CMZ annulus population. Panel B: Marginal posterior mean retinal volume estimate. Insets in Panels A & B display data from 3-17 dpf. Panel C: Marginal posterior mean of the proliferative index of the CMZ annulus, assayed by specified retinal neurons with incorporated thymidine from an 8hr pulse at the indicated ages. Panel D: Mean daily rate of volumetric increase of the neural retina, calculated as the difference in volumes between two ages over the number of elapsed days. All means are displayed in a band representing the  $\pm 95\%$  credible interval for the marginal posterior distribution of the mean.

only about 97% of the posterior mean density of the population growth rate at the 23d peak ( $230.2 \frac{cells}{d}$ ) is greater than the mean at 3dpf, which is a small negative value ( $-5.5 \frac{cells}{d}$ ).

Subsequent to the bulk of the CMZ buildup and contribution to the retina, the population of the CMZ declines at about 39% the rate of its peak ascent, with an estimated  $-90.1 \frac{cells}{d}$  by 90dpf. This

process thins out the CMZ population to below its size immediately after embryogenesis, spread out over a much larger peripheral annulus, for a much less dense mature CMZ. Interestingly, the 95% CI on the marginal posterior mean of estimated retinal volume at 180 dpf ( $1.10 \pm 0.04e9 \mu\text{m}^3$ ) completely encompasses the 95% CI on volume at 90dpf ( $1.11 \pm 0.03e9 \mu\text{m}^3$ ), while we assess a 99.6% probability of the 360dpf mean ( $1.73 \pm 0.10e9 \mu\text{m}^3$ ) being greater than 180dpf. This suggests a possible second period of quiescence from approximately 90-180dpf, followed by steady contribution to the retina without a buildup in the CMZ population subsequently.

Given this rough description, it seems obvious that the ontogeny of the CMZ as a stem cell niche is characterised by different phases of activity, with different rates of proliferation and specification. It is not immediately clear what sort of periodization is justified by the data. It seems plausible that as few as two phases could explain the data well enough: an initial phase of logarithmic growth, with short cell cycle time and lower exit rate of RPCs from the CMZ into the specified neural retina, followed by a second phase of decay with longer cycle time and higher exit rate. On the other hand, perhaps some of the data features noted above justify a more bespoke model that captures, for instance, the initial quiescent period, or the post-180dpf growth of the retina. Because this is straightforwardly a question of how much model structure is justified by our data, we may address it as a model selection problem, using the system of Bayesian inference provided by nested sampling, and it is to this we now turn.

### 4.3 Periodization of postembryonic CMZ activity by Galilean Monte Carlo Nested Sampling

To perform our model selection task, we wish to simulate the time-evolution of CMZ population and retinal volume. This requires us to supply initial values for the size of the simulated CMZ populations and the volumes of the simulated retinae they are associated with. Given the findings presented in Figure 4.1, we are justified in initializing CMZ population and retinal volume by independent samples from the log-Normal models of their interindividual variability at 3dpf, at the end of embryogenesis and the beginning of CMZ-driven retinal growth. In order to produce new, simulated values of CMZ population and retinal volume, we apply a system of difference equations as follows, where  $pop_n$  is the population at  $n$  dpf,  $CT$  is the mean cell cycle time of the population in hours, and  $\epsilon$  is the proportion of the population at time  $n - 1$  that exits cycle and contributes to the volume of the specified neural retina, and  $\mu_{cv}$  is the mean volume per cell contributed to the retina in  $\mu\text{m}^3$ :

$$p_n = p_{n-1} \cdot 2^{\frac{24}{CT}} - p_{n-1} \cdot \epsilon \quad (4.1)$$

$$v_n = v_{n-1} + p_{n-1} \cdot \epsilon \cdot \mu_{cv} \quad (4.2)$$

A model "phase" can then be defined by the  $CT$  and  $exit$  parameters it applies to update the population and volume over its length. The full parameterisation of a model with  $p$  phases is given by  $p$  pairs of  $CT$  and  $exit$  values,  $p - 1$  phase transition times, and  $\mu_{cv}$ , which is taken to apply equally to all phases. Given an initial population and volume sampled from the log-Normal models of their interindividual distributions, Equation 4.1 and Equation 4.2 may be applied to these values difference equations can be applied to produce simulated sample values at the times actually observed. Many such samples obtained by Monte Carlo can be used to estimate log-Normal distributions for the model

parameters, which can be used to score the model against observations. By defining prior distributions over the model parameters, we may sample from the prior to initialize a model ensemble. The ensemble can then be compressed by nested sampling, moving each model-particle over the parameter space by Galilean Monte Carlo, as described in ???. Using the typical procedures applied in nested sampling [Ski06], we calculated the Bayesian evidence for each of 1, 2, and 3-phase models. These results are presented in, with output from maximum a posteriori parameterizations for each model plotted against observations in.

## 4.4 Characterisation of asymmetrical CMZ population and retinal contribution dynamics

In the course of the preceding investigations, it became apparent that the population asymmetry mentioned in Chapter 3 was not a static phenomenon, with the dorsal lobe of the CMZ annulus being consistently more populous than the ventral lobe, as generally implied by the sources covered in Chapter 1. Rather, both the extent and orientation of asymmetry seem to evolve over time. Sectional population totals for the dorsal and ventral CMZ are presented in Figure 4.3, Panel A, alongside the related intra-individual asymmetry ratio in Panel B. The initially pronounced dorsal population and reduced ventral population both seem to go through the overall boom-bust progression of CMZ population, but their relative proportion within individuals reverses itself over the period from 17-90dpf.

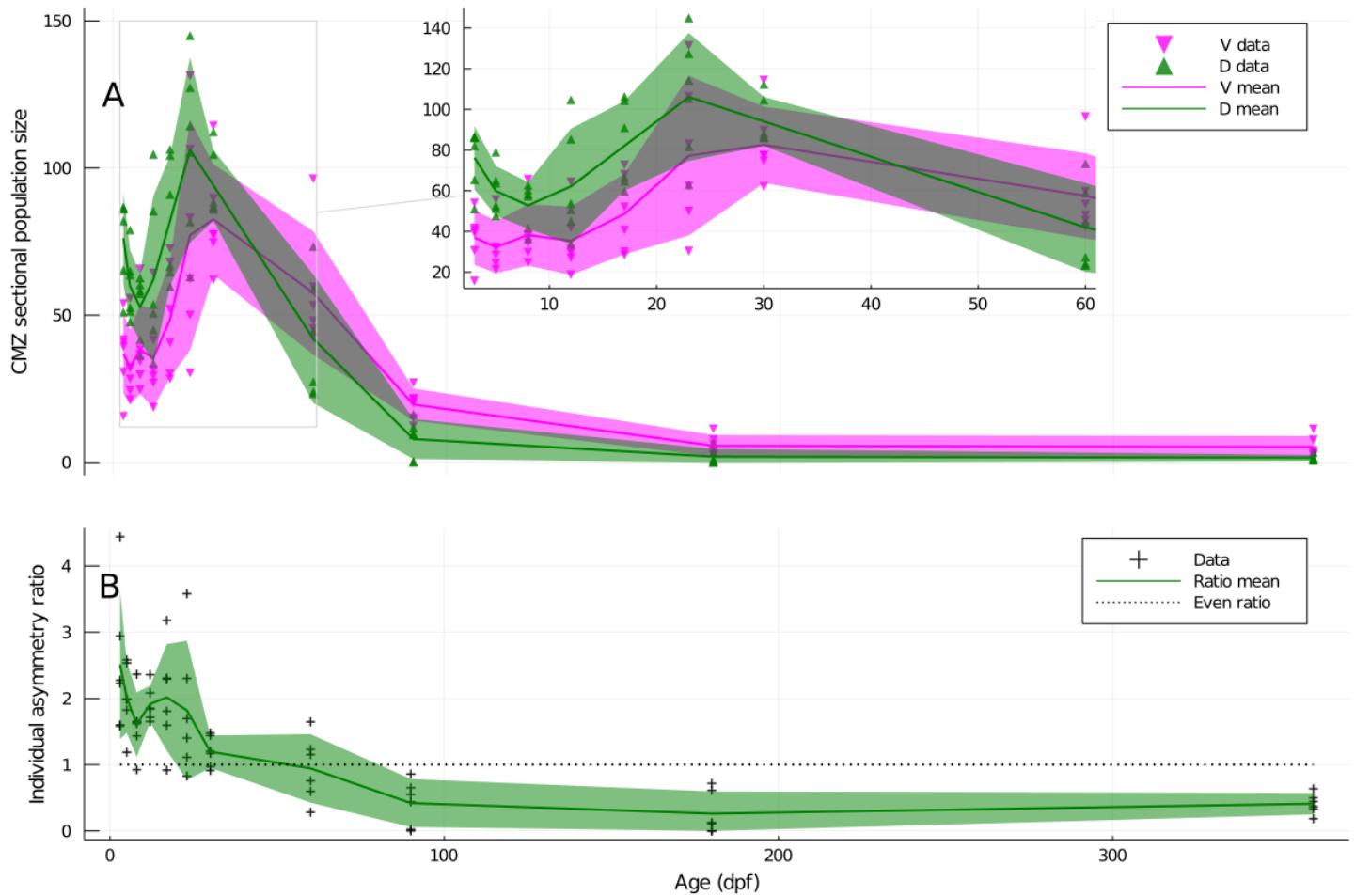
Inspected closely, these data provide a possible rationale for the reversal of asymmetry in the proliferative dynamics of the niche itself: the sectional (or "slice") population of the dorsal CMZ is increasing beyond its postembryonic minimum by 12dpf, while the ventral CMZ takes until 17dpf to exhibit a noticeable increase in size; moreover, the peak dorsal population is achieved by 23dpf, whilst ventrally the peak is only achieved at 30dpf. This strongly suggests that the dorsal and ventral CMZ populations undergo similar, time-shifted processes of proliferation from different starting populations. If this is so, an explanation for this time-shifted phenomenon could have fundamental relevance to predicting and controlling the proliferative behaviour of peripheral RPCs and stem cells.

As a crude way of assessing the possibility of variable cell cycle attributes across the dorsal-ventral axis, we used the Empirical Bayes approach to estimate the evidence for separate Nowakowski-style [NLM89] linear models for the dorsal and ventral CMZ in a cumulative thymidine analogue labelling experiment at 3dpf, against a model for both subpopulations combined. While this model is inadequate for reasons described in ??, it can serve as a guide as to whether further inferential calculations are likely to be productive here. These results are summarized in Table 4.4, with the relevant linear regressions displayed in ??.

Table 4.4: Evidence favours whole-CMZ linear cycle models over separate D/V models

Model	Implied $T_c$ (hr)	Implied $T_s$ (hr)	logZ
Dorsal	$14.7 \pm 1.6$	$1.38 \pm 0.76$	7.778
Ventral	$14.0 \pm 1.2$	$0.8 \pm 0.58$	15.202
Combined	$14.6 \pm 1.1$	$1.25 \pm 0.53$	<b>26.165</b>

$T_c$ : calculated cell cycle time.  $T_s$ : calculated s-phase length. logZ: logarithm of p(D), the marginal likelihood of the data, or model evidence. Largest evidence values bolded.



**Figure 4.3: Developmental progression of dorso-ventral population asymmetry in the CMZ.** Marginal posterior distribution of mean dorsal (D) and ventral (V) population size in  $14\mu\text{m}$  coronal cryosections (panel A) or intra-individual D/V count asymmetry ratio (panel B),  $\pm 95\%$  credible interval,  $n=5$  animals per age. Data points represent mean counts from three central sections of an experimental animal's eye.

The estimated log evidence for the combined model, treating dorsal and ventral CMZ as a single, uniformly proliferative population, is 26.165. Compared to the joint log evidence for two models taking dorsal and ventral CMZs as separate homogenously proliferative populations, 22.980, there are approximately 3.19 orders of magnitude of evidence in favour of the combined model. There is no evidence, on this basis, for differing cell cycle characteristics across the D/V retinal axis of asymmetry at 3dpf. 3dpf was selected here for pragmatic reasons; the D/V asymmetry is large and the dorsal population seems to be, perhaps, proliferating more slowly, given its decline relative to the ventral population. It is possible that asymmetric cell cycle times would be better detected during the candidate "time-shift" period, around 15dpf. It is, in any case, plain from the implausibly short calculated cycle times that this model is inadequate.

## 4.5 Investigating CMZ RPC lineage outcomes

By labelling CMZ RPCs with the thymidine analogue EdU in an 8 hour pulse at 3, 23, and 90 dpf, followed by histochemical analysis for known zebrafish retinal neural lineage markers, we investigated the possibility that RPC lineage outcomes change over the life of the organism. This hypothesis is of particular interest, as differences in the mosaic organisation of embryonically-contributed central retina and CMZ-contributed peripheral retina remain unexplained [ABS<sup>+</sup>10]. We used antibodies raised against Pax6 and Isl2b to mark retinal ganglion cells (RGCs) of the ganglion cell layer and amacrine cells of the inner nuclear layer. Anti-glutamine synthetase (GS) and anti-PKC $\beta$  were used to mark Müller glia (MG) and bipolar cell (BPC) populations of the INL. The unique flattened nuclear morphology of horizontal neurons was used to identify them. Lastly, the antibody Zpr1, directed against an unknown antigen present in photoreceptors with double cone morphology, was used to mark these cells.

Observations were collected in "staining groups", which combined histological markers; representative confocal micrographs from this study in animals pulsed at 23dpf are displayed in Figure 4.4, while data from all ages are plotted in Figure 4.5. It is worth noting that the relative position of the CMZ-contributed cohort is very different in older animals, with 7 days of chase time being just enough for the majority of the 90dpf cohort to be reliably located within the specified neural retina, as depicted in supplementary ???. Because we suspect that CMZ-contributed retinal cohorts are subject to a process of attrition, and that this turnover might be higher near the CMZ, as documented in Section 4.6, if this hypothetical turnover process has differential effects on specified retinal neurons of different lineages, results may not be directly comparable between ages. With that caveat, we proceed to the lineage data.

These data take the form of fractions of the presumptive CMZ-contributed, thymidine-labelled cohort entering each of the three cellular layers (panel A of Figure 4.5), or the subfraction of the cohort within a given layer expressing a particular cellular marker (Panels B-I). While some variability is apparent in all of the measurements, it is unclear whether it is well-described as time-dependent in most cases. In order to address the question of whether lineage outcomes differ over time, we needed to assess the joint evidence for separate models of each measurement at each age, against the evidence for a single model of the measurement for all of the assessed ages. Because it is not immediately obvious that these fractional measurements are better described log-Normally as the underlying population counts are, we first assess the joint evidence for Normal and log-Normal models of the data, both separated by age and as age-marginalised measurement sets.

## 4.6 Microglial apoptotic fate of *D. rerio* retinal neurons & functional significance of CMZ activity

Recently, extensive neural death has been reported in older zebrafish retinae, described as a "neurodegenerative pathology" and suggested as a model of age-related neurodegeneration [VGV<sup>+</sup>18]. In the course of our thymidine analogue pulse-chase studies, we noted that it often appeared that CMZ-contributed cohorts had been "thinned out" noticeably only a month or two after their entry into the neural retina, even in juveniles of 30-90 days of age. If neural retinal turnover is a general phenomenon throughout the life of the organism, this has fundamental implications for the view that this phenomenon should be treated as pathological, rather than constitutive. However, the thinning phenomenon could be explained by processes involved in the changing morphology and geometry of the neural retina during this period.

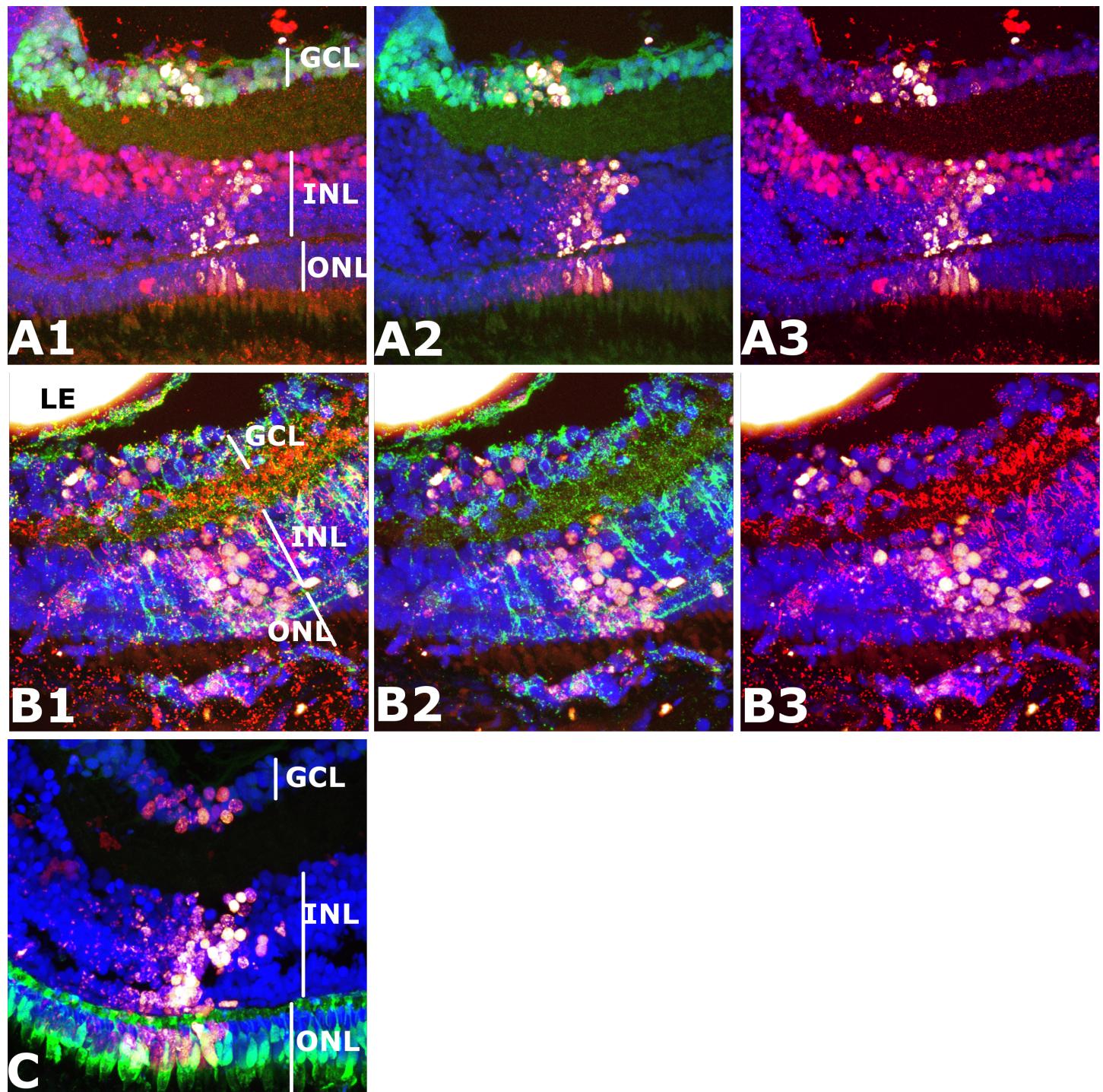
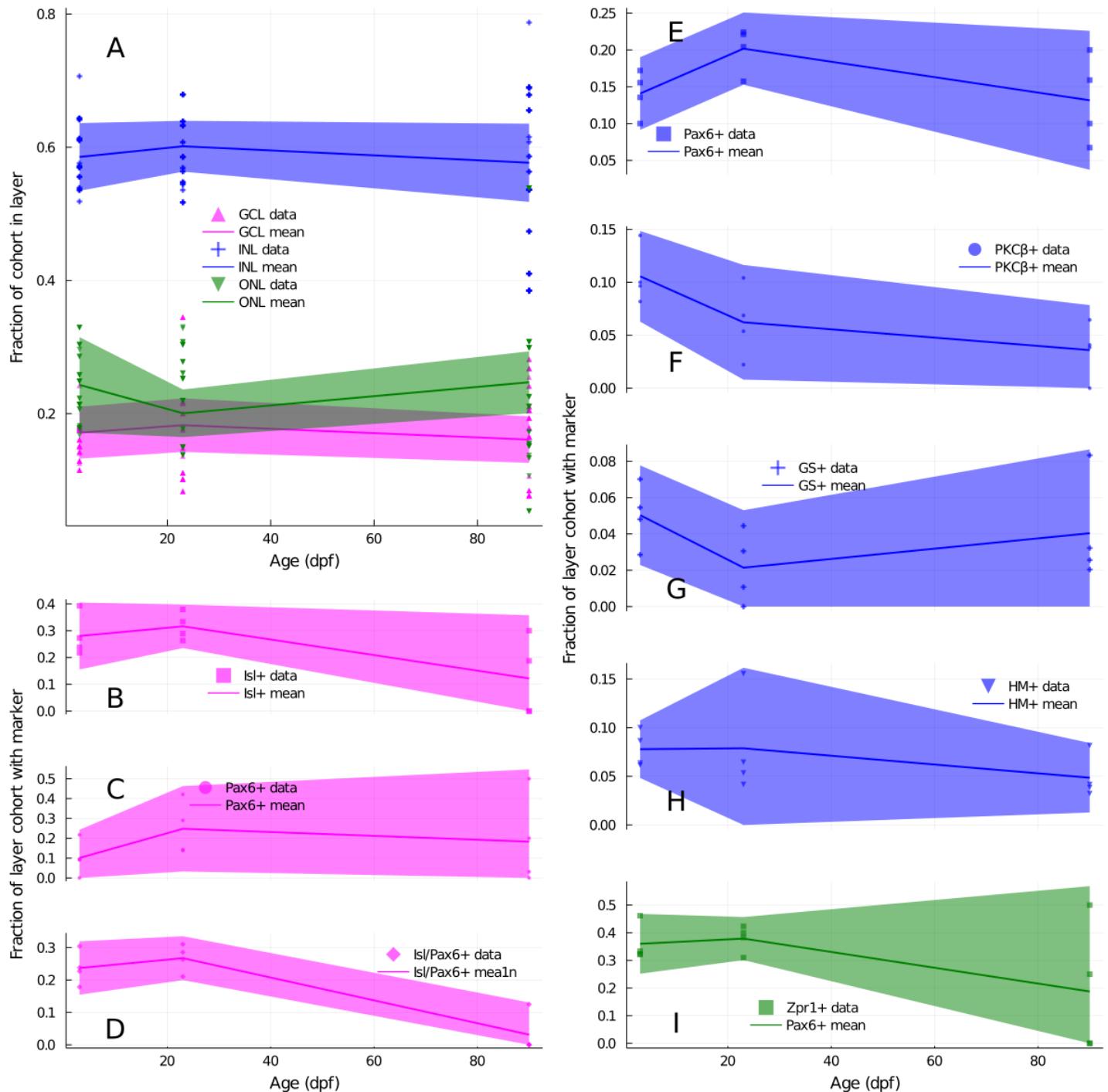


Figure 4.4: Representative 23dpf lineage marker confocal micrographs  
Panel A1: RGC/Amacrine staining group. A2: Isl2b channel. A3: Pax6 channel.  
Panel B1: MG/BPC staining group. B2: PKC $\beta$  channel. B3: GS channel.  
Panel C: Double cone staining group. Zpr1 channel.  
GCL: Ganglion cell layer. INL: Inner nuclear layer. ONL: Outer nuclear layer.



**Figure 4.5: CMZ contributions to the neural retina over time by layer and lineage marker**  
 Marginal posterior distribution of mean dorsal (D) and ventral (V) population size in  $14\mu\text{m}$  coronal cryosections (panel A) or intra-individual D/V count asymmetry ratio (panel B),  $\pm 95\%$  credible interval,  $n=5$  animals per age. Data points represent mean counts from three central sections of an experimental animal's eye.

In particular, the neural retina thickens noticeably over this time period, as displayed in supplementary ???. Although this increase is due in large part to the lengthening of photoreceptor outer segments, the inner nuclear layer is also significantly thickened. It is plausible that, for instance, labelled CMZ cohorts are "invaded" by unlabelled neighbours during this process, giving the appearance of "thinning" without its quantitative reality.

In order to investigate this phenomenon, we administered 24hr pulses of EdU to 1dpf embryo, and followed with a 24 hr pulses of BrdU at 23dpf to mark a CMZ-contributed cohort near the height of its activity. By taking both coronal and transverse sections through animals at 30, 60, and 90 dpf,we sampled these cohorts from both morphological axes of the retina and counted labelled sectional totals.

#### 4.7 Toward cell-based computational models of the CMZ

# Chapter 5

## Mutant npat results in nucleosome positioning defects in *D. rerio* CMZ progenitors, preventing specification but not proliferation

### 5.1 Introduction

The zebrafish (*D. rerio*) circumferential marginal zone (CMZ), located in the retinal periphery, contains the retinal stem cells and progenitors responsible for the lifelong retinal neurogenesis observed in this cyprinid. Analogous to CMZs in other model organisms, such as *X. laevis* [PKVH98], it has been of particular interest to us since the discovery of quiescent stem cells at the mammalian retinal periphery [Tro00], as an understanding of the molecular mechanisms regulating this proliferative zone may shed light on whether these mammalian cells might be harnessed for the purpose of regenerative retinal medicine. While significant progress has been made in this direction [RBPP06], molecular lesions in a plethora of zebrafish mutants displaying defects in CMZ development and activity remain largely uncharacterised. We examine here one such microphthalmic line identified in an ENU screen, rys [WSM<sup>+</sup>05], characterised by Wehman et al. as a Class IIA CMZ mutant, with an apparently paradoxically enlarged CMZ.

Mapping revealed the causative rys mutation lay in the zebrafish npat gene, the nuclear protein associated with the ataxia-telangiectasia locus in mammals [IYS<sup>+</sup>96]. Although npat is heretofore uncharacterised in zebrafish, its mammalian homologues, human NPAT and mouse Npat, have been extensively examined. These studies have demonstrated that NPAT plays a critical role in coordinating events associated with the G1/S phase transition in proliferating cells [YWNH03]. S-phase entry requires tight co-ordination between the onset of genomic DNA synthesis and histone production, in order to achieve normal chromatin packaging and assembly. NPAT, found in the nucleus [STN<sup>+</sup>02] and localised, in a cell-cycle dependent manner, to histone locus bodies [GDL<sup>+</sup>09], induces S-phase entry [ZDI<sup>+</sup>98] and activates replication-dependent histone gene transcription by direct interaction with histone gene clusters [ZKL<sup>+</sup>00] in association with histone nuclear factor P (HiNF-P) [MXM<sup>+</sup>03]. The protein's effects

on S-phase entry and histone transcription are associated with distinct domains at the C-terminus and N-terminus, respectively [WJH03]. NPAT is also known to associate with the histone acetyltransferase CBP/p300 [WIEO04] and directs histone acetylation by this enzyme [HYSL11].

NPAT is known to be a component of the E2F transcriptional program [GBB<sup>+</sup>03] and a substrate of cyclin E/CDK2 [ZDI<sup>+</sup>98], although E2F-independent activation by cyclin D2/CDK4 in human ES cells has also been described [BGL<sup>+</sup>10]. The expression of NPAT protein peaks at the G1/S boundary [ZDI<sup>+</sup>98], as does its phosphorylation, which promotes its transcriptional activation of replication-dependent H2B [Ma00] and H4 [MGv<sup>+</sup>09] genes, while its effect on low, basal levels of H4 transcription is phosphorylation-independent [YWNH03]. Of particular interest, NPAT has recently been found to be required for CDK9 recruitment to replication-dependent histone genes (Pirngruber and Johnsen, 2010); CDK9 and monoubiquitinated H2B are essential for proper 3' end processing of stem-loop histone transcripts [PSS<sup>+</sup>09]. NPAT and HiNF-P have also been found associated with the U7 snRNP complexes that perform this function [GDL<sup>+</sup>09]. The replication-dependent activities of NPAT are thought to be terminated by WEE1 phosphorylation of H2B, which excludes NPAT from histone clusters [MFKM12].

All of these studies have been conducted in tissue culture contexts, perhaps due to the challenges associated with studying this critical protein *in vivo*; mouse embryos with provirally inactivated Npat arrest at the 8-cell stage, for instance [DFWV<sup>+</sup>97]. The availability of rys, a zebrafish npat mutant which develops well into the larval stage, is therefore of considerable interest, as it allows for the study of npat's function within complete tissues. We demonstrate here that npat is critical for the normal proliferation and differentiation of CMZ neural progenitors.

The altered npat regulation of histone transcription and cell cycle progression in rys results in a dramatically shortened S-phase and an increase in the abundance of core histone transcripts, including polyadenylated transcripts, and their associated proteins. These effects are associated with disorganised nucleosome positioning and altered protein expression and progenitor identity, failure of rys CMZ cells to contribute to the neural retina, and subsequent cell death, suggesting that npat may play a critical role in the coordination of genomic events required for normal execution of differentiation programmes.

## 5.2 Results

### 5.2.1 The microphthalmic zebrafish line rys is an npat mutant

In order to determine the causative mutation responsible for the rys phenotype, we performed linkage mapping to identify candidate genes, followed by PCR analysis of the transcript products of these candidates. This study revealed a single G>A transition at position 24862961 on chromosome 15 (NC\_007126.5, Zv9), annotated as the first base of intron 9 in the zebrafish npat gene, in a canonical GU splice donor site. The predicted protein sequence expressed from the mutant transcript is truncated by a stop codon at residue 283, which precludes the translation of predicted phosphorylation sites and nuclear localisation signals cognate to those identified in human NPAT (Figure 1D) [Ma00, STN<sup>+</sup>02].

Because *D. rerio* is a teleost known to have undergone genome duplication in an ancestral clade, we used the Synteny Database tool to identify any possible duplicates; plausibly, a duplication in zebrafish npat could explain the lessened severity of the mutant phenotype when compared to mammalian proviral inactivants. The results of this analysis are displayed in supplementary ???. While npat is in the midst of a region which appears to be duplicated on chromosomes 5 and 15, relative to the unduplicated

homologous syntenic run on *H. sapiens* chromosome 11, it is not, itself, duplicated.

We observed that the putative causal *rys* transition mutation results in retention of this intron in npat transcript in *rys* mutants, with 6 days post fertilisation (dpf) heterozygotes expressing both wild type and mutant transcripts and homozygotes expressing solely mutant transcript. We performed in situ hybridisation using a probe directed to the wild type npat transcript to confirm that the gene is indeed expressed in wild type CMZs; we found that npat expression is progressively restricted to the CMZ from 4 to 6 dpf in wild-type fish (Figure 1C).

### 5.2.2 *rys* CMZ cells display altered nuclear morphology and progenitor marker expression

We next set out to examine the *rys* CMZ phenotype in more detail. *Rys* has previously been described as having an enlarged CMZ [WSM<sup>+05</sup>], but whether this is a consequence of an enlarged proliferative population, altered cellular morphology, or some combination of both, remained unclear. We first, therefore, quantitated the number of cells present in 5dpf *rys* CMZs relative to their siblings. PCNA immunostaining was selected as a marker of proliferating cells in the CMZ, as it is present throughout the cell cycle in proliferating retinal neurons. It was observed that *rys* CMZs contain similar numbers of PCNA-positive cells per central coronal cryosection relative to their siblings ( $n=5$  sib eyes,  $n=6$  *rys* eyes) (Figure 2A) [OR It was observed that *rys* CMZs contain 15% (?) more PCNA-positive cells per central coronal cryosection relative to their siblings ( $n=?$ ) (Figure 2A) – Mariam’s data]. As most of the volume of the CMZ consists of closely-packed nuclei, we examined the Hoechst-stained nuclear volume and sphericity of PCNA-positive cells in the CMZ; this analysis revealed that *rys* CMZ nuclei have 17% greater volume than those present in their siblings, and are significantly less spherical ( $n=449$  sib nuclei,  $n=547$  *rys* nuclei; both measures  $p<.01$ ) (Figure 2B). Suspecting that *rys* CMZ cells may display altered progenitor identity, we examined pax6a immunostaining of this population, which is normally restricted to putative progenitors in the peripheral and middle CMZ, as well as the retinal ganglion cell layer [RBBP06]. The Pax6-stained region was enlarged in *rys* CMZs relative to their siblings (Figure 2C). As *vsx2* is also known to be a marker of retinal progenitor cells in the CMZ [RBBP06], we also generated a transgenic *vsx2::eGFP* *rys* line using a fragment of the zebrafish *vsx2* promoter which drives eGFP expression in the utmost retinal periphery in siblings. Mutant fish from this line displayed substantially expanded eGFP expression in the CMZ (Figure 2D).

### 5.2.3 Proliferating cells in *rys* embryos display altered cell cycle parameters; those in the CMZ have a dramatically truncated S-phase

As NPAT has been extensively implicated in proper coordination of events required for S-phase entry and cell cycle progression [GBB<sup>+03</sup>, YWNH03], we hypothesized that *rys* CMZ cells would display defective cell cycle progression, such as cell cycle arrest. In order to survey the effects of the *rys* mutation on proliferative cells throughout the developing zebrafish embryo, we performed a DNA-content/thymidine analogue (EdU) fluorescence activated cell sorting (FACS) analysis on dissociated cells from 3dpf *rys* and sibling embryos. By staining for PCNA and gating out PCNA negative cells, we were able to restrict our FACS analysis to the proliferative population of these animals. Doing so revealed [significant/no significant changes] in the cell cycle parameters of the proliferative population of *rys* embryos as a whole compared to their siblings (Figure 3A). We subsequently narrowed our focus solely to the proliferative

population in the CMZ by conducting cumulative EdU immunohistochemical labelling of PCNA-positive CMZ cells, and subsequently quantitating the number of EdU-labelled cells over time. As this entire population is definitionally proliferative, we assumed as our model a single asynchronously dividing population as previously described [NLM89]. This model allows the determination of total cell cycle length ( $T_c$ ) from the slope of the proportion of cells labeled vs. time regression:

#### **5.2.4 rys embryos express higher levels of core histone transcripts, particularly polyadenylated transcripts, and histone proteins**

Since NPAT is known to regulate histone transcription and is critical for coordinating the correct expression of replication-dependent histone transcripts required to package genomic DNA during S-phase [ZKL<sup>+00</sup>], we hypothesized that the altered nuclear morphology and truncated S-phase we observed in proliferating rys CMZ cells may be a consequence of perturbed regulation of histone transcription. To test this, we performed qPCR on random-hexamer-primed cDNAs produced from 6 and 8dpf wild type, sibling, and rys embryo mRNA extracts. These qPCR assays were performed using degenerate primers directed toward all members of the zebrafish core histone gene families H2A, H2B, H3 and H4. In 6dpf fish, we observed statistically significant increases in the levels of H2A, H2B, and H3 family transcripts in rys cDNA compared to wild type (Figure 4A). By 8dpf, the levels of all examined histone classes were significantly elevated in rys cDNA relative to both siblings and wild-type; siblings in turn displayed significantly elevated levels of all examined histone classes relative to wild-type (Figure 4B). As a role for NPAT in 3' end processing of histone transcripts has recently been identified [PJ10], we also set out to determine whether 3' end processing of histone transcripts is altered in rys. By repeating the above-described qPCR assays on oligo-dT-primed cDNAs, we were able to examine the population of polyadenylated histone transcripts in isolation. We found that in 6dpf embryos, rys had significantly increased levels of polyadenylated H2A and H2B transcripts relative to both sibling and wild-type embryos (Figure 4C). In 8dpf embryos, rys had further exaggerated levels of polyadenylated H2A and H2B transcripts, as well as significantly elevated polyadenylated H4 transcript (Figure 4D). We also examined the expression of histone proteins and determined effects on protein expression(Figure 4E).

#### **5.2.5 rys embryos display altered nucleosome positioning and chromatin organisation**

Our observation of perturbed histone expression in rys embryos led us to suspect that the aberrant nuclear morphology observed in rys CMZ progenitors arises from disrupted chromatin organisation, which is a possible consequence of altering the dynamic composition of the histone pool available for nucleosome formation during cell cycle. We therefore sought to characterise nucleosome positioning in rys and wild-type siblings by micrococcal nuclease (MNase) digestion of pooled genomic DNA (gDNA). Nucleosome-protected fragments from the MNase digests were sequenced and positions called as described previously.

We first sought to characterise the genomic disposition of nucleosome positions within wild-type siblings and rys. We found nucleosome positions distributed relatively evenly over sib genomic material, with only tiny deviations from a neutral assumption of a uniform distribution of the total position number over the full length of the genome, as displayed in Fig. 5.1, panel A. Rys chromatin displays minor differences from the proportions of positions found in each sib scaffold (panel B). Sib nucleosome positions are differentially occupied across scaffolds, with Chr 10 and scaffolds not yet mapped to chromosomes

(NC) being notably more occupied than expected from scaffold length alone (panel C). Similarly to the distribution of positions, rys chromatin is somewhat more evenly occupied than sibs; scaffolds with positions that are more heavily occupied in sibs tend to be depleted in rys and vice versa (panel D).

By mapping the called nucleosome positions in rys to those in sibs, we observed that there is a subset of sib positions which are never found to be occupied in rys, and likewise a subset of rys positions which are unique to the mutants (Fig 5.2). Intriguingly, sib positions which are not observed in rys are compensated for quite evenly across the genome by new positions gained in the mutants, with a small excess of new rys positions on most chromosomes.

This result suggested that a particular subpopulation of wild type nucleosomes are mislocalised in rys. If the pool of available histones in proliferating rys progenitors is substantially different from that of siblings, rys nucleosomes forming from this pool may have altered physico-chemical interactions with DNA, which could result in nucleosomes with unusual sequence preferences. If so, this would explain the disappearance of a subset of positions in rys and the appearance of a new, similarly sized subset of positions (the ‘differential set’).

To formally address this hypothesis, we calculated the Bayesian evidence ratio for separate emission processes for sib and rys position sequences against a combined process for both sets of sequences. If the molecular process resulting in the differential set arises from altered nucleosome composition in proliferating rys cells, we expect these sequences to provide greater support for separate emission processes than for a single, combined process. On the other hand, if aberrant DNA-nucleosome interaction is not causally relevant, and the reason for the apparently displaced nucleosomes is another macromolecular process, the evidence ratio should favour a combined process for the emission of the differential set—that is, there should be no difference between the unique sib and rys sequences that would justify the additional complexity of separate models.

To perform this calculation, we needed to construct background models of *D. rerio*’s genomic “noise” from which repetitive sequence signals characteristic of nucleosome positions could be extracted. Following the suggestion of Down and Hubbard [DH05] that a principled approach to the selection of background models is to train and test a variety of them on relevant sequence, we used the Julia package BioBackgroundModels (BIORXIV CITE GOES HERE) to screen a panel of 1,2,4, and 6-state HMMs against 0th, 1st, and 2nd order encodings of samples from the zebrafish genome, partitioned grossly into exonic, periexonic, and intergenic sequences. We found that each of these partitions is best represented by 6-state HMMs trained on a 0th order genome encoding (i.e. the HMMs emit the 4 mononucleotides), as determined by model likelihood given an independent test sample, as displayed in Fig. ??

In order to assess this, we performed micrococcal nuclease (MNase) digestion of genomic DNA (gDNA), and determined gross effects on polynucleosome organisation from gel analysis (Figure 5X). We subsequently performed MNase-Seq in order to map the specific positioning of nucleosomes on rys and sibling gDNA, and determined mapping analysis (Figure 5X). To confirm that chromatin organisation is indeed disrupted specifically in the CMZ, we acquired high-resolution transmission electron microscopy imagery of these cells in rys and sibling CMZs, observations from EM(Figure 5X).

### 5.2.6 The rys CMZ does not contribute to the differentiated neural retina; the central retina of rys is disorganised

Having characterised the effects of the rys npat mutation on the identity, nuclear organisation, and cell cycle of proliferating cells in the rys CMZ, we next addressed why rys mutant embryos are microphthalmic

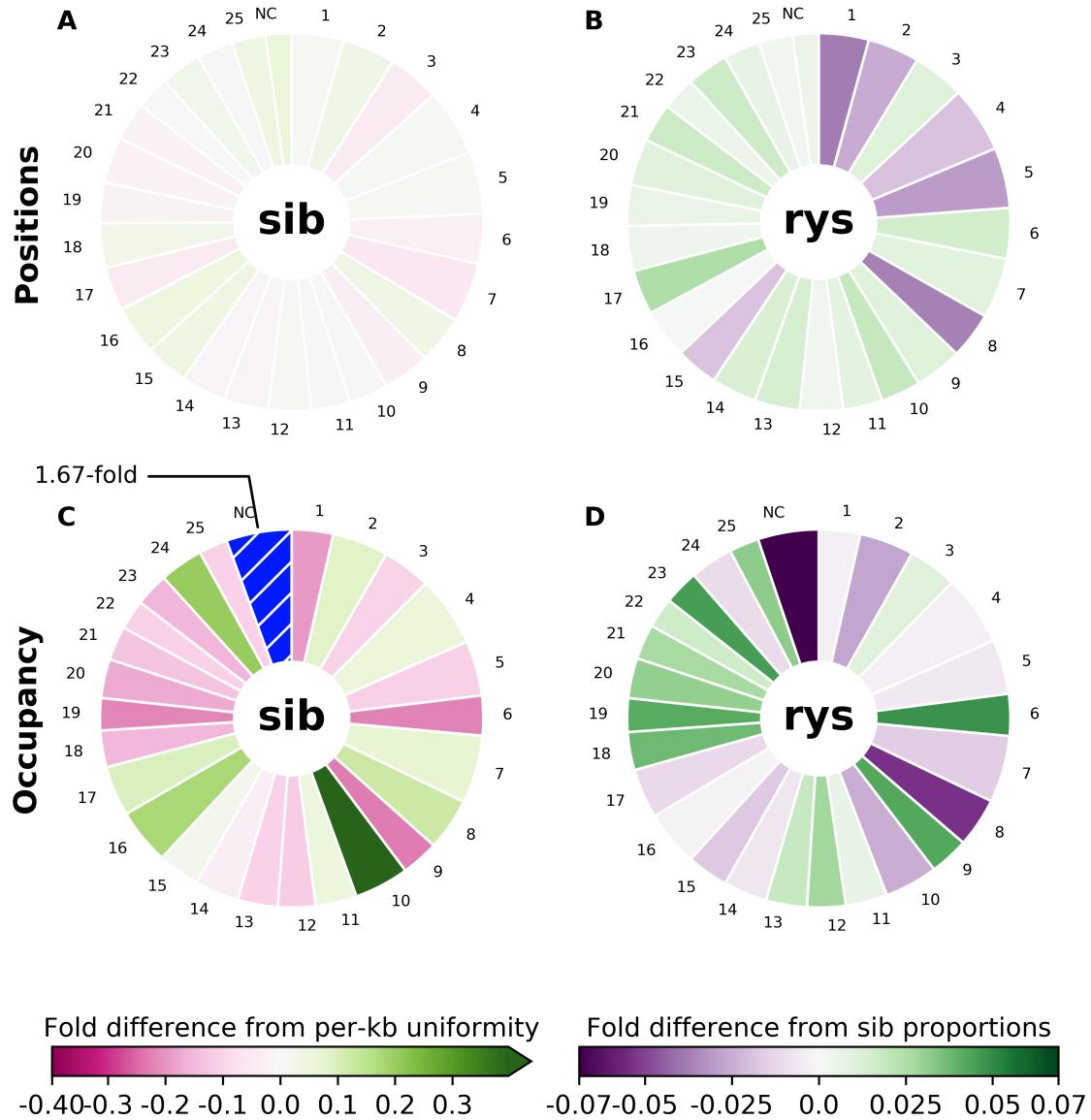


Figure 5.1: *rys* chromosomes are differentially enriched and depleted of nucleosome position density and occupancy.

Pie charts of nucleosome position density and occupancy by chromosome. Width of pie slices in panels A and B indicates the fraction of the total number of positions occurring in the numbered chromosomes and nonchromosomal scaffolds (NC). In panel A, depicting the sib genome, slices are colored according to the extent of deviation from an assumption of even distribution of nucleosomes across the genome. In panel B, depicting the rys genome, slices are colored according to the extent of deviation from the sib distribution. The width of slices in slices C and D indicate the fraction of the total nucleosome occupancy signal detected. Slice coloration depicts deviations from nucleosome occupancy distributions analogous to position distributions in A and B. Blue and white diagonal bars in the NC slice of panel C denote an out-of-scale positive deviation from the assumption of even distribution, i.e., sib NC scaffolds have 1.67-fold more of the total nucleosome occupancy signal than is expected from their length alone.

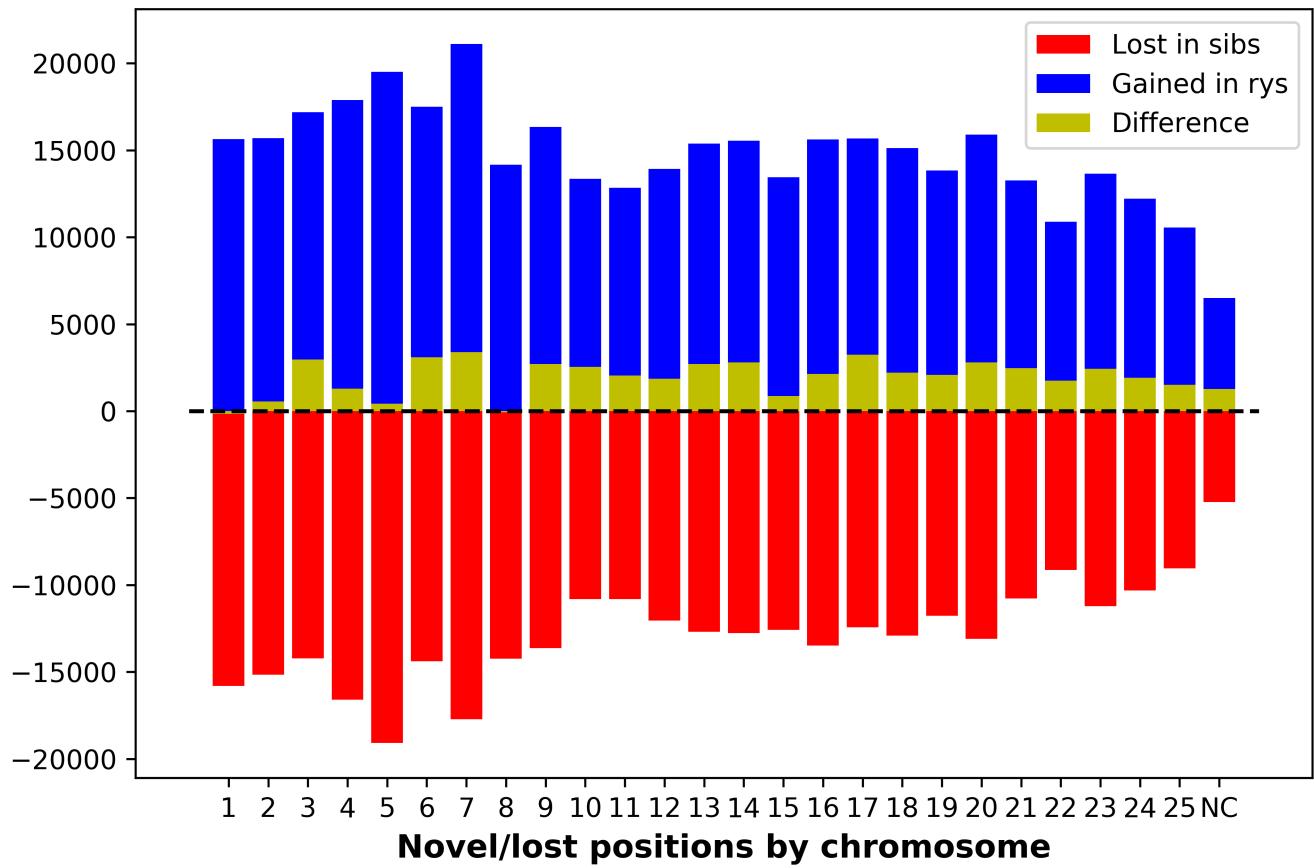


Figure 5.2: Novel nucleosome positions in rys occur in similar numbers to those lost from sibs.

Counts of positions found in sib but not rys (red bars, represented as negative numbers, as these are ‘lost’ in rys) and those found in rys but not sib (blue bars, ‘gained’). The magnitude of the difference between the counts is represented with a yellow bar.

despite having a cycling pool of CMZ progenitors. To do so, we performed an EdU pulse-chase analysis of 3, 5, and 7 dpf rys and sibling embryos, pulsing these animals for 8 hours and subsequently chasing for 1 week before assaying for EdU reactivity. While rys siblings contribute new cohorts to all three layers of the neural retina normally at all time points (Figure 6A), the majority of rys mutants do not contribute to the differentiated neural retina at all at any timepoint (Figure 6B). [more specific quantitative analyses required] We also investigated whether mutant rys embryos have an associated central retinal phenotype by assaying for specific markers of differentiated retinal cell types (Table 1). We found that, while all examined markers are present in the correct layers, many cell types are disorganized or patchy relative to siblings (Figure 6C). more specific analyses of particular cell types? Eg zpr-1 photoreceptors

### 5.2.7 The fate of rys CMZ cells is caspase-dependent apoptosis and subsequent phagocytosis by microglia

Given that rys CMZ progenitors do not contribute to the neural retina, and that the population of rys CMZ cells is not expanded commensurately with their ongoing proliferation and failure to differentiate, we set out to determine what the fate of rys CMZ progeny might be. To investigate the possibility that rys CMZ cells may be dying *in situ*, we examined high-resolution TEM micrographs of rys CMZs for karyorrhectic nuclei and apoptotic bodies. We consistently observed these features in rys mutant but not sibling CMZs (Figure 8A). Caspase-3 immunoreactivity of rys CMZs was also assayed; we found significantly more caspase-positive cells in rys mutant relative to sibling CMZs ( $n=??$ ;  $p<??$ ) (Figure 8B). Finally using an antibody directed to 4C4, an epitope associated with zebrafish microglia, in combination with TUNEL staining, we observed TUNEL-labelled bodies internalised in 4C4-positive microglia in rys mutants significantly more frequently than in rys siblings ( $n=??$ ;  $p<??$ ) (Figure 8C).

## 5.3 Discussion

The zebrafish rys mutant model provides a heretofore unique opportunity to study the role of an NPAT homologue in a complete, developing tissue. This has previously proven difficult using mouse Npat, proviral inactivation of which resulted in early embryonic arrest, prior to tissue formation (Di Frusco 1997). The survival and development of rys mutant embryos beyond this stage may reflect the presence of wild-type npat transcript contributed maternally [HSK<sup>+13</sup>]; rys mutants nevertheless do not survive beyond metamorphosis (~21dpf), so npat seems to be similarly obligatory for normal development in zebrafish, if over a longer timeframe. As teleost fish are known to have undergone whole-genome duplication subsequent to their radiation from vertebrates, it is possible that zebrafish may have multiple npat paralogues. We have excluded this possibility due to our failure to identify any significantly similar CDS sequences in the Zv9 zebrafish genome using BLAST. The zebrafish npat gene does have a substantially different genomic context from human NPAT, however: it is not associated with the eponymous ATM locus (which has been duplicated, and is present in this duplicate form elsewhere on chromosome 15). ATM's 5' position is, in zebrafish, occupied by hif1al, with the intergenic region lacking canonical E2F promoter sites. Of the 80 vertebrate genomes currently available from Ensembl, this organisation is shared only with the cave fish (*A. mexicanus*), with the human-like ATM-NPAT association preserved in all other species. This apparently evolutionarily novel genomic organisation for npat may be responsible for some of the differences in npat function we report here relative to its homologues. The rys mutation in zebrafish npat is associated with gross abnormalities in CMZ nuclear organisation and altered expression of progenitor markers. These changes were observed with a concomitantly shortened S-phase, a phenomenon which has not previously been reported in any NPAT mutant, knockout, or knockdown context. A shortened S-phase is nonetheless a viable alternative interpretation of some prior work; Gao et al. [GBB<sup>+03</sup>], for instance, reported a decline in the number of cells in S-phase in a population of synchronised U2OS cells treated with siRNAs directed against NPAT, coupled with an increase in the number of cells in G2/M. This was interpreted as a decreased propensity of NPAT-knockdown cells to proceed through the G1/S barrier, but could also be explained by a shortened S-phase resulting in fewer cells in S and faster progression to G2/M phase. Unfortunately, no experiments have been conducted with the robust cumulative thymidine analogue labelling paradigm we use here, so it remains unclear

whether the observed truncated S phase is unique to rys or if it might be present in other contexts as well. We subsequently identified alterations in histone mRNA transcription and 3' end processing as candidate mechanisms by which the rys phenotype might be produced. In both 6 and 8 dpf rys larvae as well as npat-morpholino treated 1dpf embryos, we observe increased abundances of histone and, specifically, polyadenylated histone transcripts, although the composition of affected histone families differs between these contexts. The specific mechanism by which these effects are produced in rys remains unclear, and the increases in histone transcript abundance observed in both mutant and morpholino-treated animals is at odds with observations that NPAT knockouts display decreases in histone transcription [YWNH03] and that the destruction of CDK phosphorylation sites on NPAT protein, which has putatively occurred in the rys npat protein, or treatment with a CDK inhibitor, result in similar declines in histone transcript abundance [Ma00, MGv<sup>+</sup>09]. This may imply the role of zebrafish npat in regulating histone transcription is be structurally different from what has been described for human NPAT. We are unable to determine from these data whether the overall increases histone transcript abundance are a consequence of increased histone transcription or of the greater stability of improperly polyadenylated histone transcript, however. The increased abundance of polyadenylated transcript in rys mutants and npat morpholino-treated embryos is consistent with the observed role of NPAT in recruiting CDK9 to replication-dependent histone gene clusters, known to be important in generating the normal stem-loop structure at the 3' end of these transcripts [PSS<sup>+</sup>09], suggesting that this role is conserved in zebrafish. It is also possible that particular histone genes give rise to polyadenylated transcripts in zebrafish, as observed in a variety of cell lines [KKT<sup>+</sup>13]; increased transcription of these particular genes might also account for the observed increase in polyadenylated histone transcript abundance. We speculate that the increased abundance of histone transcript in rys allows for more rapid, if dysfunctional, chromatin packaging, and consequently the shorter S-phase observed in the CMZ of these animals. Interestingly, although increased abundance of polyadenylated histone transcript has been associated with both cell cycle arrest and differentiation [KKT<sup>+</sup>13], we found that in the rys mutant CMZ, this was associated with altered cell cycle parameters, failure to differentiate, and ultimately, cell death by apoptosis. This suggests that the presence of polyadenylated histone transcripts are not directly related to particular cell cycle states or differentiated fates, but rather that a particular regime of coordination and control of the expression of polyadenylated and replication-dependent, stem-looped histone transcripts is required to achieve appropriate transitions between these states. In the case of rys, this seems to be related to the effect of altered histone expression on chromatin structure; we observed in the rys mutant CMZ altered nuclear morphology, chromatin ultrastructural features, and nucleosome positioning. As the spatiotemporal dynamics of chromatin architecture are known to be critically involved in both replication timing [GTR<sup>+</sup>10] and cell identity and fate [SVT13], the significant alterations we observe in rys mutant chromatin architecture, accompanied by our observations of altered cell identity marker expression [proteomics analysis references] provide a plausible molecular mechanism by which mutated npat protein can produce the observed components of the rys mutant CMZ phenotype via altered histone expression. Further experimentation is required to determine how specific changes in chromosome organisation (e.g. alterations in 3-dimensional chromatin organisation of gene expression, number of replication foci, etc.) produce specific features of the rys phenotype. The rys npat mutant has also provided an opportunity to elucidate some heretofore obscure features of the zebrafish retina. We found that the failure of the rys mutant CMZ to contribute to the differentiated neural retina was associated with a phenotype of progressive disorganisation in the central retina, which may suggest either that CMZ contributions are

required to maintain the normal ordering of the retina over time, or that npat has a previously unrecognised role in patterning and maintaining the viability of differentiated retinal neurons. Furthermore, we identified microglia as the cells responsible for the phagocytosis of dying cells at the retinal periphery in rys mutants, an activity which has not previously been ascribed to these neural immune cells. This behaviour on the part of microglia may be necessary for the homeostasis of the CMZ proliferative niche, and may in part explain the ability of the rys mutant CMZ to remain relatively intact for an extended period of time in spite of the failure of its progeny to differentiate and their eventual apoptotic fate.

# **Chapter 6**

## **Inferring and modelling nuclear dynamics in retinal progenitors**

### **6.1 Intro**

transition from discussion of modelling postembryonic CMZ and using nuclear shape to infer identity for modelling purposes to discussion of the implication of the correlations of nuclear state, cellular identity, and cellular function

### **6.2 Rys**

introduce zebrafish eye mutants in general and rys in particular

### **6.3 Histones and DNA sequence**

Background material on histone positioning and the effect of DNA sequence identity

### **6.4 Inferential logic**

briefly outline the chain of inference made in the paper and the influence of technical limitations, explain calculating evidence ratios by nested sampling as a form of Bayesian model selection

### **6.5 Technical caveats**

List important caveats to keep in mind when reading the paper

### **6.6 Separability of cell cycle and specification**

practical import of this is high- compare the conflation of Harris models to possible model framework which separates and modulates these independently for eg. transplants

## 6.7 Frontiers of nuclear dynamics

discuss recent experiments measuring chromatin conformation, pulsatile transcription etc

## 6.8 Integrating nuclear dynamics into models of CMZ- abstraction and biosemiotics

abstraction necessary due to computational limits, biosemiotics allow for the calibration of this theoretically

## **Part II**

# **Software Technical Reports**

# Chapter 7

## GMC\_NS.jl

`GMC_NS.jl` implements a basic Galilean Monte Carlo nested sampler [Ski12] in pure Julia. It is not capable of multiprocess likelihood calculations like `BioMotifInference.jl` and is not intended for high-dimensional problems requiring dynamic ensemble size adjustment [HHHL19] or multi-ellipsoidal sampling [FHB09]. `GMC_NS.jl` accepts arbitrary user-defined functions for model calculations, so could be used for any moderately-sized sampling problem with a continuous parameter space.

### 7.1 Implementation notes

The basic sampling scheme followed is, first, to initialize an ensemble of models with parameters sampled evenly from across the prior, then to compress the ensemble to the posterior by applying GMC moves to the least-likely model in the ensemble at a given iterate, constrained by its current likelihood (which is the ensemble's likelihood contour for that iterate). GMC tends to move model-particles across the parameter space very efficiently, particularly in the initial portion of compression across the uninformative bulk of the prior mass. That said, GMC particles may get "stuck" in widely separated minima, so there are instances when the least-likely particle has no valid GMC moves. In this case, the trajectory is not continued. Instead, a new trajectory is begun by resampling from the remaining prior mass within the ensemble. This is accomplished by one of two means. Firstly, an attempt is made to resample by "diffusion". This means a random particle remaining within the ensemble is selected, and random velocity vectors are sampled from this particle to find a starting location for the new trajectory (this vector is conferred to the new particle). If diffusional sampling fails (extraordinarily rare), a primitive ellipsoidal sampler serves as a backup. This draws an ellipsoid around the ensemble's current models in parameter space and samples evenly from the ellipsoid. While this is an acceptable method of sampling evenly from within the existing likelihood constraint, it is wildly inefficient for posteriors with any kind of interesting topology, which is both why multi-ellipsoidal sampling exists [FHB09], and why this single-ellipsoid sampler is confined to a backup role.

`GMC_NS.jl` follows the common convention of transforming parameter space to a unit-standardized hypersphere. The user is responsible for ensuring that the density of the prior is uniform, that is  $\pi(x) = 1$ , over the -1:+1 range of the hypersphere dimension. Utility functions `to_unit_ball` and `to_prior` which accept the prior and transformed positions and the relevant prior probability distribution, returning the unit hypersphere or prior coordinate, respectively. In future versions, this scheme is likely to be modified

to a 0:1 unit hypercube, which does not require additional operations beyond obtaining the cumulative probability of a parameter value on the prior distribution. The user may also specify a sampling "box" to bound particles from illegal or nonsensical areas of the parameter space (eg. negative values for continuous distributions on positive-valued physical variables). The box reflects particles specularly in the same manner as the likelihood boundary, with the exception that because the orientation of the nth-dimensional box "side" is known (eg. is the plane at n=0.), the reflection is performed by stopping the particle at the box side and giving it a new velocity vector identical to the old but with reversed sign in dimension n.

`GMC_NS.jl` attempts to maintain efficient GMC sampling by per-particle PID tuning of the GMC timestep. GMC treats models as particles defined by their parameter vectors, which give their positions in parameter space, as well as a velocity vector, normally chosen isotropically and only changed when the particle "reflects" off the ensemble's likelihood contour. In GMC, when a model-particle is to be moved through the parameter space (in this case, because it is the least likely particle in the ensemble), a new position is proposed by extending some distance along its velocity vector through the parameter space. This distance is determined by scaling the velocity vector by a "timestep" value, which can be thought of as the speed with which the particle is covering parameter space. As the model ensemble is compressed down to the posterior, this timestep must decline fairly evenly in order for GMC to be efficient, as the amount of available parameter space declines rapidly. Additionally, GMC particles may enter convoluted regions of parameter space that form small likelihood-isthmuses to other, more likely regions. In order to deal with this, each trajectory is assigned its own PID tuner, which maintains an independent timestep for the trajectory, tuned to target a user-supplied unobstructed move rate. In short, this will reduce the timestep if the particle is repeatedly reflecting off the likelihood boundary (in which case it is crossing the remaining prior mass without sampling much from it, or it is in a highly convoluted area of the local likelihood surface), and extend the timestep if it repeatedly moves without encountering the boundary, so that particles that are closely sampling an open region without encountering the boundary begin to "speed up". This scheme tends to produce a "shell" of swirling GMC particles reflecting obliquely off the likelihood contour, which gives good linear sampling of the entire likelihood range.

## 7.2 Ensemble, Model, and Model Record interfaces

## 7.3 Usage notes

## Chapter 8

# BioBackgroundModels.jl

# Chapter 9

## BioMotifInference.jl: Independent component analysis motif inference by nested sampling for Julia

### 9.1 Introduction

While many methods to detect overrepresented motifs in nucleotide sequences, only one is a rigorously validated system of statistical inference that allows us to calculate the evidence for these motif models, given genomic observations: nested sampling [Ski06]. Computational biologists almost immediately benefitted from Skilling’s original publication, with the release of the Java-coded nMICA by Down et al. in 2005 [DH05]. Unfortunately, this code base is no longer being maintained. It is, moreover, desireable to take advantage of modern languages and programming techniques to improve the maintainability and productivity of important bioinformatic code. We therefore re-implement Skilling’s algorithm for inference of DNA motifs in Julia [?] and characterise its function and performance.

### 9.2 Implementation of the nested sampling algorithm

Generally speaking, these methods are not available to us because the PWM representation of sequence signal in the ICA PWM model (IPM) is not readily susceptible to ordinary parameter space representation. This is for at least four reasons:

These issues make it difficult to see how one would mathematically describe the position of existing ensemble models within the parameter space at all, precluding sampling from a hypersphere or the use of GMC. Therefore, BioMotifInference (BMI), like it’s predecessor nMICA, samples new models by applying one of a collection of permutation functions to an existing model. Additionally, although BMI is supplied prior distributions on ICA parameters from which the initial ensemble is sampled, there is no way to calculate a sensical posterior from the models generated by the nested sampling process. The primary outputs of interest are the estimation of the Bayesian evidence for model structure given the data, given as its logarithm, as well as the models in the final ensemble (which constitute samples around the posterior mode or modes).

Because the more conventional methods to address the inefficiency of nested sampling by Monte Carlo used in cosmological models described above are unavailable, BMI uses an ad hoc method of adjusting the mixture of permute patterns that are applied to models. The basic permute logic (encoded by `Permute_Instruct` arguments to the `permute_IPM` function) is as follows:

### 9.3 Usage notes

### 9.4 Recovery of spiked motifs

### 9.5 Recovery of promoter motifs

## Part III

# Supplementary Materials

# Chapter 10

## Supplementary materials for Chapter 2

### 10.1 Materials and methods

#### 10.1.1 Zebrafish husbandry

Zebrafish used in this study were of a wild type AB genetic background. Embryos were derived from pairwise crossings. Larvae were collected upon crossing and held in a dark incubator at 28°C. At 1dpf, embryos were treated with 100 $\mu$ l of bleach diluted in 170ml of embryo medium for 3 minutes, rinsed and dechorionated. Embryo medium was changed at 3dpf. After this, all animals were maintained at 28°C on a 14-hour light/10-hour dark cycle (light intensity of 300 lux) in an automated recirculating aquaculture system (Aquaneering). Animals were reared using standard protocols [Wes00]. Fish were sacrificed by tricaine overdose at the appropriate timepoints indicated in the text. All animal experiments were performed with the approval of the University of Toronto Animal Care Committee in accordance with guidelines from the Canadian Council for Animal Care (CCAC).

#### 10.1.2 Proliferative RPC Histochemistry

##### 10.1.2.1 Anti-PCNA histochemistry

In order to assay the number of proliferating cells in zebrafish CMZs, we used anti-PCNA histochemical labelling, as PCNA is, in zebrafish, detectable throughout the cell cycle in proliferating cells. After sacrifice as described above, fish (or their razor-decapitated heads, in the case of fish older than 30dpf) were fixed in 1:9 37% formaldehyde:95% ethanol at room temperature (RT) for 30 minutes, followed by overnight incubation in a fridge at 4°C. Subsequently, the samples were removed from fix and washed 3 times in PBS. They were then cryoprotected by successive 30 minute rinses at RT on a rocker in 5%, 13%, 17.5%, 22%, and 30% sucrose in PBS. Samples were allowed to incubate overnight at 4°C in 30% sucrose. The next day, samples were removed from the sucrose rinse and infiltrated with 2:1 30% sucrose:OCT compound (TissueTek) for 30 minutes at RT. Samples were then embedded for cryosectioning and frozen. 14 $\mu$ m coronal cryosections were cut through the fish's heads and collected on Superfrost slides (Fisherbrand). These were stored in a freezer at -20°C until staining.

Staining was begun by allowing the slides to dry briefly at RT. Sections were outlined with a PAP pen, then rehydrated in PBS for 30 minutes at RT. Subsequently, sections were blocked in 0.2% Triton X-100 + 2% goat serum in PBS for 30 minutes at RT. This was followed by incubation in mouse monoclonal (PC10) anti-PCNA primary antibody (Sigma), diluted 1:1000 in blocking solution, at 4°C overnight. The next day, primary antibody was removed, and slides were rinsed five times in PDT (PBS + 1% DMSO + 0.1% Tween-20). Cy5-conjugated goat-anti-mouse secondary antibody (Jackson Laboratories), diluted 1:100 in blocking solution, was applied to the sections, which were incubated at 37°C for 2 hours. Secondary antibody was removed, followed by five rinses in PDT as above. Sections were counterstained with Hoechst 33258 diluted to 100 µg/mL in PBS for 15 minutes at RT. Counterstain was then removed, five rinses in PDT were performed as above, followed by a final rinse in PBS. Slides were then mounted in ProLong Gold antifade mounting medium (ThermoFisher Scientific), coverslipped, and sealed with clear nail polish. Slides were kept at 4°C until imaging.

#### 10.1.2.2 Cumulative thymidine analogue labelling for estimation of CMZ RPC cell cycle length

In order to obtain an estimate of cell cycle length in CMZ RPCs at 3dpf, we used cumulative thymidine analogue labelling following the method of Nowakowski et al.[NLM89]. 3dpf zebrafish were divided into ten groups of n5 animals and held in 10 mM EdU for 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 7.5, 8.5, 9.5, and 10.5 hours, after which they were sacrificed as described above. Samples were processed as described under “Anti-PCNA histochemistry”, except that goat-anti-mouse Cy2-conjugated secondary antibody (Jackson Laboratories) was used to label anti-PCNA, after which the Alexa Fluor 647-conjugated azide from a Click-iT kit (ThermoFisher Scientific) was added to the sections for 30 minutes at room temperature to label EdU-bearing nuclei. This was followed by five PDT rinses as described above, and resumption of the protocol with counterstaining, mounting, etc. The raw data is available in \empirical\_data\cumulative\_edu.xlsx

#### 10.1.2.3 Whole retina thymidine analogue labelling of post-embryonic CMZ contributions

We used thymidine analogue labelling as an indelible marker of CMZ contributions to the retina in Fig 2.5. Because thymidine analogues are incorporated into DNA during S-phase, postmitotic progeny of proliferating RPCs which take up the label may be detected at any point after this treatment. Dosing zebrafish with a thymidine analogue for a limited period of time ensures that the label is diluted to undetectable levels in cells which continue to proliferate. Repetitive dosing thus gives rise to labelled cohorts which represent a limited set of generations of cells which become postmitotic after the dose is provided. n=3 fish were held in 10mM BrdU (Sigma) diluted in facility water for 8 hours at 30, 60, 90, 120, and 150 days post fertilisation. 10 days after the last BrdU incubation, the fish were sacrificed as described above. One representative whole retina dissection is presented.

### 10.1.3 Confocal microscopy and image analysis

All histochemically processed samples (coronal sections of retinas and whole retinas) were imaged on a Leica TCS SP5 II laser confocal imaging microscope, using the Leica LAS AF software for acquisition. For coronal sections, central sections were identified by their position in the ribbon of cryosections. For Fig 2.7, the central section was imaged together with the flanking section on either side of it. Confocal

data was analysed using Imaris Bitplane software (v. 7.1.0). Cell counting analyses were performed by segmenting the relevant channels using the Surfaces tool to produce counts of PCNA positive (Figs 2.7, S4 Fig) or PCNA/EdU double-positive (S4 Fig) nuclei. Lens diameters were measured in the DIC channel using the linear measurement tools, across the widest point of the section of spherical lens.

#### 10.1.4 Estimation of CMZ annular population size

In order to estimate the total population of the CMZ in zebrafish eyes sampled throughout the first year of life, we counted PCNA-positive cells in the dorsal and ventral CMZ of central and flanking sections of these eyes as described above. We added the dorsal and ventral populations and took the average of the central and flanking section per-section populations to reduce the possibility of sampling error. We treated this as a sample of a torus mapped to the  $14\mu\text{m}$  sphere segment of lens intersected by the average section. As the surface area of this zone is given by  $S = \pi \times d_{lens} \times h_{section}$ , and the total surface area of the lens by  $A = \pi \times d_{lens}^2$ , where  $d$  is the diameter of the lens and  $h$  the section thickness, we may calculate an estimate of the total population by  $pop_{total} = \frac{pop_{section}}{S} * A$ . We therefore obtained our annular CMZ population estimates using our measurements with the simplified formula  $pop_{total} = \frac{pop_{section} \times d_{lens}}{h_{section}}$ , calculating  $pop_{total}$  on a per-eye basis to account for covariance of the measurements, and subsequently averaging across n=6 eyes per sampled timepoint (3, 5, 8, 12, 17, 23, 30, 60, 90, 120, 180, and 360 dpf). These data are presented in Fig 2.7.

#### 10.1.5 Modelling lens growth

To simulate the annular CMZ population, we wanted to be able to simulate the stem cells at the periphery of the retina occasionally undergoing symmetric divisions, so as to keep up the same density of stem cells in the first ring around the lens. We did this by normalising our lens diameter measurements to the 72hpf value, to produce a measure of relative increase in lens size over the first year of CMZ activity. We performed a linear regression of the logarithm of this relative increase vs the logarithm of time using the Python statsmodels library, using the constants derived from this regression for a power-law model of lens growth. This model was used to supply the `WanStemCellCycleModel` with a target population value as described below.

#### 10.1.6 Estimation of 3dpf CMZ cell cycle length

To produce an estimate of the length of the cell cycle in 3dpf RPCs, we first took the total count of EdU-positive cells in dorsal and ventral CMZ from our cumulative labelling experiment described above, and divided by the total dorsal and ventral CMZ RPC population, as measured by the number of PCNA positive cells. This gave the labelled fraction measurements displayed in S4 Fig. Following Nowakowski et al. [NLM89], we performed an ordinary least squares linear regression on these data using the Python statsmodels library. We then calculated the total cell cycle time  $T_c = \frac{1}{slope}$  and  $T_s = T_c \times intercept_y$ . While this method is not suited for heterogenous populations of proliferating cells, its use is justified here, as the He SSM to which the estimate is being applied makes similar assumptions as Nowakowski et al., in particular the homogeneity of the population. It should not be considered more than a rough estimate.

### 10.1.7 CHASTE Simulations

#### 10.1.7.1 Project code

All simulations were performed in an Ubuntu 16.04 environment using the CHASTE C++ simulation package version 2017.1 (git repository at <https://chaste.cs.ox.ac.uk/git/chaste.git>). The CHASTE package is a modular simulation suite for computational biology and has been described previously[MAB<sup>+</sup>13]. All of the code, simulator output, and empirical data used in this paper is available in the associated CHASTE project git repository at <http://github.com/mmattocks/SMME>, as well as in the supplementary archive [S1 File](#). In brief, the approach taken was to encapsulate the SSMs examined here as separate concrete child classes inheriting from the CHASTE `AbstractSimpleCellCycleModel` class. An additional model, representing the simple immortal stem cell proposed in Wan et al.[WAR<sup>+</sup>16], was similarly produced. Each model is generic and provided with public methods to set their parameters and output relevant simulation outcomes (and per-lineage debug data, if necessary). While these may be used in the ordinary CHASTE unit test framework, permitting their use in any kind of cell-based simulation, we wrote standalone simulator executables (apps, in CHASTE parlance) to permit Python scripting of the simulator scenarios used in this study. This allows for the multi-threaded Monte Carlo simulations performed herein. Therefore, the Python fixtures available in the project’s `python_fixtures` directory were used to operate the single-lineage simulators (`GomesSimulator`, `HeSimulator`, and `BoijeSimulator`) and single-annular-CMZ simulator (`WanSimulator`) in `apps/`, including the Cell-CycleModels and related classes in `src/`; the simulators output their data into `python_fixtures/testoutput/`. These data, along with the empirical observations (both from the Harris groups’ papers and our own described herein) available in `empirical_data/`, were processed by the analytical Python scripts in `python_fixtures/figure_plots/` to produce all of the figures used in this study.

We have attempted to make the project fully reproducible and transparent. CHASTE uses the C++ boost implementation of the Mersenne Twister random number generator (RNG) to provide cross-platform reproducibility of simulations. All of our simulators make use of this feature, permitting user-specified ranges of seeds for the RNG. We have also used the numpy library’s RandomState Mersenne Twister container to provide reproducible results for the Python fixtures, notably the SPSA optimisation fixture. The output of the project code will, therefore, be identical every time it is run, on any platform.

It should be noted that none of the code Harris’ group used in their reports has been published. We have made every effort to replicate the functional logic of the models as closely as possible. Nonetheless, it is not possible to determine precisely why, for instance, the original He SSM parameterisation failed so notably in our implementation.

#### 10.1.7.2 SPSA optimisation of models

In order to make a fair comparison between the He SSM and our deterministic mitotic mode alternative model, we coded an implementation of the SPSA algorithm described by Spall [Spa98]. This is available in `/python_fixtures/SPSA_fixture.py`. This algorithm, properly tuned, will converge almost-surely on a Karush-Kuhn-Tucker optimum in the parameter space, given sufficient iterates, even with noisy output (eg. due to RNG noise from low numbers of Monte Carlo samples, permitting conservation of computational resources). It does so by approximating the loss function gradient around a point in parameter space, selecting two sampling locations at each iterate, then moving the next iterate’s point in parameter space “downhill” along this gradient.

Our loss function was a modified AIC; we weighted the residual sum of squares from the PD-type mitotic probabilities by 1.5 in order to improve convergence, as the PD data are the most informative part of the dataset regarding the critical phase boundaries (PP mitoses occur in all 3 phases of the He model, DD occur in phases 2 and 3, PD mitoses occur only in phase 2). We also compared model induction count output (that is, panels A-C in Fig 2.4) up to count values of 1000 to penalise parameters producing very large lineage totals.

The values of the SPSA gain sequence constants were selected to be as large as possible without resulting in instability and failure to converge; this ensured that the algorithm stepped over the widest possible range of the parameter space in finding optima. The  $\alpha$  and  $\gamma$  coefficients were initially set at the noise-tolerant suggested values of .602 and .101 respectively [Spa98]. 200 iterations were performed. Initially, each iterate involved 250 seeds of each induction timepoint for the count data (panels A-C in Fig 2.4) and 100 seeds of the entire 23-72 hr simulation time for mitotic mode rate data (panels D-F). At iterate 170, these were increased to 1000 and 250 seeds, respectively, decreasing RNG noise to a low level. For the last 10 iterations, RNG noise was reduced to close to nil by increasing the seed numbers to 5000 and 1250, respectively;  $\alpha$  and  $\gamma$  were set to the asymptotically optimal 1 and  $\frac{1}{6}$  for these iterates, as well. The particular seeds used were kept constant throughout in order to take advantage of the improved convergence rate this affords [KSN99].

Because the simple SPSA can assign any value to parameters, our implementation uses constraints, projecting nonsensical values for parameters (e.g. negative values, mitotic mode probabilities summing to  $> 1$ ) back into legal parameter space. The use of constraints has no effect on the algorithm's ability to almost-surely converge within the given parameter space [Sad97]. As we felt that the deterministic mitotic mode models' "sister shift" should be relatively small in order to be biologically plausible, we also constrained this parameter such that 95% of sister shifts would be less than the mean length of the shortest phase.

The numerical results of the SPSA algorithm are available in `/python_fixtures/testoutput/SPSA/HeSPSAOutput`, alongside png images of output from each iterate, enabling the visualisation of the progress of the algorithm.

**Monte Carlo simulations** Subsequent to SPSA optimisation, the parameters derived from this procedure were used to perform Monte Carlo simulations of large numbers of individual lineages, using ranges of 10000 non-overlapping seeds for each of panels A, B, C, G, H, and I, and 1000 for each of panels D, E, and F of Fig 2.4 (using the SPSA-optimised parameter sets) and [S1 Fig](#) (using the original He et al. parameterisation). This was performed using `/python_fixtures/He_output_fixture.py`.

Similarly, 100 seeds were used in Monte Carlo fashion to simulate individual annular CMZ populations using `/python_fixtures/Wan_output_fixture.py`, which scripts the WanSimulator. Each of these simulations is initialised with an RPC population drawn from a normal distribution given the mean and standard deviation of the CMZ torus estimated from our 3dpf observations, as described above. Each RPC is given a TiL value randomly chosen from 0 (newly born from a stem cell) to 17hr (exiting the CMZ, "forced to differentiate", in the terms of Wan et al.). A further population of stem cells, sized at  $\frac{1}{10}$  of the RPC population, is added using the `WanStemCellCycleModel`. This stem population divides asymmetrically except when it falls under its target population value determined by the model of lens growth described above; as long as this condition holds, the stem cells will divide symmetrically to keep up their numbers.

**Simulation data analysis** Not all of the numerical data used in He et al. and Wan et al. has been published. As such, we reconstructed the lineage data from He et al.’s Figure 5C, used by He et al. to obtain their mitotic mode probabilities; our reconstructed values are available in `/empirical_data/empirical_lineages\`. We also reconstructed the He et al. WT and Ath5 morpholino data (Fig 2.4 panels H, I) and Wan et al. lineage count probabilities (panel G) from the relevant figures. These values are entered into `/python_fixtures/figure_plots/He_output_plot.py`, where they are used in the AIC calculations described below.

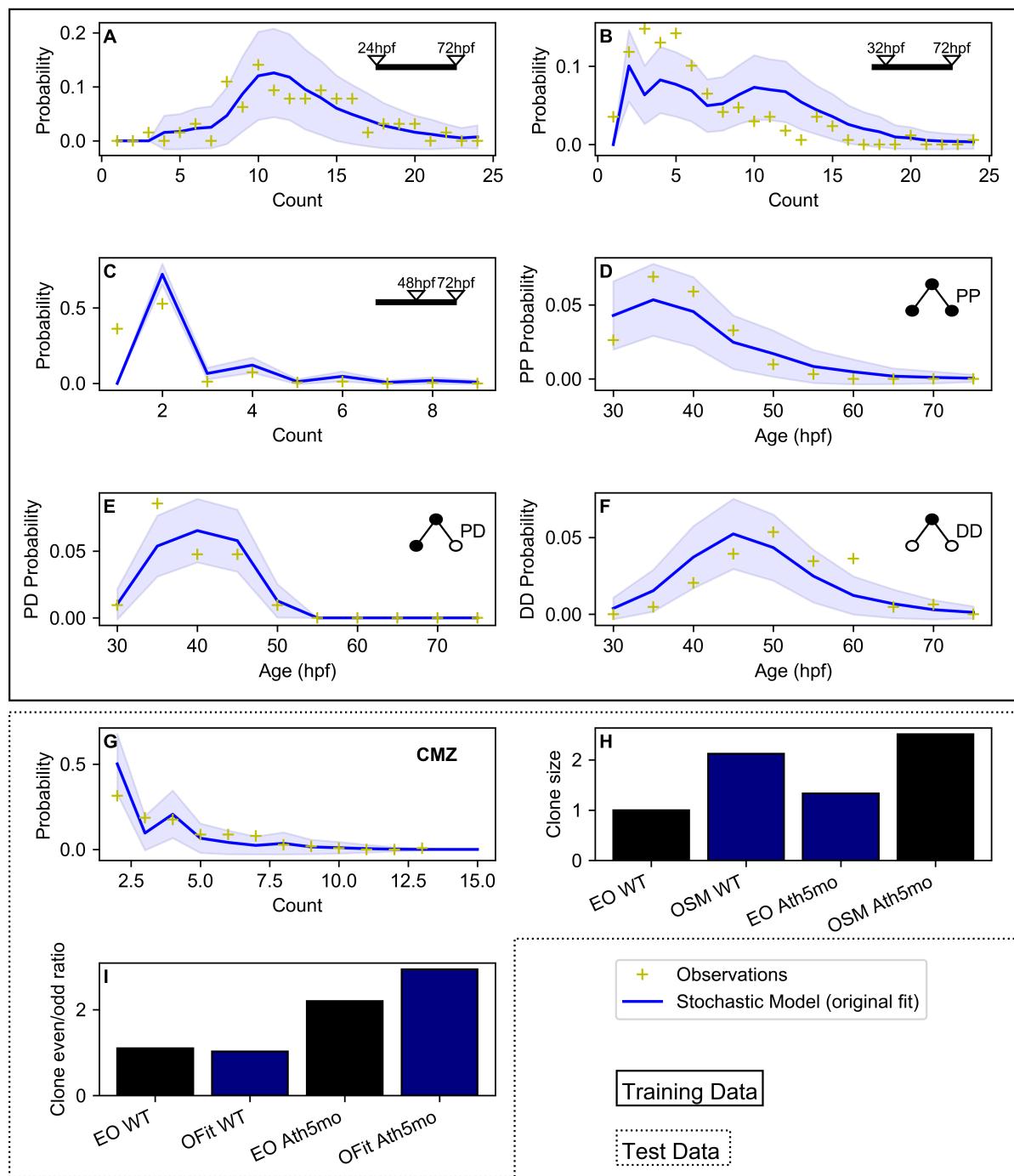
In order to assess the performance of the models used in our simulations, we used Akaike’s Information Criterion (AIC). This allows us to compare models with dissimilar numbers of parameters, as AIC trades off goodness-of-fit against parameterisation. The values reported in Table 2.1 were produced by `/python_fixtures/figure_plots/He_output_plot.py`.

The 95% CIs for the model output presented in this paper were estimated by repetitive sampling of the output data, using the same number of lineages observed empirically. 5000 such samples were performed. This provides a reasonable estimate of the confidence interval given the limited observational sample size.

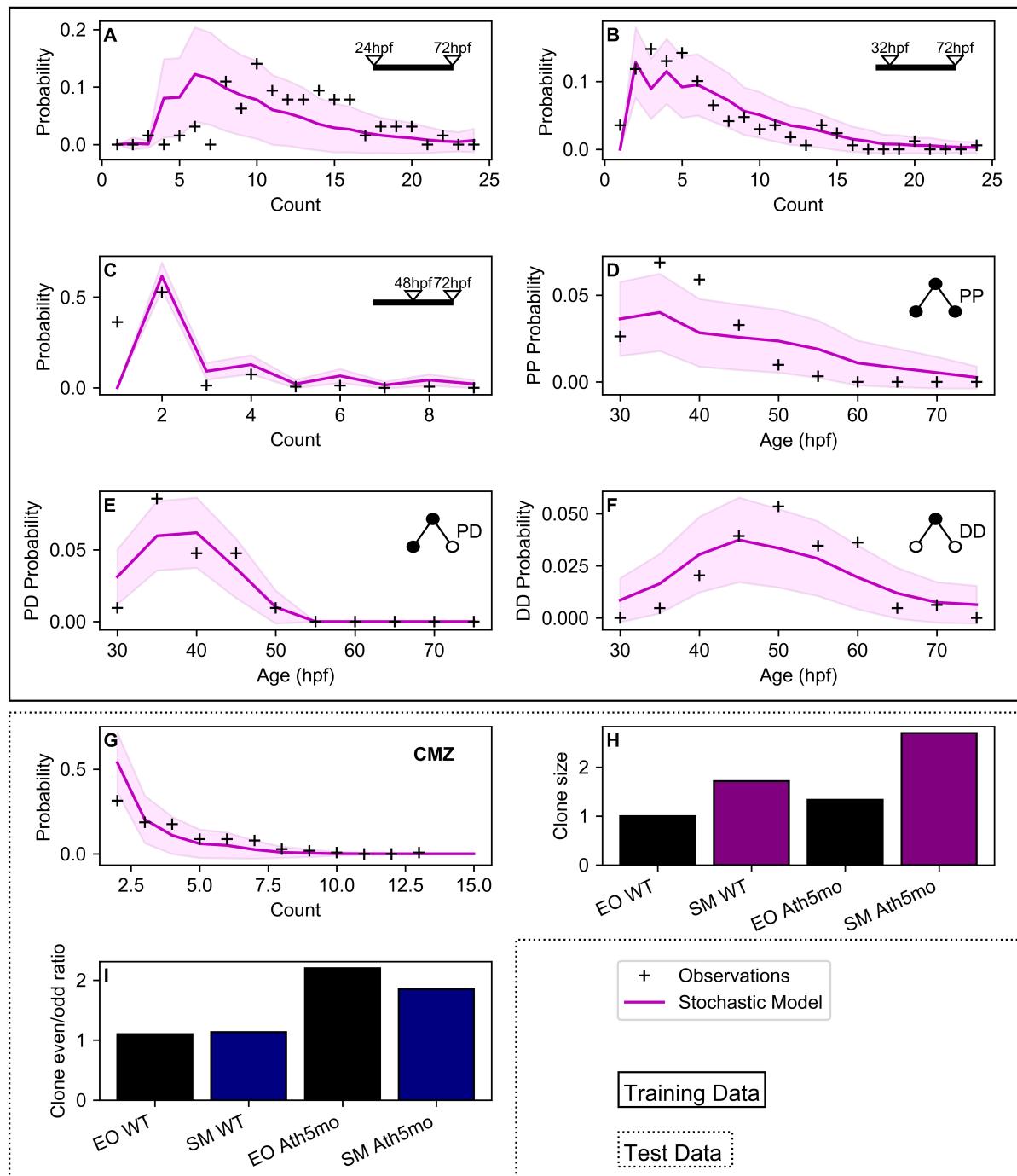
## 10.2 Supporting information

### 10.2.1 S1 File.

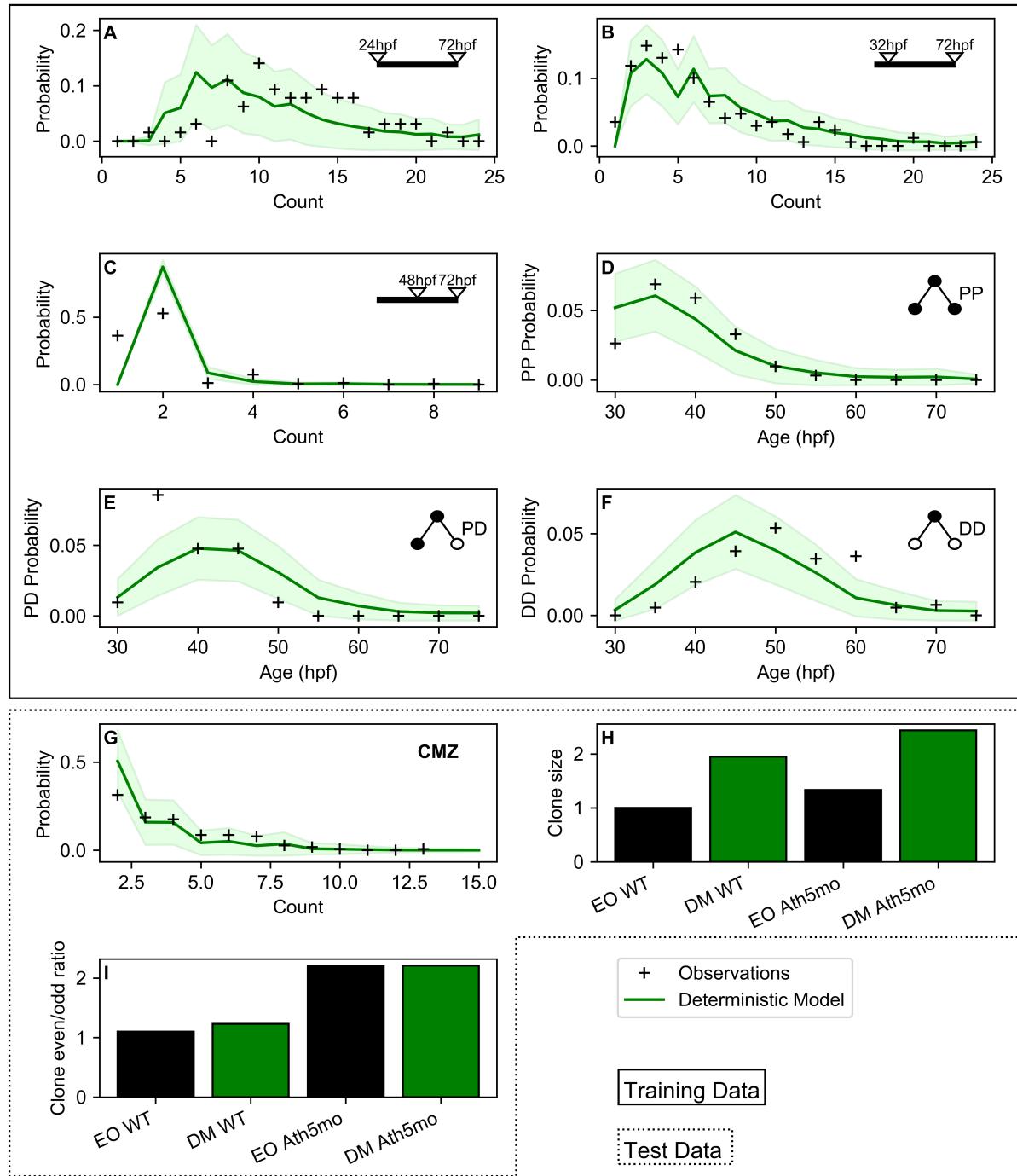
**Code archive** Archive of all code and code output used in this paper. *Available in this thesis in ??.*



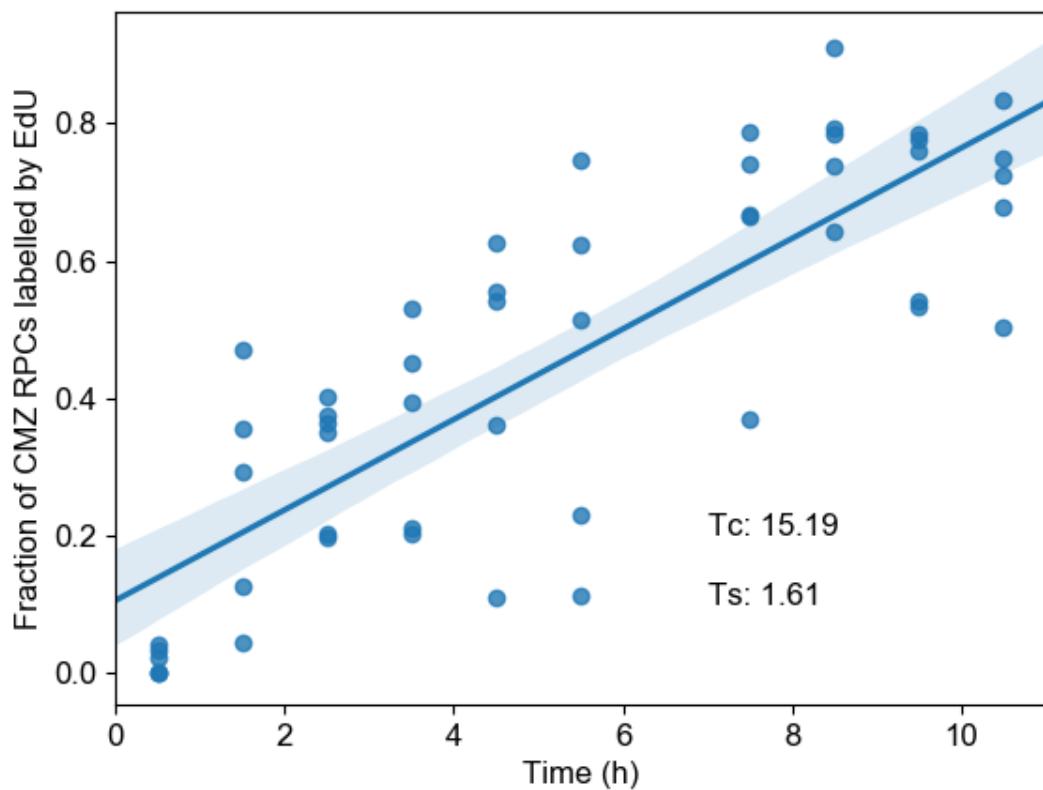
**S1 Fig. He SSM original parameterisation model output.** As Fig 2.4; model output uses the parameterisation given in He et al.[HZA<sup>+</sup>12]. Note significant deviation from observations in panel B, reflecting excess proliferative activity of modelled RPCs.



**S2 Fig. SPSA-optimised He SSM model output.** As Fig 2.4; stochastic model output displayed alone.



**S3 Fig. SPSA-optimised deterministic mitotic mode model output.** As Fig 2.4; deterministic model output displayed by itself.



**S4 Fig. Cumulative EdU Labelling of 3dpf CMZ RPCs** Fraction of PCNA-labelled CMZ RPCs which bear EdU label over time, as determined by histochemical labelling of central coronal cryosections of 3dpf zebrafish retinas. Dots represent results from individual retinas. Line is the ordinary least squares fit  $\pm$  95% CI. Cell cycle length ( $T_c$ ) and S-phase length ( $T_s$ ) are estimated from this fit following the method of Nowakowski et al.[NLM89]

## Chapter 11

# Supplementary materials for Chapter 4

- 11.1 Description of the computational cluster used in the work
- 11.2 Developmental progression of naso-temporal population asymmetry in the CMZ
- 11.3 Methods
  - 11.3.1 Statistical Analyses

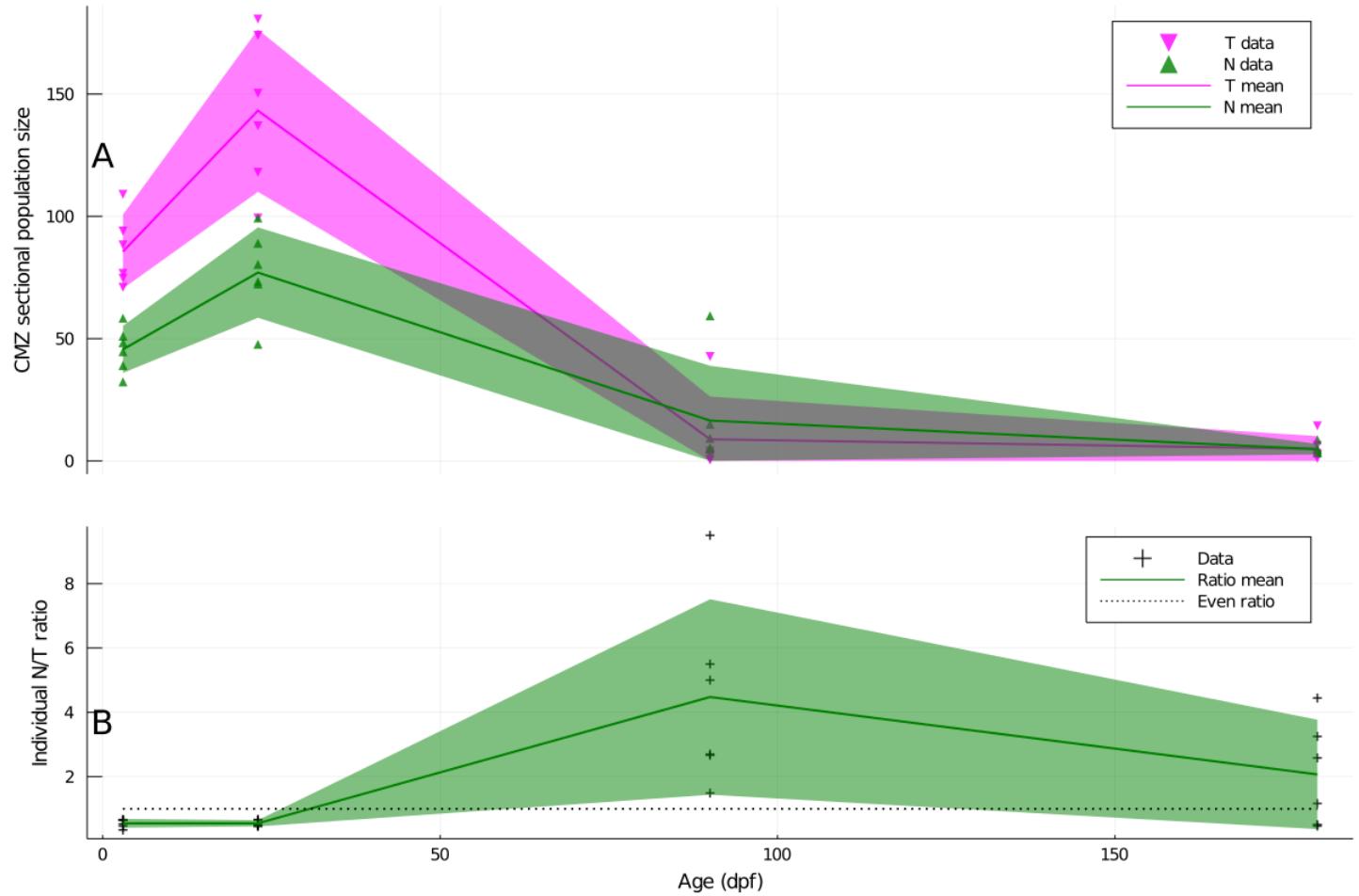


Figure 11.1: **Developmental progression of naso-temporal population asymmetry in the CMZ.**

Marginal posterior distribution of mean nasal (N) and temporal (T) population size in  $14\mu\text{m}$  transverse cryosections (panel A) or intra-individual N/T count asymmetry ratio (panel B),  $\pm 95\%$  credible interval,  $n=6$  animals per age. Data points represent mean counts from three central sections of an experimental animal's eye.

## **Chapter 12**

# **Supplementary material for Chapter 5**

# Chapter 13

## Theoretical Appendix A: Statistical methods and model theory

### 13.1 Statistical Theory

#### 13.1.1 Bayesian parameter estimation

##### 13.1.1.1 Problems with frequentist inference using normal models of sample data

Typical biological practice is to report the mean and variance of a sample, assuming a normal distribution of the error around the mean. In other words, the sample is taken to be representative of a larger population; that population is modelled by a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ; the parameters of the maximum likelihood estimate (MLE) for the normal model are the values reported. A slightly more sophisticated approach is to report the standard error of the mean of the sampling distribution the sample is taken to be drawn from, if more than one sample can be obtained, although this usually plays no role in hypothesis testing (often conducted by t-test).

This approach has a number of defects which follow from one another. We are reporting MLE parameters without any account of our uncertainty about those parameters. There is no way to incorporate prior information we have about the parameters (even just to admit total ignorance about them). This leads to overfitting of our estimates to the sample data. Practically speaking, this means our estimate of the mean is stated too precisely, and the variance is too sensitive to outliers.

Additionally, the plain-sense interpretation of the estimates are often unclear. Means are usually reported plus-minus variance,  $\mu \pm \sigma$ , and  $\sigma$  is often erroneously interpreted as uncertainty about  $\mu$  rather than an estimate of a second parameter, the variance of the normal population model. If the frequentist confidence interval for  $\mu$  is reported, it is explicitly not understood as the interval in which we have e.g. 95% confidence that  $\mu$  lies, but rather as the interval in which, in the case we repeat the experiment indefinitely,  $\mu$  will be found in 95% of samples. Hypothesis tests are given similarly confusing interpretations involving long-run repeated experiments. These interpretations are widely, if not ubiquitously, misunderstood or ignored in favour of technically incorrect but comprehensible ones [?, ?].

### 13.1.1.2 The Bayesian approach to normal models of unknown mean and variance

Bayesian methods rectify these problems by understanding the normal model as a model of our information about the population and not of the population itself. This epistemological view of statistics is explicated in [Section 14.0.2](#). Normal gaussian distributions are well-justified both by their ubiquitous success in parameter estimation and by information theoretic considerations [[JBE03](#)], and need not reflect the actual distribution of the population. However, we wish to express our uncertainty about the parameters of a normal gaussian distribution by giving further distributions over the mean  $m$  and variance of the normal distribution, with variance usually expressed as precision,  $\lambda = 1/\sigma^2$ . Typically, this is done by assuming normally distributed uncertainty on  $\mu$  and gamma distributed uncertainty on  $\lambda$ , giving rise to a joint normal-gamma (NG) distribution [?]:

An NG distribution may thus serve as a model of our prior information about the population being measured. Because my estimates are the first ones I have made about the relevant populations, and I have no specific guide as to the actual numbers of cells to expect, I have chosen to use the uninformative NG prior:

$$p(m, \lambda)$$

lies within that range, but is rather understood as the probability that, if the experiment were repeated indefinitely, 95

The appropriateness of the normal model is often in question because it is taken to represent some actually-existing population (which are often not well modelled by normal gaussians). Comparisons of these models using t-tests are given complex interpretations involving long-run rates of error

In Bayesian statistics, available information about a parameter is often modelled by a gaussian distribution over possible values of the parameter.

## 13.1.2 Bayesian Epistemological View on Model Comparison

## 13.2 Model Theory

### 13.2.1 Model optimization

#### 13.2.1.1 MLE

### 13.2.2 Overfitting, Underfitting

### 13.2.3 Monte Carlo simulation

Referenced on pages: [92](#)

### 13.2.4 The Akaike Information Criterion For Model Selection

### 13.2.5 Simple Stochastic Models

The Simple Stochastic Model is schematically summarised in Figure 13.1. This is the basic structure of the great majority of formal models in the stem cell literature, derived from post-hoc analyses of populations taken to include stem and progenitor cells. The population-level approach is usually explicit, as no differentiation is made between types of proliferating cell- in general, no particular cell is identified

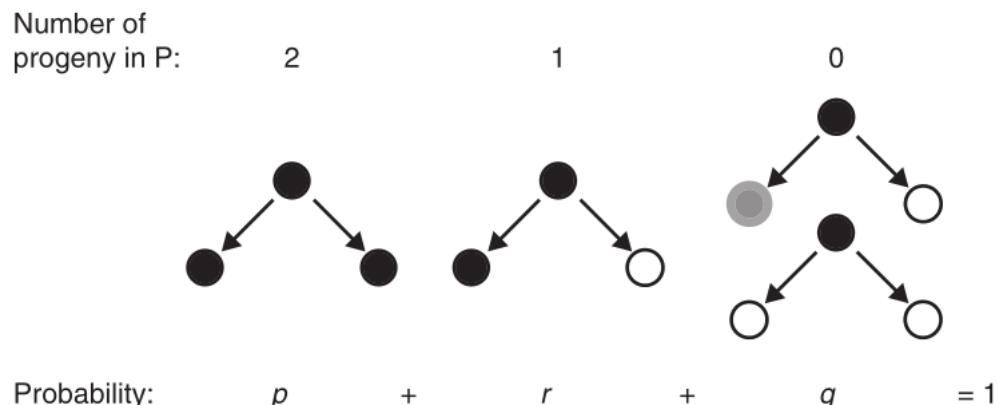


Figure 13.1: Simple stochastic stem cell model, representing probabilities of cell division events, excerpted from Fagan 2013 pg. 61. Black circles denote proliferative cells, while white and grey circles denote different types of postmitotic offspring. “Number of progeny in P” is the number of mitotic offspring produced by each type of division. The probability of each division type must sum to 1, as all possibilities are represented, granting that the division types are defined by the postdivisional mitotic history of the offspring.

with a stem cell, nor can any be identified from the necessarily retrospective population data used to infer the parameters of the model.

The central concept of the model is that divisions can be categorised by the number of progeny which remain mitotic after the division. It is important to note that a mitotic event cannot presently be categorized in this fashion except retrospectively. This must be kept in mind when analysing models of this type, as this categorisation does not necessarily imply that there is some mechanism by which the cell specifies the fate of offspring *at the time of mitosis*, although there is extensive evidence for the coupling of mitotic and specification processes at the molecular level.

In effect, then, the model compresses the process of fate specification into individual mitotic events. Since the primary distinction between cells in the model is simply whether they are proliferating or not, the model also elides any heterogeneity within the proliferating population. Beyond not identifying particular cells as “stem cells”, this may make models derived from the SSM inappropriate for proliferative populations with a large degree of heterogeneity. One may think here of the classic idea of a small number of slowly proliferating “true” stem cells and a larger population of rapidly dividing “transit amplifying” progenitors- this type of internal structure within the proliferating population can only be represented by multiple, independent SSMs (as implemented in Chapter 4).

As Fagan notes, in the SSM, “relations among  $p$ ,  $r$ , and  $q$  values entail general predictions about cell population size (growth, decrease, or ‘steady-state’), and equations that predict mean and standard deviation in population size, probability of [lineage] extinction, and features of steady-state populations are derived.”<sup>1</sup>[Fag13, p.60]

Typically, this type of model has been employed to describe population dynamics of proliferating cells in assays generating ostensibly *clonal* data, where a “clone” here refers to the population constituted by all of the offspring descended from some particular (usually “initial” and sometimes therefore taken for

<sup>1</sup>While Fagan refers to “stem cell” extinction, the model does not specifically define stem cells, nor does it imply intergenerational continuity, such that a particular intergenerationally identified stem cell should be said to have become extinct. The unit which survives or is made extinct is the lineage derived from some particular proliferative cell.

“stem”) proliferative cell. This population is the *lineage* generated by some particular dividing cell.

### 13.3 Traditional roots of the SMM: Till and McCulloch and the simple stochastic model (SSM)

At the core of the SMM are variants of the most common model used by stem cell biologists, the Simple Stochastic Model or SSM, described in more detail in Section 13.2.5. The SSM consists of a Galton-Watson branching process, a stochastic process originally intended to model the lineage extinction of surnames. Wikipedia describes a stochastic process as “a mathematical object usually defined as a collection of random variables” [Wik18]. In the case of branching process models applied to proliferative cells, the random variable determines the mode of division of each cell within the lineage, with this mode being defined by the proliferative state (construed in the model as being either mitotic or postmitotic) of progeny. For any given division, a cell may produce two mitotic, one mitotic and one postmitotic, or two postmitotic progeny, and each of these division modes is given a defined (often, but not always, static) probability. Given these values, the history of a cell lineage may be simulated; the output of many of these simulations pooled together, usually by what is referred to as the ”Monte Carlo method”, allows the statistical properties of the dynamics of population of simulated cells to be estimated.

The concept of “stochasticity” has a long and fraught pedigree in the SCBT. We find it deployed in an identical manner in the early 1960s as today, as in the classic modelling effort of Till, McCulloch, and Siminovitch<sup>2</sup> [TMS64], explaining the variability in the size of ectopic spleen colonies formed by hematopoietic stem cells in irradiated mice by means of a Monte Carlo simple stochastic model (SSM). “Stochastic” is used in an ambiguous manner here, and, importantly, we find precisely the same ambiguity in Harris’ a half century later. This ambiguity arises from the application of the term “stochastic” to the biological *process* under investigation, to the process’ *outcomes*, and to the *model* constructed to describe the process. This is particularly obvious in this early work, entitled “A Stochastic Model of Stem Cell Proliferation”, which states that “variation [in clonal lineage size] may be generated by a well-known probabilistic (‘stochastic’) process, the ‘birth-and-death’ process”, and that this “process is operative when an entity, for example, a single cell, may either give rise to progeny like itself (‘birth’), or be removed in some way (‘death’) and these two events occur in a random fashion.” [TMS64] As the significance and import of the SMM directly depend on what the meaning of “stochastic” is, and since the manner in which Harris uses this concept is directly descended from the usage of Till et al., we must sort through this ambiguity before proceeding.

We may clarify the matter by returning to the biological issue at hand, which Till et al. frame in the terms of classic cybernetic control theory:

[T]he development of a colony involves processes of differentiation occurring among the progeny from a single cell. Analysis of the cell content of the resulting colony might be expected to cast light on any control mechanisms which act during colony formation. If rigid control mechanisms are operative, acting on cells of a relatively constant genotype, all colony-forming cells might be expected to behave in a similar fashion, and colony formation should be a relatively uniform process, giving rise to colonies with very similar characteristics. Alternatively, if control is lax, colonies with widely differing characteristics might be

---

<sup>2</sup>It is worth noting that none of these investigators actually performed the relevant modelling calculations, leaving these to the U of T computer scientist L. Cseh. This division of labour remains lamentably common, and is likely responsible for many of the problems biologists have in understanding the meaning of their mathematical models.

expected to develop. Results which may bear on this problem are available from experiments in which colonies were analyzed for their content of colony-forming cells. It was found that, while most colonies contained these cells, their distribution among colonies was very heterogeneous, with many colonies containing few colony-forming cells, and a few containing very many. This result suggests that control is lax. [TMS64]

This makes quite clear that the essential question here is how the process which produces stem cell proliferative and specificative behaviours is structured. We may think of Waddington's topological model of cellular specification, itself inspired by cybernetic theory. Rigid and lax control schemes would, if modelled in this topological fashion, present themselves as very different "landscapes". The abstract concept At the logical extreme, rigid control of stem behaviour would be represented by a single deep channel, down which the "ball" (behaving something like a ball bearing) representing "cellular state" rolls, reliably, at the same rate, in every single instance. Conversely, extremely lax control would be represented by a landscape resembling a Galton board<sup>3</sup>, a broad, flat slope with many pegs which may bump the ball this way or that, tending to cluster the balls in the general center of the slope but never preventing some of them from ending up at either extreme (a properly constructed Galton board produces a Gaussian distribution of balls at the bottom of the slope). This metaphor immediately reveals that, *contra* Till et al.'s interpretation of

---

<sup>3</sup>The Galton board, named after its inventor, Sir Francis Galton, is significant because the statistical methods Till et al. use to solve the 'birth-and-death' process were also invented by him.

## Chapter 14

# Theoretical Appendix B: Metaphysical Arguments

If one objects that these are not properly “biological” considerations, I can only respond by saying that this is no longer the case, if it ever were. What I have called here the ”Systems Biology Encounter” (SBE) has made plain the reality described by the great German-American philosopher of science Nicholas Rescher as follows:

The ramifications and implications of philosophical contentions do not respect a discipline’s taxonomic boundaries. And we all too easily risk losing sight of this interconnectedness when we pursue the technicalities of a narrow subdomain. In actuality, the stance we take on questions in one domain will generally have substantial implications and ramifications for very different issues in other, seeming unrelated domains. And this is exactly why systematization is so important in philosophy - because the way we *do* answer some questions will have limiting repercussions for the way we *can* answer others. We cannot emplace our philosophical convictions into conveniently delineated compartments in the comfortable expectation that what we maintain in one area of the field will have no unwelcome implications for what we are inclined to maintain in others. [Res05, p.97]

The introduction of “Systems” methods (mainly drawn from various branches of the complexity sciences) has forced us to contend with the “implications and ramifications” of the methodological, epistemological, and metaphysical contents of the scientific traditions they are drawn from. If we do not understand what some sophisticated mathematical method assumes about the system to which we apply it, we are bound to make errors in doing so, and we cannot know what “limiting repercussions” the use of these methods to answer some questions will have for future investigations<sup>1</sup>. To fail to take heed of this is simply to concede that we do not really care about ensuring that what we say makes sense, that it is not contradictory, spurious, or simply meaningless. Refusing to make this concession, and lacking any intuitive genius that would allow me to proceed without an carefully laid plan, I have made resort to borrowings from philosophers, mathematically- and theoretically-inclined biologists, statisticians, and so forth, to formulate one. I have attempted to do so in a disciplined way; my intent here is not to obfuscate

---

<sup>1</sup>These “limiting repercussions” are not limited to a restriction in the kinds of methods that can consistently be used, given some particular “systems biological” approach. If we make no effort to understand these repercussions, we run the risk of having an internally contradictory or degenerate research program, with all of the associated wastages and opportunity costs.

with unnecessary philosophical speculation, but rather to clearly document the conceptual background used to tackle this particular problem in its context.

The meta-method I outline in this chapter assumes that Paul Feyerabend's view of science as a constellation of different *traditions* as substantially correct. This is no longer a particularly contentious view- for the molecular biologist confronting the profoundly foreign, esoteric, and opaque utterances of mathematicians and physicists from the complexity sciences, it is a lived reality. I also accept Feyerabend's contention that historical scientific development occurs *counterinductively*, when traditions show up one another's implicit natural assumptions and metaphysical content. By doing so, scientific traditions make available new ways to think about natural phenomena. I have used this basic view to structure my approach- my objective is to proceed in a way that maximises the counterinductive potential of the historical moment.

The meta-method itself consists in analysing the appropriate "metascientific unit" of molecular biological practice, the Extended Heterogenous Joint Mechanistic Explanation (EHJMEx), Melissa Fagan's JMEx concept [Fag15b] slightly modified, and extended in time to allow accounting for the development of explanations over several different primary papers. By doing this with a *counterinductive general model* in mind, I hope to reveal some of the "metaphysical ingredients" implied by Harris' explanation and its models, and to suggest how different ones might provide better approaches.

The general modelling approach I have chosen conceives of cells as a type of semiotic agent. This agent-based modelling approach allows the traditional models of population-level stem cell modelling to be expressed as a subset of the larger global model. This in turn permits models with different global metaphysical implications to be compared for local explanatory value. Because the claims I make revolve around the validity of these comparisons, I have explained and defended this at some length.

I have subsequently documented the most important mathematical, methodological and metaphysical "ingredients" present in Harris' explanations and in my own. I have attempted to place these within the context of the SBE and offer some remarks regarding the epistemological basis of statistical generalisations of complex systems to make this more meaningful for the biologist reader.

Finally, I have made an attempt to assess the relative sustainability of the different approaches to systems modelling implicated by this discussion, as first attempt to guide a research program with an explicitly scientific, realistic futurology in mind, relying heavily on Nicholas Rescher's explications of the practical and in-principle limits of scientific and technical progress.

I have sought to make this chapter useful for my own future reference, and for any colleagues coming out of various parts of the molecular biology tradition, who are by now confronted with an astonishing variety of ways to interpret biological phenomena, and few clear guidelines on how they might structure their research programs in light of the SBE. It is nevertheless, by necessity, confined to considerations relevant to retinal stem cells in zebrafish. Due to the limited space and time available, I have made what historians and philosophers of science probably should consider gross oversimplifications. I consider my overall line of reasoning here to be merely one way to "tell the story" of what is happening to us as biologists, in a way that seems to provide a productive way to think about this problem.

As I have indicated in this chapter, I intend it in the spirit of what Feyerabend called "open exchange"- I do not intend to dictate the terms of future scientific exchange, or to replace one model with another, but rather to suggest one possibility for how we might compare explanations and responsibly guide research programs given the complexity of the present and the uncertainty of the future.

### 14.0.1 Metaphysics, Epistemology

Throughout this chapter, I have used the terms “metaphysics” and “epistemology” in reference to assumptions, axioms, postulates, etc. about reality and knowledge, respectively. The asking of a question within the natural sciences always has background propositions from both of these domains that are required to make sense of any answer. That is, in order to perform any experiment, we must start with some idea about what sort of thing a phenomenon could consist of (eg. we decide a cell consists mainly of macromolecular constituents arranged in space-time, which informs the methods we use to study cellular life), as well as an idea of what a good answer might be (we must have a sense of how our explanations correspond to reality).

As Nicholas Rescher notes, speaking here of metaphysical issues specifically:

Metaphysical issues are thus ‘basic’ or ‘fundamental’ because they are the product of a methodological stance that facilitates empirical inquiry rather than being a product of our observational study of nature. What those principles of traditional metaphysics do is to provide question-generic presuppositions of factual inquiry. Natural science (physics, as the Greeks called it) provides the specific answers to our specific questions. But metaphysics is a matter of ‘first principles’: it sets out the presuppositional framework within which those answers are developed. [Res00, p.4]

Rescher further subdivides metaphysics into two broad categories:

Presuppositional inquiry in relation to science has two aspects, depending on whether we ask, ‘What presuppositions are called for if we are to do science *at all*?’ or ‘What presuppositions are called for if we are to do science the particular way in which the course of experience has ultimately taught us to proceed?’ Those initial most fundamental principles are fixed. But then the less fundamental implementation is something else again, something learned, something in relation to what the case of experience costs. At this less fundamental level the presuppositions and methodological principles of natural science must be retrospectively informed and restructured in the light of the deliverance of scientific agency itself. (At this stage we can usefully resort to Otto Neurath’s graphic image of the boat refitted and repaired while sailing in the open sea.) At this level of consideration metaphysics not merely underpins but also reflects science. [Res00, p.5]

It is the latter, lesser category of metaphysical proposition I am interested in here, and it is this sort of theoretical “at-sea” refit that I am attempting to conduct by considering the metaphysical and epistemological implications of biological theory in the light of our lived experience of scientific practice.

### 14.0.2 Bayesian Epistemological View on Statistical Generalisation

I have chosen to make extensive use of Bayesian statistical methods<sup>2</sup>, which have by now matured into a useable toolkit for the cellular and molecular biologist. While Bayesian statistics are sometimes portrayed as a marginal subjectivist trend, to be ignored by serious biostatisticians, nothing could be farther from the truth.

Under

---

<sup>2</sup>Some would describe them as Laplacian, since Laplace famously provided their first real scientific use in calculating the posterior distribution of the mass of Saturn; Bayes was calculating lottery chances. Properly speaking, Laplacian statistics are the original statistical orthodoxy

## 14.1 Study Structure in the Light of Tradition

A scientific report is, by its nature, a retrospective structuring of an investigation. Structuring a scientific study requires defining the goals and methods of the study. To the extent that the products of a research program (eg. the intellectual output of a number of labs over some period of interaction) are structured by some common set of rules for determining their goals and methods, these rules turn out to be surprisingly difficult to define. As the Kuhnian philosopher of science Philip Kitcher noted about the teaching of classical genetics:

Neophytes are not taught (and never have been taught) a few fundamental theoretical laws from which genetic “theorems” are to be deduced. They are introduced to some technical terminology, which is used to advance a large amount of information about special organisms. Certain questions about heredity in these organisms are posed and answered. Those who understand the theory are those who know what questions are to be asked about hitherto unstudied examples, who know how to apply the technical language to the organisms involved in these examples, and who can apply the patterns of reasoning which are to be instantiated in constructing answers. More simply, successful students grasp general patterns of reasoning which can be used to resolve new cases. [Kit84]

It is the effective use of these patterns of reasoning that constitutes the good or proper practice of biology. The neophyte biologist, if he inquires into the general structure of those patterns, is likely to be met with what Paul Feyerabend referred to as the “fairy-tale” underlying the special credibility afforded natural scientists:

Scientists have ideas. And they have special methods for improving ideas. The theories of science have passed the test of method. They give a better account of the world than ideas which have not passed the test. [Fey93]

This is an appealing myth, particularly for the learner motivated by a search for the truth, which may account for its use in redirecting the student to immersion in experimental practice. We often teach that there is actually only one method, “The Scientific Method” (hereafter Method), which consists in iteratively falsifying hypotheses about phenomenal reality, allowing a model of that reality (in the form of an logical argument constructed of these hypotheses) to be refined over time so that the model is, by correctly applying this procedure, inexorably brought closer to reality<sup>3</sup>.

The conscientious student, looking for serious academic treatments of the Method, is immediately forced to contend with a bewildering array of perspectives on what constitutes the Method and what procedures are legitimately employed in properly Scientific practice. Most of these are produced from accounts of the development of physics and its auxiliary sciences. Many insist that the Method consists of some formal criterion or procedure which is plainly not in use in biology or in scientific practice at large. The wide disagreement on the very concept of any such Method has, at least, the salutary function of disabusing the student of the notion that scientific practice could possibly be governed by the well-understood “special methods” of Feyerabend’s “fairy-tale” scientists.

Deprived of any “special method” recipe to apply to his problem, the biologist only has resort to the patterns of reasoning in their field. These are normally understood heuristically or intuitively- we know

---

<sup>3</sup>This instruction, strangely, often fails to note the origin of this notion with Karl Popper and its subsequent uneven reputation as a good description of, or prescription for, scientific activity. This idea now often goes by the name of “evolutionary epistemology”, with the general idea being that incorrect hypotheses are selected against in a progressive refinement of knowledge. To whatever extent this happens, the selective procedure is not iterative application of a particular methodology or set of rules.

how to make arguments about molecular systems without any training in formal logic, and without necessarily offering any account of *what we are doing* when we are making the argument. This type of understanding is sufficient when one is simply applying these patterns to new cases. What are we to do when confronted with claims by a colleague who is using novel, foreign, imported, etc. scientific methodologies? We may attempt to “muddle through” by “seeing what works”, but this concedes entirely too much to influences we generally understand to be orthogonal to systematisation of knowledge about nature. If “what works” is largely defined by what gets published, the result of “muddling through” without clear method will frequently be shaped more by the parochial interests involved in the scientific process than by the drive for better biology<sup>4</sup>.

The student may therefore return to academic discussion of scientific practice looking not for any recipe-Method, or indeed systematic prescriptions about how to structure studies and analyse data, but simply for points of agreement on what science is and what scientists are doing. As a full survey of this literature would extend across a huge range of studies in a variety of disciplines, I have instead focused on one informative point of agreement arrived at in the philosophy of science. In the second half of the 20th century the observation was made by Thomas Kuhn, Paul Feyerabend, Imre Lakatos, and others, that science plainly did not proceed by iterative hypothesis falsification, as asserted by the more credible scientific realists of the first half of the 20th century. Kuhn, Feyerabend, and Lakatos conceived of different scientific “paradigms”, “traditions”, or “research programmes”, respectively, making competing claims on truth in a historical process of scientific development. Rather than collectively producing one huge cultural artefact called “Science”, it became clear that the only way to explain the apparently “revolutionary” character of science, as well as the chaotic succession of different scientific theories, was to understand science as the activity of people committed to a plurality of different methodologies interacting in particular historical contexts.

These ideas and their relation to one another took time to digest, but by the end of the 20th century, most philosophers of science agreed that scientific theories must be analysed in their historical, social, and intellectual context, and that diverse scientific schools of thought advance theories and models as competing explanations for phenomena, rather than proceeding by any mutually agreed upon Method. Scientists participate in and draw from these schools as they perform their work, convert from one school of thought on an issue to another as the latter becomes more fleshed-out and persuasive, and so on. From a biologist’s perspective, this is a realistic description of scientific practice as a type of ecosystem, rather than the rote application of some logician’s rules. The agreement of the philosophers suggests that this is genuinely a better description of what scientists are actually doing than the quasi-mythological “received view” of the past, particularly in the context of the bitter disputes on virtually every other topic.

I have therefore chosen to define a framework derived from these basic insights in order to concretely examine what I take to be the leading research program in my field. In order to structure this examination, I have chosen to make use of a number of concepts from the philosophy of science: Feyerabend’s “tradition”, Fagan’s “Joint Mechanistic Explanation (JMEx)”, the latter modified by Schaffner’s “ex-

---

<sup>4</sup>For instance, both the academic careers of scientific personnel staffing with granting agencies and editorial boards, and the commercial success of manufacturers selling expensive scientific equipment, reagents etc. depend directly on the perception that the methodology used by the personnel with the equipment and reagents is in some way *orthodox*. This perception may be justified with reference to some philosophical rule to establish the boundary between science and non-science, such as Popper’s “pseudoscience” demarcation. In the particular case of stem cell biology, as Melissa Fagan has noted, even basic conceptual divisions (eg. “adult” vs. “embryonic” stem cells) can become politically charged [Fag13, p.47]. Given the lucre associated with promising approaches to stem-cell based regenerative medicine, the influence of these “orthogonal influences” is particularly strong in the SCBT- the perception that a particular theoretical approach undergirding some therapy is mainstream, well-supported science is critical for the therapy’s adoption.

tended theory". I briefly describe some of the relevant background of these concepts with occasional clarifying examples from our field. I then make modifications to these concepts for my own use and demonstrate how they inform the structure of this study.

### 14.1.1 Feyerabend's Scientific Traditions

The central, seminal insight of Paul Feyerabend's *Against Method* is that the observed succession of scientific theories occurs by counterpositional advancement of incompatible opposing theories, because only counterinductive comparisons *between* theories are capable of showing up their implicit assumptions and allowing them to be challenged. Feyerabend explains:

... it emerges that the evidence that might refute a theory can often be unearthed only with the help of an incompatible alternative: the advice (which goes back to Newton and which is still very popular today) to use alternatives only when refutations have already discredited the orthodox theory puts the cart before the horse. Also, some of the most important formal properties of a theory are found by contrast, and not by analysis. A scientist who wishes to maximize the empirical content of the views he holds and who wants to understand them as clearly as he possibly can must therefore introduce other views; that is, he must adopt a *pluralistic methodology*. He must compare ideas with other ideas rather than with 'experience' and he must try to improve rather than discard the views that have failed in the competition.[Fey93, p.20]

*Against Method* takes as its historical exemplar Galileo's advancement of the heliocentric Copernican model against its orthodox Aristotlean competitor championed by the Catholic Church. Although we commonly think of the so-called Copernican Revolution as the replacement of an obviously defective theological explanation by a properly formed theory from the empirical sciences, Feyerabend shows that this conceals the actual means by which Galileo makes his persuasive case for the (itself badly defective) Copernican model. Centrally, it is process of comparing the heliocentric and geocentric theories that makes the implicit, unstated "natural assumptions" of the geocentric theory clear. Feyerabend elaborates on the kind of structures that counterinduction can reveal:

Methodological rules speak of 'theories', 'observations' and 'experimental results' as if these were well-defined objects whose properties are easy to evaluate and which are understood in the same way by all scientists.

However, the material which a scientist *actually* has at his disposal, his laws, his experimental results, his mathematical techniques, his epistemological prejudices, his attitude towards the absurd consequences of the theories which he accepts, is indeterminate in many ways, ambiguous, and never fully separated from the historical background. It is contaminated by principles which he does not know and which, if known, would be extremely hard to test. Questionable views on cognition, such as the view that our senses, used in normal circumstances, give reliable information about the world, may invade the observation language itself, constituting the observational terms as well as the distinction between veridical and illusory appearance. As a result, observation languages may become tied to older layers of speculation which affect, in this roundabout fashion, even the most progressive methodology. (Example: the absolute space-time frame of classical physics which was codified and consecrated by Kant.) The sensory impression, however simple, contains a component that expresses the physiological reaction of the perceiving organism and has no objective correlate. This 'subjective' component often merges with the rest, and forms an unstructured whole which must be subdivided from the outside with the help of counterinductive procedures. (An example is the appearance of a fixed star to the naked eye, which contains the effects of irradiation diffraction, diffusion, restricted by the lateral inhibition of adjacent elements

of the retina and is further modified in the brain.) Finally, there are the auxiliary premises which are needed for the derivation of testable conclusions, and which occasionally form entire *auxiliary sciences*.

...

Consideration of all these circumstances, of observation terms, sensory core, auxiliary sciences, background speculation, suggest that a theory may be inconsistent with the evidence, not because it is incorrect, *but because the evidence is contaminated*. The theory is threatened because the evidence either contains unanalysed sensations which only partly correspond to external processes, or because it is presented in terms of antiquated views, or because it is evaluated with the help of backward auxiliary subjects.

...

It is this *historico-physiological character of the evidence*, the fact that it does not merely describe some objective state of affairs *but also expresses subjective, mythical, and long-forgotten views* concerning this state of affairs, that forces us to take a fresh look at methodology. It shows that it would be extremely imprudent to let the evidence judge our theories directly and without any further ado. A straightforward and unqualified judgement of theories by 'facts' is bound to eliminate ideas *simply because they do not fit into the framework of some older cosmology*. Taking experimental results and observations for granted and putting the burden of proof on the theory means taking the observational ideology for granted without having ever examined it.

[Fey93, p.52]

Feyerabend argues that it was the cognitive contrast between the Copernican model and orthodox Aristotlean geocentric and geostatic conceptions which showed up precisely this kind of implicit assumption. In this case, the assumption in question gave rise to the unchallengeable "fact" that the Earth could not be moving rapidly through space because objects are observed to fall straight down- if the Earth were in motion, falling objects would appear to be moving in a slanting trajectory.

We start with two conceptual sub-systems of 'ordinary' thought ... One of them regards motion as an absolute process which always has effects, effects on our senses included. The description of this conceptual system given here may be somewhat idealized; but the arguments of Copernicus' opponents, which are quoted by Galileo himself and, according to him, are 'very plausible', show that there was a widespread tendency to think in its terms, and that this tendency was a serious obstacle to the discussion of alternative ideas.

...

The second conceptual system [the Copernican model] is built around the relativity of motion, and is also well-entrenched in its own domain of application. Galileo aims at replacing the first system by the second in *all* cases, terrestrial as well as celestial. Naive realism with respect to motion is to be *completely eliminated*.

...

Viewing natural phenomena in this way leads to a re-evaluation of all experience, as we have seen. We can now add that it leads to the invention of a *new kind of experience* that is not only more sophisticated *but also far more speculative than* the experience of Aristotle or of common sense. Speaking paradoxically, but not incorrectly, one may say that *Galileo invents an experience that has metaphysical ingredients*. It is by means of such an experience that the transition from a geostatic cosmology to the point of view of Copernicus and Kepler is achieved.

[Fey93, p.69]

In other words, the conceptual frame of the Galilean observer has shifted so that fundamental, common-sense natural perceptions (objects which are not observed to be in motion cannot be in motion)

are overturned, and “experience now ceases to be the unchangeable fundament which it is both in common sense and in the Aristotelian philosophy. The attempt to support Copernicus makes experience ‘fluid’ in the very same manner in which it makes the heavens fluid, ‘so that each star roves around in it by itself’. An empiricist who starts from experience, and builds on it without ever looking back, now loses the very ground on which he stands.” [Fey93, p.72]

Thus, the Copernican Revolution succeeded because the counterinductive comparison of the older Aristotelian explanation with Galileo’s theory revealed the flawed natural assumption of the geostatic model- the assumption that all motion is operative, that objects not in apparent motion for some observer are indeed at rest in an absolute sense. Only by thinking of motion as a relative phenomenon that only produces observable effects when bodies are moving *with respect to one another*, does the problem with the assumption of absolute motion become obvious.

This success was not the result of any of the usual “rules” offered as candidates for the mono-Method. Galileo’s theory substantially contradicted the available evidence, erroneously asserted the reliability of telescopic observations, made extensive use of ad hoc hypotheses, and was advanced by propagandistic and even dishonest means. Much of this was unavoidable. Ad hoc hypotheses are necessary for new theories because the auxiliary sciences associated with them have not been developed- scientific development is intrinsically uneven, obligating the use of these makeshift theoretical devices. Without a certain level of dishonesty and rhetorical sleight of hand on Galileo’s part, the Copernican program would have succumbed to the greater development and argumentative weight of the scholastic tradition. As Feyerabend notes, if any of the typically suggested Method criteria were applied, the Church would have won the debate and we might still have an Aristotelian cosmology!

For Feyerabend, then, science, like other human social practices (including the arts, administration, and so on) consists of a variety of interacting traditions with differing assumptions, methods, sensory interpretations, and so on. The development of science is thus “not the interaction of a practice with something different and external, *but the development of one tradition under the impact of others.*” [Fey93, p.232]

#### 14.1.2 Classical and Molecular Genetics: Shared Traditions, Plural Methodologies

To bring the implications of Feyerabend’s views into some familiar relief, let us briefly examine a case from the field: the “molecularisation” of zebrafish genetic experimentation. Because the zebrafish genome took some time to be released in reliable revision, the use of the Mendel/Morgan “classical genetic tradition” (CGT) was widespread in mapping experiments before this. One might locate an allele of interest by recombination experiments, mapping its position in (by now unfamiliar) units of centimorgans to reflective relative recombination frequency between loci. This is, of course, increasingly unusual, as the practices of the Watson/Crick “molecular genetic tradition” (MGT) are more and more accessible due to the previously unavailable genomic data<sup>5</sup>. The zebrafish geneticist now has access to two traditions to perform their work, and will make recourse to either as they assess circumstances warrant (without any reference to an explicit external rule about which tradition should supercede the other in particular circumstances). For instance, if I am planning to cross two fish that are heterozygous at some allele, I

---

<sup>5</sup>It can be an edifying experience to ask a student to “convert” between CGT units like centimorgans and MGT units like megabases. At the very least, the cognitive discomfort produced will tend to reveal a problem in the question’s assumptions to the student. At best, one can gain a practical understanding of how scientists make sense of and experience “switching” between different traditional explanatory frames.

will use typical CGT assumptions to calculate the expected number of various offspring genotypes. Even if I was able to offer a reasonable description, in molecular terms, of zebrafish meiosis, recombination, fertilisation, and so on, I would never choose to do so- an MGT explanation is unnecessary where Mendel's laws suffice, and will begin to lack explanatory power the moment I apply it to another species, unlike the Mendelian CGT framework.

One might object that the claims of the CGT tradition “reduce” in some meaningful way to the MGT tradition, or “cover” for it, so that CGT theories are in fact just abstractions of MGT theories. In other words, the fact that MGT claims are not useful and lack explanatory power in some areas does not change the fact that MGT *could* provide good theoretical explanations for all of the claims of CGT. As Kitcher shows, this is not correct<sup>6</sup>. The CGT description of the *process* of meiosis as a particular type in which “paired entities (i.e. chromosomes) are separated by force so that one member of each pair is assigned to a descendent entity”[Kit84, p. 349] (which Kitcher calls a PS-process), is not describable solely in terms of the molecules participating in the process. Because a PS-process like meiosis can be realised in any number of ways at a molecular level, PS-processes cannot be accounted for in a general way by MGT theories. The manner in which different traditions frequently address different levels of biological organisation in similar non-interchangeable ways is discussed further below.

We can extend this example to see that the typical biologist has access to a great plurality of methodological traditions. Biologists are intuitively used to “translating” between traditions and to switching freely between them when translations are cumbersome or useless. Where there are no deep conflicts between the implicit aspects of these traditions, no counterinductive process occurs. That is, we do not “refute” CGT with MGT in the same way the Copernican Revolution “refuted” Aristotlean geostatic theory because MGT rarely makes the structure of CGT seem untenable to us; an attempt to replace CGT with MGT would surely be quixotic. We simply appropriate the methodology of the tradition which is most appropriate for the problem at hand.

### 14.1.3 Sense of “Tradition” in this study

Thomas Kuhn was soundly criticised for the vague sense in which he used the term “paradigm”[Sch93, p.206], which left him unable to offer coherent explanations. Feyerabend was far more precise, but, as a philosopher rather than a practitioner, never had need (or desire) to guide a scientific project from within the worldview he developed. Having erected an unassailable proof that no special Method existed, and that subscribing to one would have prevented the central Galilean vignette of naive philosophy of science's core Enlightenment self-narrative, he effectively retired<sup>7</sup>.

---

<sup>6</sup>Kenneth Schaffner later insisted that Kitcher was wrong, and that CGT theories are reducible to MGT theories, because the MGT theories establish direct linkages between DNA sequences and proteins, and thus are formally equivalent to the classical concepts of genes and phenotypes [Sch93]. This now seems obviously mistaken on at least 2 counts: (1) there is no stable molecular entity implicated by “the gene”: its local context in terms of primary sequence (promoters etc.), and tertiary arrangement in a nuclear “transcription factory”, all contribute to the complexity of the causal locus involved in transcription alone, and (2) there are very few phenotypes of interest that are caused by stable, uncomplicated interactions between DNA and protein such that classical genetic phenotypes reduce cleanly or usefully to descriptions solely in terms of macromolecules. Moreover, this does not address Kitcher's point: there *are* phenotypic traits which are *not* describable in terms of their protein constituents, and that there *are* higher-level processes which may be generated by any number of configurations of molecular constituents and are thus better understood as types of processes and not as specified molecular systems.

<sup>7</sup>Feyerabend's chilly reception and subsequent exile is difficult to understand in retrospect, as the basic framework of his argument is now generally understood to be correct in some sense or another. Whether one ascribes the initial poor reaction to his combative personality or flamboyant political statements, his ideas are by now pervasive, even normative. The appearance of “post-colonial” biological studies would have been very unlikely in the Popperian scientific world of the 1960s. Curiously, the Feyerabendian molecular biological traditionalist will note that this may actually constitute an

When Feyerabend wrote of “scientific traditions”, he meant scientific practice in its broadest possible sense, a social practice with “historico-physiological character”. That is, scientific theories are conditioned by the historical moment, and by the involvement of the observer and their social, cognitive, and biological baggage in the phenomenon under study. For the purposes of making Feyerabend’s arguments against the existence or advisability of a mono-Method, this was entirely adequate, but I intend only to deal with a limited subset of these materials. The “Stem Cell Biological Tradition” defined below, in its full Feyerabendian sense, includes rumours circulated at stem cell conferences, methodological tips and tricks passed on by observation of careful practice (that could never be learned from any protocol), the commercial aspirations of particular academics, and so forth, in addition to all of the “properly scientific” materials implicated by all of the publications that have ever used the “stem cell” concept or a recognisable homologue. While my use of the term acknowledges this broader contextual sense, I will only look at the explicitly documented “traditional background” of particular, formally advanced explanatory statements or models. The objective here is mainly to account for molecular biological explanations occurring within a “composite tradition” with diverse, only partially overlapping internal “subtraditional” currents.

Therefore, in general, I have not intended to draw any hard divisions between subtraditional currents in the molecular biological tradition (MBT). While I have pointed out how MBT explanations may include components drawn from irreducibly separate subtraditions (the irreducible CGT explanation for PS-processes discussed [above](#)), I have only mentioned the philosophical, methodological, etc. implications of MBT subtraditions where necessary, as the primary concern is the extra-traditional background of the “systems” modelling approaches encountered in the data chapters. For instance, in the case of the Galton-Watson branching process model, my definition of this as an artefact (one can think of a probability distribution plot figure in a paper) from the Probability Theory Tradition (PTT) means that this is a traditional statistical solution to an early biological lineage model, which stands in for the large-scale behaviour of a simple model of lineage growth and extinction, in a larger MBT mechanistic explanatory framework (the [MEx](#), Mechanistic Explanation).

I have also made reference to the “Stem Cell Biological Tradition” (SCBT), which is intended to convey the last approximately 6 decades of research practice organised around the stem cell concept<sup>8</sup>. The question of the relationship between the MBT and SCBT is not simple. Till and McCulloch were not offering molecular mechanisms for the phenomena they were documenting [[MT60](#)], and could well have been described as operating in an earlier Cell Biological Tradition (CBT).

My conception of this relationship is that, generally, SCBT practice has been “molecularised”; most of the explanations that SCBT practitioners care about are molecular MEx, and the SCBT is therefore a subtradition of the MBT. Nevertheless, the [multilevel](#) character of all MBT explanations means that SCBT practitioners routinely use CBT concepts where a specified molecular system would be useless or inappropriate. An example is the ubiquitous abstraction of mitotic phenomena as a class of cell behaviour in SCBT explanations (referred to simply as “mitosis”, “proliferation”, “renewal” etc), rather than making any attempt to specify this central concept in explicitly molecular terms.

Therefore, in this sense, SCBT practitioners are used to “slotting in” explanatory components from a variety of traditions in order to offer a broader MEx for some phenomenon (regeneration, repair). We are in this sense classic Feyerabendian “Epistemological Anarchists”, using whatever explanatory

---

intra-traditional “colonial” venture on the part of critical theorists and continental philosophers.

<sup>8</sup>I include anti-stem cell and stemness theories in this tradition, since they are marginal heresies to a hegemonic orthodoxy and not meaningfully independent.

frames suit our purposes. This habit has, in the context of the SBE, shifted from predominantly “intra-” to “inter-traditional anarchism”, with the result that the required level of sophistication required to advance a good molecular MEx has increased steeply.

The SCBT is a particularly notable subtradition of the MBT because of its explicitly therapeutic, and therefore ethical and normative orientation. That is, while the broader developmental biology tradition does not have a unified therapeutic goal, the SCBT is structured by the objective of interacting with stem cell phenomena to medicinal and therapeutic ends. This particularly ethical aspect of the SCBT compels me to make some brief remarks about how my own study’s design is be guided by this ethical orientation.

## 14.2 EHJMEx as a Subtraditional Metascientific Unit of Molecular Biological Theory

From the practitioner’s perspective, we are interested in thinking about the “metaphysical unit” of scientific practice on a finer scale than allowed by Feyerabend’s concept of a tradition. If we take the practice of molecular biology to be about generating explanations for biological phenomena in terms of macromolecules and their organisation, we want to know enough about the structure of these explanations to allow us to compare them. While we know that we are looking for the “underlying molecular mechanisms” of (usually) cellular phenomena, it is not always clear how we might go beyond intuitive preferences for one proposed mechanism or another in, say, deciding what assay to perform next. We may reason in terms of plausibility or agreement with evidence, but we rarely do so formally<sup>9</sup>. Moreover, it is sometimes unclear what kinds of descriptions are mechanistic explanations (MEx) and when we can agree such a description has explained some phenomenon.

A deliberate counterinductive evaluation of two “units” of biological theory that is productive- one that allows us to imaginatively explore the metaphysical implications of two scientific models for explaining some phenomenon- requires a definition of the unit. Feyerabend’s description of scientific tradition suggests that we start by looking at “local” traditional practice. I have thus chosen here to rely heavily on a general conception of the “locally accepted method” proffered by the philosopher of stem cell biology, Melinda Bonnie Fagan (Fagan hereafter), who has extremely usefully summarised these ideas in her recent book “Philosophy of Stem Cell Biology: Knowledge of Flesh and Blood”, as well as a number of other useful publications[Fag13, Fag15b, Fag15a]. I have modified her idea by extending it temporally (following Schaffner’s Extended Theory concept[Sch93, p.211]) and by placing it within the what I take to be the extended context of the Molecular Biological Tradition conceived of in Feyerabend’s terms- explicitly including its auxiliary sciences and metaphysical implications.

### 14.2.1 Law-based and Causal Mechanistic Explanations

Fagan is not the first philosopher of science to note the local currency afforded to MEx in the MBT. Indeed, she has thoroughly addressed the inadequacies of earlier theorisations of biological MEx. In general, philosophers have previously advanced law-based and causal theories to explain what biologists are doing when they offer MEx.

---

<sup>9</sup>It may not even be advisable to do so in many cases, as, for instance, we often become aware of problems with particular lines of research which are not disclosed in published reports.

### 14.2.1.1 Law-based MEx

Law-based theories suggest that MEx operate by explaining a phenomenon in terms of universal, exceptionless laws of broad scope, given some initial set of conditions. I take it to be obvious that this is not what biologists are doing- MEx simply do not express this kind of law, and I have personally never encountered a biologist who understood MEx in anything like “universal, exceptionless” terms, even as a “regulative ideal” toward which MEx are asymptotically refined. Mechanistic diversity underlying broad classes of phenomena is understood to be normative, so “nomological” law-based concepts are of little use to us.<sup>10</sup>

### 14.2.1.2 Causal MEx (CauMEx)

Theories of MEx that are underlain by some understanding of causal relations described by the mechanism are more influential than law-based ones within biological practice. The most typical theory of causality taken to underlay these “Causal MEx” (CauMex, com-ex) is an interventionist one, typically expressed in Fagan’s description of “manipulability theory”. Indeed, biological studies are often divided between “interventional” and “non-interventional”, largely along the lines proposed by James Woodward. As Fagan notes of Woodward’s description:

This theory analyzes causality as a relation between values of variables, X and Y. X causes Y if and only if there is a possible manipulation of some value of X, under idealized experimental conditions, such that the value of Y changes. Woodward’s is a counterfactual account of causality, hinging on what *would* happen to the values of variables in an idealized experiment, or *intervention*. An intervention I on variable X with respect to variable Y is a causal process that determines the value of X in such a way that, if the value of Y changes, then the change in Y occurs only in virtue of the change in X.

The concept of an intervention provides a regulative ideal for experiments aimed at discovering causal relations: approximate the conditions of ideal experimental interventions. Insofar as they meet this standard, experiments reveal genuine patterns of counterfactual dependence among sets of entities and their properties, i.e. causal relations in the world. Because invariance [in the correlative relationship between X and Y] is required only under *some*, not all, interventions, causal explanations need not include general laws. The range of interventions under which a dependency relation between values X and Y holds may be broad or narrow. Within their range, ‘fragile’ dependency relations are no less causal than those with a much wider range of invariance: universal causal laws. Experiments, if correctly designed, reveal relations between values of variables that are invariant under some interventions. The relevant counterfactuals for causal relations are tested by realizing particular values of X and observing the values of Y under controlled circumstances that fall within the range of invariance for X and Y. [Fag13, p.99]

I quote this description at length because it well-describes the essential orthodox view implied when we teach students the canonical experimental typology of the MBT. This is a highly plausible account

---

<sup>10</sup>Fagan refutes recent advocacy of the law-based view. Schaffner previously advocated for a more sophisticated conception in which biological explanations were only “universal” or “exceptionless” in the specific context where they were formulated (species, tissue, cell type, etc). Biological explanations might then consist of laws which support counterfactual refutation. As Fagan points out, MEx are not even expected to be locally universal explanations for phenomena. One of the best-understood MEx, that of asymmetric mitosis in drosophila Germline Stem Cells (GSCs), only applies in full to about 80% of GSC divisions which “have a plane of division parallel to the hub-GSC binding site.” [Fag13, p.97] A proponent of the idea that biological theories make use of “laws” might have resort to “legal pluralism”, in which they assert that more than one law may pertain to a particular case, depending on some contextual factor. I doubt a concept of this sort can really be any more useful than Fagan’s Joint MEx description adopted here.

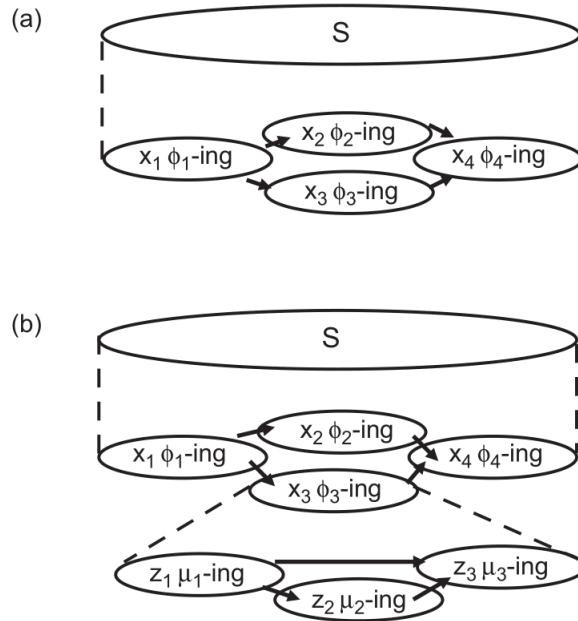


Figure 14.1: The consensus view of mechanisms, excerpted from [Fag13, p.96] (a) Two levels: component x's and overall mechanism S. (b) Downward expansion: three mechanistic levels.

of how these experiments “work” to explain the underlying system. We can think of a typical dose-dependence curve in pharmacological research: we apply some hormone over a range of physiologically-reasonable doses X to some cells, and observe the response of some functional readout Y (say, the cell-normalised luminance of a luciferase reporter downstream of a receptor for X). If there is a invariant relationship between X and Y over some reasonable dose range, we may postulate that the application of dose X causes the luminance of the cells Y by resulting in hormone binding, activation of a promoter element, etc. If we perform a series of other interventions we can test counterfactuals about this relationship: we may try to observe the movements of GFP-tagged signalling proteins during the application of dose X, or apply inhibitors of enzymes, antagonists of receptors, and so on, to implicate the other elements involved in the causal chain we suppose makes up the mechanism. Therefore, CauMEx do seem to be describing useful features of our explanatory practices in the MBT.

Fagan provides this formulation (M) to summarise the conventional understanding of this view:

(M) A mechanism S consists of multiple diverse components (x's) engaging in causal relations or activities ( $\phi$ 's) such that x's and  $\phi$ 's are spatially and temporally organized so as to produce some overall phenomenon. [Fag13, p.95]

Of course, mechanistic explanations are multi-level, so components may themselves be composed of sets of these causal relations, and the overall mechanism S may in turn be a component of a higher level set of relations which describes a broader S, and so on. This overall view of CauMEx explanations is summarised in Figure 14.1.

Of this view of CauMEx, Fagan writes:

An account is “full” if and only if it describes all *relevant* components of a mechanism. Causal relevance is in turn analyzed in terms of “mutual manipulability”. A working component (x  $\phi$ -

ing) is relevant to an overall mechanism's behavior ( $S \Psi\text{-ing}$ ), if the latter can be manipulated by intervening on  $x$ 's  $\phi\text{-ing}$  and  $x \phi\text{-ing}$  can be manipulated by intervening on  $S \Psi\text{-ing}$ . A component is irrelevant to a mechanism if neither  $x$ 's  $\phi\text{-ing}$  nor  $S \Psi\text{-ing}$  can be manipulated by intervening on the other. These two sufficient conditions link the hierarchical structure of MEx to experimental practices in neuroscience, which employ both 'top-down' and 'bottom-up' strategies to detect components of mechanisms. [Fag13, p.100]

Therefore, we seem to have an appealing account of what biologists are doing with MEx that is neatly connected to our experimental practices. However, Fagan describes three fundamental problems with CauMEx.

1. **Ambiguity.** The overall mechanism  $S$  is a causal explanation: mitosis ( $S$ ) is a process of cellular reorganisation resulting in two cells being produced from the one mitosing (the phenomenon  $P$ ) (Fig. 14.2 (a)). We infer the contribution of  $S$ 's components  $x$  (eg. proteins involved in spindle formation, cytoskeletal components, etc) to  $P$  by examining the effect of altering their activities  $\phi$  (such as by mutating a spindle formation protein of interest) on  $P$  (Fig. 14.2 (c)). However, usually, we take the overall explanatory bricolage we assemble of inter-component causal relations to explain  $S$ , not only  $P$ . That is, the proteins involved in spindle formation are part of a mechanism that explains the mechanism of the mitotic process itself ( $S \Psi\text{-ing}$ ), and not only its phenomenal manifestations while  $\Psi\text{-ing}$  (chromosomal reorganisation, the appearance of mitotic furrows, etc.) or after  $S$  has finished  $\Psi\text{-ing}$  (the fact of there now being two cells).
2. **Directionality.** The assumed *direction* of causal relations in the MBT is usually implied in our casual speech: we are looking for "underlying" molecular mechanisms, which is to say, we take any "chain" of causal relations to flow upwards from a bottom level of macromolecular components. Generally speaking, "overarching" mechanisms in which the whole  $S$  feeds back onto its component  $x_s$  in some way are not produced. Moreover, it is unclear in what sense one could impinge on the overall mechanism  $S$  and "cause" a change in one of the components  $x$  in a "downward" sense-as Fagan lucidly notes, intervention into a mechanism *is* impinging on its components  $x$  [Fag13, p.103]. We are establishing a "constitutive" relationship between  $S$  and its components  $x$ , not a causal one. This makes a nonsense of the "mutual manipulability" interpretation of top-down and bottom-up approaches, since there is no manipulation of  $x$  by virtue of an intervention into  $S$  as a whole.<sup>11</sup>.
3. **Modularity.** In order for components to be defined in terms of manipulability theory, they must be independent. If we offer a mechanism consisting of generalisations about the behaviour of  $x_{1,\dots,n}$  macromolecular processes, in order for  $x_1$  to be differentiable from  $x_2$ , it must be in some way causally independent- we must be able to intervene into  $x_1$  (eg a generalisation about DNA synthesis) without affecting  $x_2$  (eg a generalisation about membrane fusion complexes). Without some form of "modular independence" between mechanism components, there is no reason to consider them separate, yet causally linked, "components" of a larger overall mechanism.

However, as Fagan notes, the "interventionist" causal paradigm implies explanations that "reveal what a phenomenon of interest depends on, answering 'what-if-things-had- been-different?' questions about the values of variables within a fixed range of invariance. But MEx are *prima facie* concerned with a different kind of question: 'how-does-it-work?'" [Fag13, p.106] In this case, the

---

<sup>11</sup>As Fagan correctly notes, top-down experimental approaches are *exploratory*, not *explanatory*.

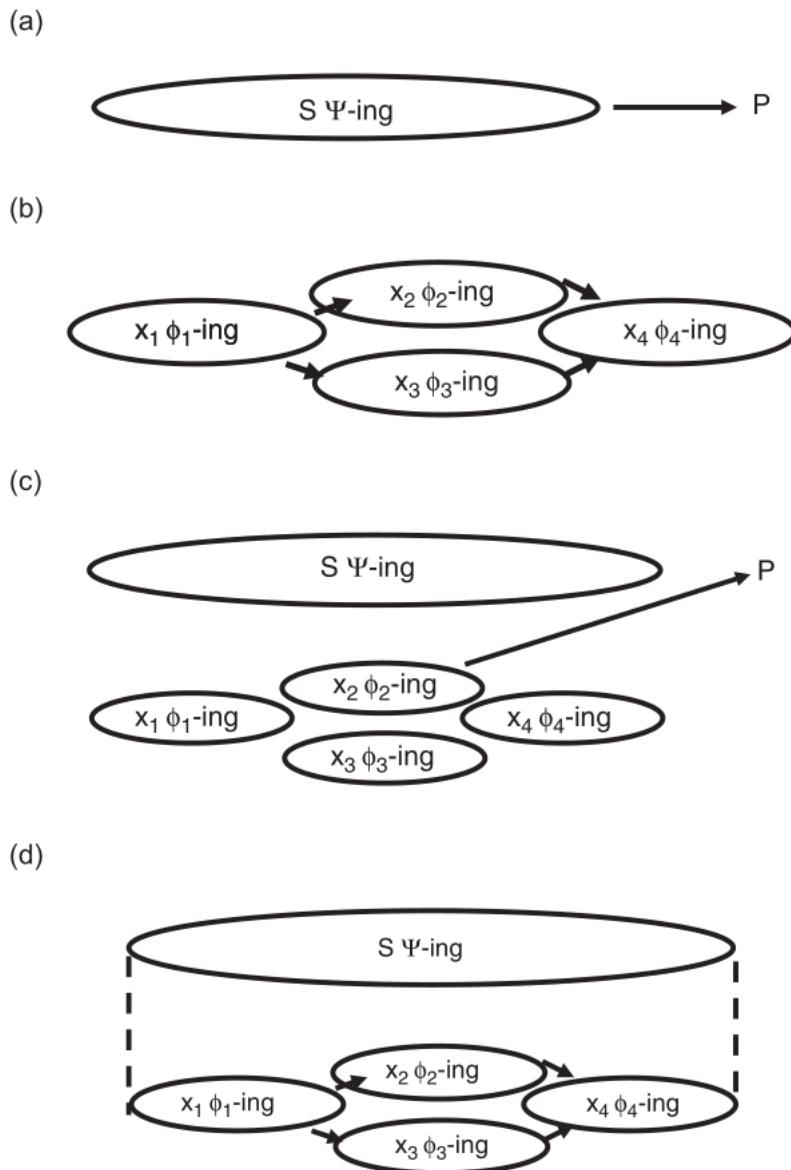


Figure 14.2: Levels of causal relations in MEx, excerpted from [Fag13, p.102]. (a) System-level: mechanism  $S \Psi\text{-ing}$  produces phenomenon  $P$ . (b) Component-level: component  $x$ 's causally interact ( $\phi\text{-ing}$ ) to effect one another. (c) Inter-level: component  $\phi\text{-ing } x$ 's make a difference to  $S$ 's outcome ( $P$ ). (d) Constitutive MEx: component  $x$ 's interact ( $\phi\text{-ing}$ ) to constitute mechanism  $S$ 's  $\Psi\text{-ing}$

causal relations that form “the ways we conceptualize them as distinct, relevant components ... *are the very same* causal generalizations that figure in MEx.” [Fag13, p.106] That is, the causal relations that provide the rationale for the division of a mechanism into independent components are *also* those that we take to be causing the activity of the mechanism- the components are thus not merely *conceptually* but “*actually*”, *causally* separable, like components of a machine. Obviously, we understand this not to be the case. Fagan concludes:

For physiological, cellular, and molecular mechanisms, as we understand them, the behavior of isolated components is not a good guide to their behavior together, and their behavior in one context is not a good guide to their behavior in others. If this line of reasoning is correct, then the causal-mechanical account of MEx leaves out something important: *interdependencies* among components of biological mechanisms. [Fag13, p.106]

As most modestly observant MBT benchworkers have noticed by now, components of biological mechanisms are not literally interchangeable gears that can be removed from one machine and slotted without modification into another “geartrain” (mechanism) missing a similar component<sup>12</sup>. *How*  $x \phi$ s depends on *how* it is intermeshed with other xs and what their  $\phi$ ing consists in. A gear only depends on the other components to the extent that it requires their presence and their intermeshing in one particular way within some tolerances<sup>13</sup> Indeed, the nature of the interdependencies between molecular components are critical to the “how-does-it-work” explanation. As indicated in the section on hierarchy theory, understanding these interdependencies is especially important in “multi-level” explanations, where xs  $\phi$ -ing at different dynamical scales do not *directly* interact in the same manner that same-level xs  $\phi$ -ing do.

Therefore, Fagan has gone on to advance what is, in my view, a much improved conception of biological MEx, the Joint MEx, to which I have made some minor additions, as described below<sup>14</sup>

#### 14.2.2 (Extended) Heterogenous Joint Mechanistic Explanations ((E)HJMEx)

While by “Extended Heterogenous Joint Mechanistic Explanation” I mean a unit of theory consisting more or less of a diachronic series of Fagan’s “Joint MEx”, I have a slightly different understanding of the “explanatory components” of JMEx (jay-mex), which I have differentiated with the term HJMEx (hedge-mex). I will first clarify how I understand Fagan’s JMEx improvement over the Law-based and

---

<sup>12</sup>There are a few theorists who still don’t understand this, but the situation has improved. We are, now, only rarely subjected to the hallucinations of “synthetic biologists” insisting that trees will be reprogrammed to grow chairs directly on their branches, for instance. The persistence of this fixation can be attributed to the mind-boggling success of the standardisation and modularity practices of modern logisticians, and their mechanical and electrical engineering colleagues. It is worthwhile to remember these are extremely recent developments- an artefact as modern as the Rolls Royce Merlin engine, the machine that won the Battle of Britain, did not have interchangeable parts (to the profound horror of Allison engineers in the USA tasked with setting up domestic American production of the engine). Each part was literally hand fitted for each engine by an English craftsman. We should note that the very English metaphor for biological “design practices” is “tinkering” and not “process systems engineering”.

<sup>13</sup>There must be at least  $b$  microns of backlash for the gears to  $\phi$  together (rotate mutually around their axes), or they will lock and abrade each other when the machine applies force by  $\Psi$ -ing, until abrasion establishes  $b$  microns between them and they are free to  $\phi$  or the mechanism fails (probably jammed with powdered gear teeth). Biological components “intermesh” in a much greater diversity of ways and with far looser coupling (“slop”) than mechanical ones, although every real mechanism needs a little slop in it, as the gear example indicates.

<sup>14</sup>Fagan is an explanatory pluralist herself; in [Fag15b] she explicitly states that she does not seek to replace “causal mechanistic” conceptions with her own JMEx. However, the conceptual problems with causal explanations she catalogues in [Fag13] remain unresolved. If there are “good” CauMEx, it seems to me they need a much deeper theory of causality than interventionist accounts provide. For the time being, Fagan’s work is adequate to stand on its own as an account of the “local explanation” in the SCBT.

Causal concepts. I will then define HJMEx and then explain what it means to “extend” one for analytical purposes, producing a EHJMEx (edge-mex)<sup>15</sup>.

#### 14.2.2.1 Joint MEx (JMEx)

Fagan has described a similar improvement to the typical Causal MEx view on three occasions: first in 2012 [Fag12], later in her 2013 book [Fag13], and most recently in a 2015 paper [Fag15b]. I have adopted the general sense of the framework described in these works, which I have called the JMEx, for Joint Mechanistic Explanation.

Fagan’s terminology varies as she emphasizes different aspects of her overall idea. [Fag15b] seems to use “collaborative” interchangeably with “joint,” which was the preferred term in [Fag13]. “Collaborative” seems to be intended to emphasize the amenity of the concept to explanatory pluralism- this is one reason I have selected it, but it is less descriptive of MEx structure. I have therefore adopted the shorter “Joint”, since a commitment to explanatory pluralism *within* MEx is part and parcel of the the entire foregoing description of the MBT, and respecifying it is redundant.

Fagan indicates that both CauMEx and JMEx are species of “Constitutive” MEx, which is to say that the mechanism consists of components that are together taken to constitute some kind of productive causal regularity which generates a phenomenon. Because a mechanism’s components must be linked *somewhat* if they are not just to be considered separate systems, some theory of causality is required to establish what these linkages might consist in. JMEx’s account of “intermeshing properties” is sufficiently general that MBT practitioners with different ideas about causality can talk to each other about these properties without needing to carefully specify either a particular local causal account with significant problems (manipulability theory) or a fully worked-out global theory of causality.

Fagan’s solution to the problems with Causal MEx dispenses with the problems her analysis of them revealed. Let us recapitulate Fagan’s formulation of Causal MEx (M) and follow it with her JMEx formulation (JM) for comparison<sup>16</sup>.

(M) A mechanism S consists of multiple diverse components ( $x$ ’s) engaging in causal relations or activities ( $\phi$ ’s) such that  $x$ ’s and  $\phi$ ’s are spatially and temporally organized so as to produce some overall phenomenon. [Fag13, p.95]

(JM) Components  $x_1, \dots, x_n$  jointly  $\Psi$  as mechanism S if and only if

- (i) each  $x_i$  has properties that mesh with one or more other  $x_i$ ,
- (ii)  $x_1, \dots, x_n$  are spatially organized and their activities  $\phi_1$ -ing,  $\dots, \phi_m$ -ing causally organized in virtue of their meshing properties,
- (iii)  $x_1, \dots, x_n$  and  $\phi_1, \dots, \phi_m$  so organized constitute mechanism S  $\Psi$ -ing, and
- (iv)  $x_1, \dots, x_n$  and  $\phi_1, \dots, \phi_m$  not so organized do not constitute S  $\Psi$ -ing

[See footnote 16 for description of clarity edit made to (JM)] [Fag15b]

Fagan elaborates on what would therefore constitute a good JMEx:

(JM) suggests a basic norm for what I shall term ‘collaborative explanation,’ namely, that a successful explanation of this kind must show that the components and system of interest

---

<sup>15</sup>While “fanciful” naming conventions and illegible acronyms are time-honoured traditional practices of the MBT, I apologise to readers for indulging out of a need to conserve space and stably refer to particular concepts.

<sup>16</sup>I have used [Fag15b] (JM) as it contains the [Fag13] definition but is more precisely worded. I have changed “system M” to “mechanism S” to conform with the other citations of [Fag13] used here, but this does not change the meaning in any way.

(the target of the model) satisfy the conditions of (JM). Given a set of components and their associated activities, and an overall system exhibiting some behavior, an explanatory model of the latter should describe:

- (i) the properties of components that allow them to interact in specific ways (I.e., meshing properties, as well as other conditions, such as spatio-temporal proximity.)
- (ii) the spatial and causal organization determined by these interactions
- (iii) how the overall system behavior is constituted by the organized components, such that
- (iv) components' organization makes a difference to the overall system behavior.

[I have expanded Fagan's footnote related to item (i) as a parenthetical comment] [Fag15b]

This strikes me as an entirely adequate definition for MEx, as well as a prescriptive norm for “good” or “complete” explanations that provides a reasonable broad guide to what sort of things MEx should eventually be able to do for the system they explain. Moreover, with JMEx, Fagan resolves all three problems with CauMEx simply by replacing manipulability theory with her conception of Joint relations between MEx components, which I find a highly convincing demonstration of the conceptual adequacy of Jointness.

1. **Ambiguity** is resolved. The component xs of JMEx are understood to be interdependent and linked by descriptions of their intermeshing properties expressed while  $\phi$ -ing (eg. binding, activation, phosphorylation, and so on). It is the assemblage of xs, linked by virtue of these properties, that constitute mechanism S  $\Psi$ -ing, so that we are explaining this overall activity of S with our JMEx, which in turn is understood to produce P.
2. **Directionality** is clarified. JMEx “bottom out” at the level of macromolecules, in general, and do not typically reference “lower levels” in the scale hierarchy, like atomic or electronic phenomena. We understand that when we intervene experimentally into the activity of some system S that a mechanism describes, we are not having a “downward” effect on the mechanism’s components x in their  $\phi$ -ing, but that intervening into S *constitutes* intervening into some x, all the way down levels of organisation to the MBT’s “explanatory foundation” macromolecular level.
3. **Modularity** is dispensed with. Given the “Jointness” account of the intermeshing properties of components x of JMEx, he rationale for their “separateness” does not depend on manipulability theory. Therefore, we are able to account for the interdependency of the mechanism’s components and avoid implying that we believe we can intervene into one component in a manner that causally isolates it from the others.

Having summarised Fagan’s JMEx concept and its fine qualities, let me pause to point out an interesting feature of JMEx: the “jointness” or “collaborative” account of causal relationships between JMEx components is adapted from social scientific accounts of *shared cooperative activity* (SCA), that is, the intentional, end-direct activity of multiple agents collaborating. As Fagan indicates, her analysis of joint activities “is modeled on (Bratman 1999)[[Bra99, pp.93-108]], with adjustments for non-intentional contexts.” [Fag15b] I return to some aspects of Bratman’s description of inter-human SCA for useful ideas about agency later, but overall, Fagan’s adjustments make perfect sense. She has chosen to “strip down” an account of joint action between people in order to produce one that fits, say, signalling pathways, which clearly have no psychological states, and of which it is meaningless to speak of “intent”. Moreover, the “stripping down” leaves an excellent skeleton which can be “refit” in a conservative, non-psychologising

manner that admits biosemiotic descriptions of cellular agents, at the appropriate levels and with regard to the appropriate “meshing properties”.

This makes JMEx a spectacularly good type of explanation for multi-level mechanisms that incorporate *bona fide semantic agents* like cells. We expect to be able to express JMEx for higher-level phenomena like tissue formation because we can plausibly represent the joint, coordinated activity of the cells that constitute the tissue as agents.

I take a substantially different view of how “systems biology” is related to JMEx, however, which I have explained below. Firstly, while Fagan is clearly a sophisticated philosopher of biology, and a masterful documenter of the “local model terrain” in the SCBT, her particular vision of JMEx that has a two limitations for my purposes. I have addressed each limitation in turn, in the context of my proposed solutions: the “H” and the “E” (proceeding outward from the core JMEx concept) of the EHJMEx.

#### 14.2.2.2 Heterogenous JMEx (HJMEx)

The first of these limitations to JMEx, as presented by Fagan, is related to an epistemological and metaphysical stance shared with the influential master historian of the MBT Evelyn Fox Keller, who offered her own thumbnail sketch of the SBE some 18 years ago:

“Out of the wickedness of war,” wrote Warren Weaver in 1949, in a paper entitled “Problems of Organized Complexity,” “have come two new developments ...of major importance in helping science to solve these complex twentieth-century problems.” The first of these was the electronic computer, built to process the masses of data generated by the procedures of modern warfare—and, perhaps most famously, to decipher enemy messages encoded in ever more elaborate encryption devices. The second development is most commonly associated with the word *cybernetics*, Norbert Wiener’s term for the study of control and communication in machines and living beings. Extrapolating from his experience with “goal-oriented” and “self-steering” devices designed to improve the accuracy of anti-aircraft artillery, Wiener and his followers envisioned the construction of purposive machines that would resemble living organisms in every way. Indeed, these machines would be built on the very principles of circular causality (“in which every part is reciprocally both end and means”) that Kant himself had invoked as the defining feature of the organism.

These two developments were clearly related—at the very least, they were related in time, in place, and in the needs from which they arose. Yet despite their persistent conjoining in the popular imagination, despite Wiener’s own hopes, and despite even John von Neumann’s efforts at integration, conspicuous differences between the two remained. In the one the emphasis was on computational power, while in the other it was on principles of organization and—increasingly over the 1950s and 1960s—of self-organization. In fact, it was not until the 1980s that the different visions embodied in these two developments would begin to resolve, and the first steps of that resolution came with the rise of connectionism, parallel processors, and neural networks. Yet Jacob’s claim that the sequence of DNA could serve as Bernard’s “invisible guide” depended absolutely on joining together these two still-disparate developments. His metaphor of a program drew directly from Turing’s original model of a computer (the reader may recall from Chapter 1 his equation of “the genetic material with the magnetic tape of a computer”), but the idea of a purposive machine was borrowed from Wiener’s cybernetic vision. The difficulty is that, in locating the program in the genome, much of the cybernetic vision of goal-seeking and self-organization was lost. And so was the recognition of the importance of reliability and with it, an appreciation of the kinds of organizing principles that would be needed to maintain such reliability. Redundancy, for example, is a basic principle of design for building reliable systems, and it is hard to imagine

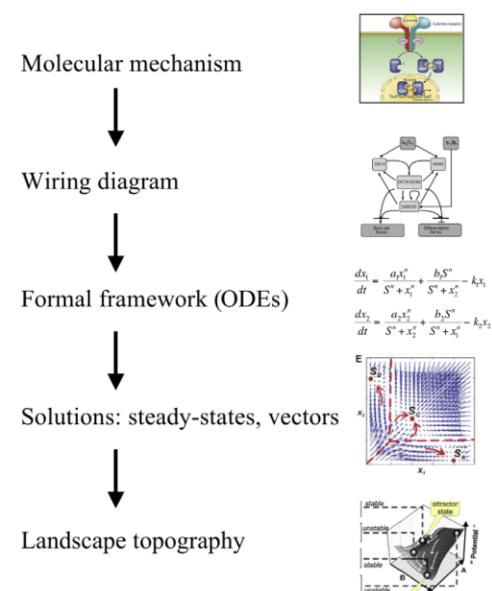
how, were it not for this amnesia, recent findings of extensive redundancy in developmental pathways could have been quite as startling as they have been. [Kel00, p.109-111]

That is, on Keller's view, the "molecular reductionist" trend in the MBT, myopically fixated on Crickian information flow outward from genomes, lost sight of the need to deploy the fruits of the IT revolution and of cybernetic theory to explain "goal-seeking" and "self-organisation", which is to say, agency. The logical solution to this quandary is to do precisely what Fagan suggests- reintroduce the *cybernetic formula*, levering the computational power of modern semiconductor devices to solve systems of ordinary differential equations (SODEs), treating cells as a type of complex chemical system with Dynamical Systems Theory (DST). Both Keller and Fagan believe this is a reasonable way to find the macromolecular basis for Waddington's classic theory of cellular specification. Fagan makes this view explicit in her definition of mathematical "Systems" approaches to JMEx:

A cellular systems model consists of a finite set of molecular elements  $\{X_1, X_2 \dots X_n\}$ , representing DNA, RNA, proteins, and small molecules. Complexes of multiple components and functionally-distinct forms of one molecule are represented as distinct, so the set may be larger than a simple parts list. Each molecular element in the set is characterized by a value of a state variable  $\{x_1, x_2 \dots x_n\}$  at time  $t$ . In these models, the cell is defined as a complex system which at any time  $t$  is in a state  $S(t)$  that is fully determined by the values of a set of variables  $\{x_1, x_2 \dots x_n\}$  representing the state of each molecular component. The values of these variables exhibit numerous and diverse dependency relationships. Cell behavior, including development, is conceived as the result of changes in the values of state variables. A set of molecular elements, each described by a state variable, and dependency relations between the values of those variables, comprises a cell network model.

In this way, systems biologists aim to derive predictions about cell behavior, including developmental phenomena, from mathematical descriptions of interacting molecules. Insofar as these predictions are confirmed by experiments, the mathematical models that entail them can be said to explain the phenomena in question. The process of model construction begins with detailed description of a molecular mechanism. This mechanistic description is then simplified into a wiring diagram, which is next translated into a formal framework. Solutions within a formal framework correspond to vectors and attractors, and these vectors and attractors in turn define a 'landscape.' [Fag13, p.204]

Fagan has advanced this prescriptive view of what Systems biological approaches to JMEx consist in on several occasions, so that she has produced a useful summary diagram, reproduced in Figure 14.3. The most obvious feature of the diagram for the student familiar with MBT models is its linearity: there is *only one* Systems approach represented here.



While most MBT practitioners and students are familiar with this approach, what is most notable about it is that it is, in practice, rarely applied in this form. Part of the reason for this is technical problems in relating complex macromolecular systems to formal SODEs, which I have summarised in the section relating to that method and its theoretical underpinnings. In general, however, I have gone against this "Systems formalisation scheme" for the following reasons:

- 1. Actually Existing MEx Pluralism.** Harris' theory, like most others, does not use DST to produce topographies that represent some kind of "Waddington space".

While much effort has gone into DST-type formalisations of mechanisms, there are other types of “Systems explanations” that must be accounted for.

2. **Unclear General Utility.** While SODEs offers a plausible formalisation of Waddington’s view, it is not at all clear that all or most cellular processes involve transitions between attractors in some state space whose dimensions express protein expression levels or estimates of activity. Waddington’s view of the cell is of a probability landscape of *chemical tendencies* extended in time, underpinned by genes and the interactions of their products (see Fig. 14.5). This is the conceptual conceit that allows macromolecular relations to be treated as a system of ODEs; their interactions are understood ordered chemical reactions in cellular “bioreactors”. I see no reason to presume that this is an adequate modelling frame to explain phenomena where, for instance, the specifics of local inter- and intracellular spatial organisation is as or more salient than chemical kinetics, as in many emerging MEx related to nuclear behaviours, transcription factories, tissue formation, etc.
3. **Commitment to Open Exchange.** As noted in Section ??, I am taking some care not to dictate what post-Systems biology ought to look like. I do not know what “systems” disciplines will ultimately prove to be successfully integrable into the MBT. Given the potential of, for instance, physical explanations to contribute as components of MEx [Mor11] or of limits on them, it seems unwise to exclude these by adopting a MEx concept that does not allow for their inclusion, or of other methods that may prove useful. I am in particular concerned with the formation of subtraditions committed to particular mathematical approaches and not able to meaningfully compare their model formalisations.

I have, instead, treated “Systems” approaches as methods of generating extra-traditional generalisations about the components of mechanisms, so that a “Systems” JMEx does not necessarily summarize the activity of the entire mechanism ( $S \Psi\text{-ing}$ ) using one particular mathematical method, but may draw from any number of traditions in generalising about  $x$ ’s  $\phi\text{-ing}$ .

There are at least two concerns with this sort of ad hoc modification: the possible loss of biological macromolecules as the lowest level of organisation represented in MEx, and the possible loss of the multi-level logical relations constituting the “Joint action” of macromolecular  $x$ ’s  $\phi\text{-ing}$  together to constitute an  $S \Psi\text{-ing}$ . I address these in turn:

### HJMEx may specify their “bottom” macromolecular levels polysemously

Firstly, Fagan’s JMEx concept usefully explains the fact that MBT explanations tend to “bottom out” at the macromolecular level. This is generally recognised and is often offered as a reason for considering the MBT to be meaningfully independent from physical and chemical traditions. We rarely make reference to physico-chemical properties of biological molecules, except in defining their activity ( $\phi$ -ing) or their relations (how  $x_1$  jointly  $\phi$ s with  $x_2$ ’s  $\phi$ -ing). I do not want to lose this feature of Fagan’s JMEx.

At the same time, we often offer generalisations about molecular behaviours that we intend *polysemously*. Let us take the example Fagan offers of the mechanism for *Drosophila* germ stem cell (GSC) specification- as Fagan notes, most (80%) but not all mitoses proceed by employing this mechanism. However, if we want to model the cellular population dynamics of a whole *Drosophila* embryo, we probably will offer a generalisation about a mitosis as a *process*- that is, we are interested in the *total number of divisions*, irrespective of their molecular mechanisms. If we were interested in (a)symmetric divisions, we would actually want to specify something like Kitcher’s PS-process concept, which could conceivably be accomplished by any number of mechanisms. Therefore, a JMEx that made reference to “mitoses” in general, or “symmetric” divisions in particular, could well have a description of some  $x$  that  $\phi$ s on a higher-than-molecular level of organisation, that does not reduce directly to any particular set of molecular  $z$ s  $\mu$ -ing, to use the terminology of Fig. 14.1. Instead, it may refer to any number of MEx, and is thus “polysemous” in the sense that it is constituted by more than one entity at the “macromolecular basement” level<sup>17</sup> One could conceivably add more detail “below” this level if necessary, but in general, there is every reason to expect MEx to have components that are not usefully broken down beyond a certain level, and which, when applied in an abstract or theoretical manner, may not even refer to stable molecular entities at all<sup>18</sup>.

However, as I hope is clear, I do not mean to imply that a CGT traditional concept like a PS-process has some reality *independent* of the specific macromolecular arrangements that actually constitute cells dividing and partitioning their contents between two descendants. The fact that the CGT concept does not reduce cleanly to a set of one or more MGT descriptions of macromolecular processes in nuclei may or may not be relevant to the explanatory quality of the JMEx which uses the CGT concept as a stand-in for many different MGT-specifiable cellular dynamics. To a large extent, this depends on what we suppose the intermeshing properties of the mitotic process and the other components of the mechanism to be. That is, if we suppose mitosis (call it  $x_1 \phi_1$ -ing) to be mainly independent from other relevant cellular behaviours we are modelling ( $x_{2\dots n} \phi_{2\dots n}$ -ing), it may not matter whether we have any account of the specific macromolecular arrangements ( $z_{1\dots n} \mu_{1\dots n}$ -ing together), which we take to constitute  $x_1$ ’s activity, to expand our HJMEx “downward”.

Moreover, we understand that any time we propose an experiment which intervenes onto some  $x \phi$ -ing, where the HJMEx uses a polysemous generalisation including many macromolecular arrangements of  $z$ s  $\mu$ -ing jointly, that intervention is onto some subset of actually-occurring, specifiable-in-their-particulars

---

<sup>17</sup>This polysemy essentially specifies all actually-present mechanisms  $S$  that  $\Psi$  such that some  $P$  occurs. It is very likely that eg. mitosis, in most contexts, is driven by a smaller number of mechanisms that account for a majority of divisions, but a very large number of mechanisms that account for rare ones (these could even be completely unique). If we are accounting for cell number in a tissue, we do not care if a mitosis was produced by a miraculous, unique mechanistic causal locus or not. A division is a division.

<sup>18</sup>Arguably, a PS-process is a basic requirement for a living thing, defined as a replicating semiotic agent. Semiosis implies an interior and an exterior, so a membrane or wall of some kind, forming a cell. “Replication” in the strict sense must then proceed by duplication of the agent’s contents and segregation by dividing its contents into two membrane compartments. To the extent that the selection of macromolecular materials used in living systems is historically contingent, there will be a correspondingly large number of possible material substrates and intermeshing properties that produce such a process.

macromolecular  $z_{1\dots n}$  and  $\mu_{1\dots n}$ -ing together. That is, the use of polysemous generalisations in no way implies that an intervention is not “directly” onto the macromolecular materials constituting an organism. For example, if there is more than one macromolecular process which produces mitosis in some cultured cells, and we apply a chemical inhibitor of mitosis, we understand that we are going to intervene into that polysemous distribution of possible macromolecular developments that result in mitosis. Which mechanisms are actually affected depends on how common those macromolecular trajectories<sup>19</sup> are and how the chemical inhibitor affects the components and intermeshing properties of each one. Chemical inhibitors of mitosis that broadly affect commonly used mechanisms would be promiscuous inhibitors of a macromolecularly polysemous behaviour, validly described by polysemous CGT generalisations about mitosis. We may be interested in specific inhibitors if they are available, but inhibitors of rarely employed mechanisms will probably never be discovered in the first place, due to the population-based assays typically used in drug discovery.

Conceivably, this could lead to a class of MBT HJMEx that are composed entirely of generalisations which cannot be reduced to particular biological macromolecular relations at all. This seems like a plausible outcome of the SBE- a class of models that deal exclusively with phenomena at higher levels of biological organisation, polysemously referring to huge arrays of unspecified molecular arrangements (perhaps consisting of statistical generalisations about assemblages of possible mechanisms  $S$  that  $\Psi$  to produce  $P$ ). It seems very unlikely to me that these will be anything more than toy models and curiosities without strong grounding in the MBT’s foundational explanatory level, however, so this does not seem like a likely risk to MBT traditional stability.

### HJMEx retain constitutional directionality and allow causal nuance

The second potential problem involves the potential confusion involved in incorporating extra-MBT causal generalisations at higher levels of biological organisation. For instance, there is much interest in tension and shear-induced signal transduction involved in morphogenesis [ML15, p.5] as contributors to self-organisation and biological complexity. Therefore, HJMEx might include physical explanations at very high levels of organisation. It is unclear that all such explanations can be understood to proceed from the “bottom up”.

Let us take one such HJMEx that already circulates “in the wild”: behavioural impingement on osteoblast development and bone dynamics. Bone, being a highly dynamic tissue, actively remodelled in response to stress, has structural and functional aspects that are determined by organismal behaviour, psychological states, etc. This is obvious when we consider what we have to do to “model” the effects of microgravity conditions on bone dynamics in rodents: the hindlimb elevation assay (HEA?), ie. restraining the animal such that its organised behaviour cannot create normal stresses on the bone, which in turn results in bone mineral density loss<sup>20</sup>. This occurs via the sensing of tissue-level strain by osteocytes, which in turn direct the action of osteoblasts and osteoclasts<sup>21</sup> in remodelling the overall structure of the bone to better suit the actually-experienced load.

---

<sup>19</sup>I mean this in the loosest possible, non-Newtonian, non-quantitative, non-state space sense.

<sup>20</sup>I have placed “model” in scarequotes because it is not at all clear to me that you can study the effects of microgravity except in a 1g laboratory. Microgravity-related BMD loss being a behavioural disease of a tiny number of humans (astronauts), taking on this risk with full informed consent, it is hard to see how it can be ethical to torture thousands of rats with these assays in order to “study” this “medical issue”.

<sup>21</sup>Osteoblasts lay down bone, osteoclasts destroy it. The local balance of osteoblast/clast activity determines whether bone is plated out or broken down at a given site.

A component of a HJMEx describing bone remodelling might therefore include a statistical generalisation of behaviourally-conditioned bone stress dynamics (say, a finite element analysis of a rat's bone under normal behavioural strain over some period of time). It is not immediately clear how such a HJMEx should be hierarchically arranged. Conceivably, a description of the rat behaving could be a "mechanistic" system S  $\Psi$ -ing, which includes the details of the molecular mechanisms underlying osteocyte strain sensing as low-level  $z$ s  $\mu$ -ing in mainly-macromolecular MEx. This ambiguity leaves space for those with directional theories of causation to claim that one may intervene onto the rat behaving (S  $\Psi$ -ing) and so cause a downward cascade through mechanistic levels and ultimately impinge onto the basement level in this manner, implying a similar problem to the one we identified with the mutual manipulability account of MEx.

I respond to this in two ways:

1. Firstly, HJMEx retain "constitutional" directionality in two senses:

(a) there is a "basement" level but no "ceiling" ("sky's the limit", organisationally). That is, HJMEx assume that the biological scale hierarchy *necessarily starts* at the level of molecular biological material *simplicter* and proceeds upwards through macromolecular assemblages, organelles, organelle lineages, cells, cell lineages, tissues, organisms, organismal lineages, local ecosystems, planetary ecosystems, and then, hypothetically, interplanetary ecosystems, intersolar ecosystems, and so on<sup>22</sup>.

(b) HJMEx imply that interventions at higher levels *constitute* intervening at the lowest level. If humanity intentionally destroys the planetary ecosystem by raising the average temperature of the planet to the point where plants no longer reliably germinate, this *is* an intervention at the lowest possible biological level of organisation of all organisms. Unloading the rat's limbs *is* intervening onto macromolecular processes in osteocytes.

This is completely sufficient for understanding the structure of a causal locus as I have defined it, and it retains JMEx's accurate reflection of the "bottoming out" of biological MEx at the macromolecular level.

2. Secondly, HJMEx deemphasize simplistic "directional" accounts in favour of detailed exploration of the intermeshing properties of JMEx at different levels of organisation.

Let us again take up the example of osteoclast strain sensing. There is a significant problem in the potential representation of a Mechanical Engineering Tradition (MET) finite element analysis (FEA) as a generalisation of the effects of rat behaviour on bone loading. The FEA is a description of the physical dynamics of tissue, measured on a spatial scale of (perhaps) millimeters, and a temporal scale of seconds. Now, macromolecular dynamics occur on spatial scales of nanometers and temporal scales of microseconds! There is therefore a dynamical barrier in this scale hierarchy, discussed further in [the section on hierarchy theory](#).

What is interesting to us then is not a question of causal priority or "directionality" but a question of "how-does-it-work", a request to elucidate the intermeshing properties between the extra-traditional MET FEA and whatever model we have for macromolecular osteoclast strain sensors.

The interesting question is *how* information about strain is brought across the dynamical barrier

---

<sup>22</sup>Theories of panspermia typically postulate intersolar ecosystems with planetary ecosystem components connected by their "intermeshing property"- their propensity to fling debris into space under the impact of asteroids, etc.

by the strain sensor, and *how* the resultant activities of osteocytes, osteoblasts, and osteoclasts act to alter the bone's structure such that its stress/strain properties evolve under load.

Therefore, HJMEx allow us to accurately represent contemporary “Systems” theories, while retaining the MBT’s understanding of the explanatory “macromolecular basement”. By emphasizing the intermeshing properties of mechanisms, and their part-whole relations, we can move past problematic causal generalisations into specific descriptions of these properties.

**HJMEx defined** Fortunately, little needs to be done to Fagan’s JMEx (JM) definition to differentiate HJMEx. I have indicated these clarifications with HJM-n notations.

(HJM) Components  $x_1, \dots, x_n$  jointly  $\Psi$  as mechanism S if and only if

- (i) each  $x_i$  has properties that mesh with one or more other  $x_i$ ,
- (ii)  $x_1, \dots, x_n$  are spatially organized and their activities  $\phi_1$ -ing, ... $\phi_m$ -ing causally organized in virtue of their meshing properties,
- (iii)  $x_1, \dots, x_n$  and  $\phi_1, \dots, \phi_m$  so organized constitute mechanism S  $\Psi$ -ing, and
- (iv)  $x_1, \dots, x_n$  and  $\phi_1, \dots, \phi_m$  not so organized do not constitute S  $\Psi$ -ing

HJM-1: Each  $x_i$   $\phi_i$ -ing consists of a description of the behaviour of a biological system at the macromolecular level of organisation or higher.

HJM-2: At any level of organisation higher than the macromolecular “basement”, each  $x_i$   $\phi_i$ -ing may describe one or more lower level mechanisms S  $\Psi$ -ing polysemously, and need not specify any of these S in any particulars.

HJM-3: Polysemous xs  $\phi$ -ing describe general properties of a population of Ss  $\Psi$ -ing that intermesh with other xs  $\phi$ -ing, and these need not arise as a consequence of specifically defined macromolecular mechanisms S.

HJM-4: Polysemous xs  $\phi$ -ing may have general properties that intermesh with zs  $\mu$ -ing on lower level, such that there is an interlevel exchange of information (boundary conditions are placed on an assemblage of zs  $\mu$ -ing, for instance), and these need not arise as a consequence of specifically defined macromolecular mechanisms S.

HJM-5: All descriptions x are drawn from some scientific tradition, and so carry with their use an array of traditional natural assumptions, epistemological and metaphysical axioms, methodological considerations, and so on. No x is *a priori* excluded from inclusion in HJMEx on the basis of its traditional identification alone.

HJMEx are therefore somewhat more complex than JMEx, and require careful attention to the physical and biological plausibility of between-component and interlevel relations. That said, the “basic norm” that Fagan suggests for JMEx is upheld for HJMEx. Good HJMEx describe (i) the properties of components that allow them to interact in specific ways, (ii) the spatiotemporal-*cum*-causal organisation produced by these interactions, (iii) the sense in which the overall system’s behaviour is constituted by the components, (iv) in a manner that explains how the organisation of components makes a difference to the system’s behaviour.

Having so-defined HJMEx, let us proceed to consider how we might assess the development and deployment of these explanations in scientific research programs.

#### 14.2.2.3 Extended HJMEx (EHJMEx)

I have further modified the HJMEx concept by suggesting extending it in time over its career, iteratively deployed as an explanation in different studies. I have taken this idea from Kenneth Schaffner’s idea of the “Extended Theory”. As Schaffner says of his own “metascientific unit”:

I am using the notion of an *extended theory* to introduce a diachronic unit that (1) permits *temporal* theory change and (2) also allows for some *logical* changes - namely, some of the assumptions within the diachronic unit will change while the integrity of the whole is preserved. [Sch93, p,211]

Schaffner was working with a substantially different conception of a biological theory<sup>23</sup>, but my intent is similar. There are numerous reasons to describe a theory's development in time. I have done so because I was unable to assess, by inspection, the broader significance of changes to components of the theory as expressed in different primary reports over time.

By representing a changing HJMEx over time diagrammatically, I am able to better understand the relationship of changing theoretical structure to the model output and explanatory power, since I can compare the output of different HJMEx to empirical data in the [global model selection approach](#). Moreover, this diachronic approach provides insight into the dynamics of a particular theoretical construct over time, so that we can form an idea of whether a given EJMEx is becoming more or less useful, consistent, precise, etc., and possible reasons for this.

**EHJMEx defined** There is little additional nuance to EHJMEx, so let us simply define one as a series of HJMEx explanations, arranged in chronological order, such that the EHJMEx reveals the diachronic explanatory dynamics of the changing HJMEx being offered by a particular research program<sup>24</sup>. I suggest that presentations of EHJMEx would normally be marked up in such a way that the changing components of the explanation and the relevant traditional implications are highlighted.

### 14.3 The Feyerabendian Modeller

If we understand the value of the interaction of scientific traditions in Feyerabendian terms, we become sensitive to the metaphysical contrasts induced by when we cognitively compare explanations produced by two or more traditions. A valid contrast between traditional explanatory frames occurs when we come to correctly understand *how* the two traditions' metaphysical implications differ<sup>25</sup>. If we are pursue this as a deliberate practice, particularly with the regard to the details of (E)HJMEx, it would be helpful to have a systematic way of making these contrasts. It is difficult to cognitively explore the full implications of some outside-MBT "Systems" artefact, like a solution to a system of ODEs, without an equally systematic exploration of the implication of that explanation *and its alternatives* for the phenomenon at hand.

Therefore, we are interested in an approach that can reveal the *systematic implications* of arbitrary HJMEx for the dynamics of our system. By this, I mean we want to make formal comparisons between model output for some group of proposed HJMEx, with to their explanatory power for some set of real data. We would also like to ensure that the approach is sufficiently general that it can be used to generate comparable output for HJMEx irrespective of the traditional or methodological makeup of their "explanatory components". In other words, if someone comes to us with a verbal description of

---

<sup>23</sup>In effect, Schaffner wanted to be able to compare Lakatosan "research programmes" consisting of lawlike generalisations arranged in hierarchies of centrality to the overall programme.

<sup>24</sup>One might also conceive of a complex HJMEx phylogeny, branching as different researchers adapt the HJMEx for their own local purposes, rather than a linear sequence.

<sup>25</sup>Invalid metaphysical contrasts can be drawn, as from misunderstanding or malice, and there may even be rhetorical sleights of hand that make these errors seem consistent, so one must be alert during this process. Biologists are better aware than anyone of how easy it is to "fool oneself", hence our attachment to blinding procedures.

rules that a retinal progenitor cell might follow on its morphogenetic trajectory towards integrating into the neural retina, this should be comparable to any complicated set of statistical generalisations about the same kinds of cellular behaviours.

### 14.3.1 Laws, Rules, Models, and Games

As should be clear from our foregoing discussion, we are not expecting to encounter enough “laws” in HJMEx to describe the overall explanation as “Law-based”. Nonetheless, one obvious consequence of admitting physical explanations into JMEx is that we may encounter traditional claims which include ostensible *bona fide* laws<sup>26</sup>, alongside our typical diagrammatic macromolecular JMEx. These may be integrated into explanations that involve something like verbally conveyed heuristic rules (“no differentiation for the first 2 rounds of mitosis”), as well as sophisticated mathematical simulations of phenomena, or statistical generalisations about higher-level phenomena (ie. “mathematical models” in general) whose molecular underpinnings are unknown or not relevant.

There tends to be some ambiguity in the way that we use these terms, and a necessary anthropomorphic projection onto a phenomenal reality of transcendent complexity and informational density. If by “law” we mean that the mathematical description of some natural regularity offered by one of the physical subtraditions is *actually existentiated* as some kind of iron fisted algorithmic tyrant, we have a very strange sort of ontology<sup>27</sup>. We know, of course, that our physical descriptions of the world are also “models”, that Newton’s Laws are not God’s Laws, nor Nature’s Laws, but a set of finely honed mathematical approximations of macroscopic mechanical phenomena local to a particular gravity well. We do not speak of Quantum Laws but of Quantum Theories, of the Standard Model of physics, and so on. We should nevertheless be clear that no reasonable person any longer believes that there is a “true model”, a generative algorithm “actually out there”, that can be discovered by comparing different candidate formulations until the “correct” one is found, thus settling the issue of what the “true” set of laws, set of rules, model, or whatever “actually” is. Systematised knowledge about phenomenal reality is *made possible* by abstracting away from the unique momentary particulars of each example of some general class, and any type of theory or model involves abstractions. That is to say, any biological explanation is by necessity *not* a complete recapitulation of the system that generated the original phenomenon.

In this thesis I have, for the most part, described EHJMEx as “explanatory programs” or “bodies of theory”, HJMEx as “explanations” or “theories”, and reserved the term “model” for particular mathematical formalisations used either (a) as components of HJMEx (“component model”) or (b) as summaries of the overall output of HJMEx (“summary model”) for some parameters of interest. I am not very interested in enshrining any of the diverse components of HJMEx or the explanations themselves as

---

<sup>26</sup>Various traditions may also posit a priori logical or reasoned truths. We tend to agree that typical logical syllogisms do not need justification, but we rarely examine the logical background of our claims (why are they justified? what is their structure?). The science of logic seems underutilised in the biological sciences.

<sup>27</sup>Richard Swenson has memorably described the definition of genes and memes as “selfish”, Ideal algorithmic replicators (which cannot be the case, as neither is a valid unit of selection) as a form of “Neopythagorean Reductionism”. [Swe97]. This sort of covert Idealism is a byproduct of the unresolved Cartesian fissure mentioned in the discussion of the SBE. It is, ironically (since advanced by militant atheists), the product of Descartes’ bifurcated *Christian* ontology. Unable to fully digest Greek philosophy properly rooted within its Abrahamic heritage, the Greek Pauline church was unable to confer upon its Catholic and Protestant progeny a viable monist ontology capable of providing a reasoned approach to study of causal regularity in phenomenal reality, and was forced to divide existence between the fallen “natural” world and an infinitely transcendent “supernatural” plane, interacting indirectly through providence (the workings of which were irretrievably compromised by the identification of First Cause with a particular human being). It was inevitable that this divide would be reproduced in Rene Descartes’ ontology as “matter” and “mind”, and thenceforth to the absurd spectacle of “scientists” denying the reality of their own minds while covertly presupposing an invisible mathematical realm where algorithms vie for dominance.

“laws”, but we can usefully speak of “rules” that connect them- the intermeshing properties of various HJMEx components will often be simple prose descriptions of how one depends on the other. These “rules” are more like human-generated legislation than Divinely-*cum*-Naturally ordained Law- we do not expect cells to always and everywhere “follow the law”, subject to a tyrannical deity or algorithm. They will, of course, break some of any conceivable set of rules if one observes enough cells long enough.

I therefore suggest that we think of HJMEx in this context as a the “rules” to a type of simulated game. This is appropriate, given that systematic Western modelling of complex biological phenomena has deep roots in the evolution of games like Kriegspiel. Indeed, the widely-known toy Conway’s “Game of Life” is a type of cellular automaton closely related to the spatially-situated NetLogo models employed in this thesis. Biologists are thus generally familiar with this type of simulative “game”. In comparing summary models of our explanations, we are not asking “which of these models is actually true”, but “which of these games is most ‘realistic’”. We mean something more quantitative than this, of course, which is related to information theoretical measures of information loss between empirical reality and model, as discussed below. It is, nevertheless, necessary to differentiate between the *rules* of the “game” (the overall HJMEx offered in some publication), the *specific interpretation* of the rules that gives rise to the actual implementation of the game (eg. the procedural logic of some summary model that we use to elucidate what the HJMEx “game” implies about the system it represents)<sup>28</sup>, and actual living phenomena (cells are not referring to a rulebook in their “play”). This is an appropriate perspective to keep, as we take our models seriously as “realistic” simulations of phenomena while not ascribing reality to any of them, since they are, after all, games and not Law. We may consider analogically the type of computer simulation used by modern militaries- while everyone is entirely conscious of the essential unreality of these virtual representations of combat, their “explanatory power” with regard to the technical particulars of warfare is taken extremely seriously.

### 14.3.2 NetLogo Agent Model Simulation

For the seasoned modeller, this suggests some general procedural or rule-based simulation space. We can simply assemble the rules and statistical generalisations that we find in HJMEx and simulate the behaviour of a cell or population of cells that satisfies this overall systemic description. This need not be particularly sophisticated to begin with, since none of the current explanations in the field involve detailed explanations of the production of eg. cellular morphology, or anything that would require much spatial detail. That said, there is no reason why a simulation framework should not be “portable” to whatever simulation contexts are required, including those that may eventually include fine-grain three dimensional spatial detail (eg. synapse formation, nuclear dynamics, etc.).

There is, fortunately, an accessible, appropriately-scaled, open source, academic modelling software package that is ideal for expressing the kinds of “cellular logics” that we find in Harris’ HJMEx- NetLogo. Developed by Uri Wilensky as a revamp of the Logo programming language, it is specifically intended to handle simulations of emergent phenomena which arise from the interaction of many agents with interdependent internal dynamics. It is relatively simple to code and to understand the code of others, and is designed to be “low threshold and no ceiling”, meaning we can start with very simple models

---

<sup>28</sup>Any set of rules requires an interpreter who implements them. Particularly annoying and tendentious interpreters of game “rules” are often called “rules lawyers”, and the same sort of hermeneutic or interpretative activity is always present when confronting the necessary ambiguities in a ruleset. Generally speaking, one tries to keep to the intent of the person offering the ruleset, in accord with the Principle of Charity, and avoid “rules lawyering” that tends to distort this.

without fear that we will be unduly constrained in implementing more complex ones later<sup>29</sup>. Furthermore, NetLogo is widely used in theoretical modelling work, and so is a broadly acceptable approach in the spirit of open exchange.

Because we can simulate arbitrary cellular logics in 2d or 3d with NetLogo, there are few restrictions on the HJMEx which can be used to explain the behaviour of retinal progenitor cells. Both traditional SCBT modelling approaches (clonal model, SSM) are fully expressible in NetLogo, including in their “unrealistic” features (eg. we could decide to simulate “immortal” stem cells that “self renew” without progeny). We can simulate comparable population-level outcomes for HJMEx which draw on different traditions, because the logic of the different explanations is “translated” into the code which drives the model’s agents. This feature of simulating agent models means that dissimilar models may be brought into a comparable context, where their relative explanatory value can be assessed given some empirical dataset.

### 14.3.3 A Global Approach to Model Selection

If we take NetLogo as our modelling frame in which we will express all of our diverse HJMEx, we need a more concrete idea of how to approach this task. I proceed here with a broad suggestion taken from the statistical theory of model selection. As Burnham and Anderson state in the introduction to their helpful text “Model Selection and Multimodel Inference”:

Often, one first develops a global model (or set of models) and then derives several other plausible candidate (sub)models postulated to represent good approximations to information in the data at hand. This forms the *set of candidate models*. Science and biology play a lead role in this a priori model building and careful consideration of the problem. [BA02, p.2]

The development of global models is generally prior to the consideration of any particular dataset or to the use of statistical methods. The selection of a global model is something which is guided primarily by the modeller’s objectives and by their detailed area knowledge:

Building the set of candidate models is partially a subjective art; that is why scientists must be trained, educated, and experienced in their discipline. The published literature and experience in the biological sciences can be used to help formulate a set of a priori candidate models. The most original, innovative part of scientific work is the phase leading to the proper question.

...

Development of the a priori set of candidate models often should include a global model: a model that has many parameters, includes all potentially relevant effects, and reflects causal mechanisms thought likely, based on *the science of the situation*. The global model should also reflect the study design and attributes of the system studied. Specification of the global model should not be based on a probing examination of the data to be analyzed. At some early point, one should investigate the fit of the global model to the data (e.g., examine residuals and measures of fit such as  $R^2$ , deviance, or formal  $\chi^2$  goodness-of-fit tests) and proceed with analysis only if it is judged that the global model provides an acceptable fit to the data. Models with fewer parameters can then be derived as special cases of the global model. This set of reduced models represents plausible alternatives based on what is known or hypothesized about the process under study. Generally, alternative models will involve differing numbers of parameters; the number of parameters will often differ by at

---

<sup>29</sup>I am indebted to the interesting paper of Rafael Bravo and David E. Axelrod, [BA13], for demonstrating the practical utility of this approach in simulating colon crypts *in silico*.

least an order of magnitude across the set of candidate models. Chatfield (1995b) writes concerning the importance of subject-matter considerations such as accepted theory, expert background knowledge, and prior information in addition to known constraints on both the model parameters and the variables in the models. All these factors should be brought to bear on the makeup of the set of candidate models, prior to actual data analysis.

[BA02, p.16-17]

This approach is itself “counterinductive” to usual field practice, which generally emphasizes finding the simplest model that accords with a selected set of data. As Feyerabendian modellers, we should immediately be aware that this stage of consideration necessarily implicates rules, preferences, and natural assumptions that are specific to traditions, to the intellectual and social background of the modeller, and so on. Models are fundamentally tradition-specific artefacts that arise from particular scientific and social practices:

Models arise from questions about biology and the manner in which biological systems function. Relevant theoretical and practical questions arise from a wide variety of sources (see Box et al. 1978, O’Connor and Spotila 1992). Traditionally, these questions come from the scientific literature, results of manipulative experiments, personal experience, or contemporary debate within the scientific community. More practical questions stem from resource management controversies, biomonitoring programs, quasi-experiments, and even judicial hearings. [BA02, p.16]

One continuously cited traditional consideration that we all recognise, but for which no specific rule can ever be articulated, is “biological plausibility”, which must obviously enter into biological modelling practices:

The more parameters used, the better the fit of the model to the data that is achieved. Large and extensive data sets are likely to support more complexity, and this should be considered in the development of the set of candidate models. **If a particular model (parametrization) does not make biological sense, this is reason to exclude it from the set of candidate models, particularly in the case where causation is of interest.** In developing the set of candidate models, one must recognize a certain balance between keeping the set small and focused on plausible hypotheses, while making it big enough to guard against omitting a very good a priori model. While this balance should be considered, we advise the inclusion of all models that seem to have a reasonable justification, prior to data analysis.[BA02, p.17]

[boldface original]

There is no rule or set of rules by which “biological plausibility” might be established. Stem cells are themselves a clear example of this: the modern history of stem cell research describes the successive expansion of “possible” cell behaviours in an unpredictable and enduringly surprising fashion. Moreover, plausibility judgements are heavily influenced by the sort of thing that one thinks the system to be, by the attributes we take it to have and the behaviours we already know it undergoes.

For instance, while the phenomenon of cells undergoing “fusion”, resulting in durable genetic reprogramming of the resultant “offspring” had been described in culture, there was no particular reason to believe that this could, or did, occur in any systematic way *in vivo*. However, there has recently been description of transgenic hematopoietic stem cells (HSCs) migrating to the eyes of mammals in response to retinal damage, where they enter this immunologically privileged site and fuse with Müller glia, resulting in reprogramming and functional repair of the lesioned retina.[PBS<sup>+</sup>18, SSV<sup>+</sup>16] There are any

number of reasons to believe that the underlying phenomenon (HSCs infiltrating the retina and fusing with retinal glia) is “biologically implausible”. Someone with a view of the mammalian retina as a highly static tissue might not have ever considered the possibility. Whether one accepts the authenticity of the report and its interpretation or not, prior to its publication, excluding this type of “extra-organic” invasion by distant stem cells in modelling approaches to the induction of Müller glial-mediated retinal repair, on the grounds that this behaviour was implausible, would not have raised much objection.

We should therefore seek to be aware of what is implied by our traditional models, and what may be gained by the use of a general model framework in which they can be expressed as submodels. One advantage of proceeding deliberately in this way is that we avoid the common pitfalls of “shotgun” approaches and “data dredging”, which tend to produce descriptions of spurious regularities:

It is not uncommon to see biologists collect data on 50-130 “ecological” variables in the blind hope that some analysis method and computer system will “find the variables that are significant” and sort out the “interesting” results (Olden and Jackson 2000). This shotgun strategy will likely uncover mainly spurious correlations (Anderson et al. 2001b), and it is prevalent in the naive use of many of the traditional multivariate analysis methods (e.g., principal components, stepwise discriminant function analysis, canonical correlation methods, and factor analysis) found in the biological literature.

...

A model is fit, and variables not in that model are added to create a new model, letting the data and intermediate results suggest still further models and variables to be investigated. Patterns seen in the early part of the analysis are “chased” as new variables, cross products, or powers of variables are added to the model and alternative transformations tried. These new models are clearly based on the intermediate results from earlier waves of analyses. The final model is the result of effective dredging, and often nearly everything remaining is “significant.” Under this view, Hosmer and Lemeshow (1989:169) comment that “Model fitting is an iterative procedure. We rarely obtain the final model on the first pass through the data.” However, we believe that such a final model is probably overfitted and unstable (i.e., likely to vary considerably if other sample data were available on the same process) with actual predictive performance (i.e., on new data) often well below what might be expected from the statistics provided by the terminal analysis (e.g., Chatfield 1996, Wang 1993). The inferential properties of a priori versus post hoc data analysis are very different. [BA02, p.18-38]

Therefore, we seek to build an a priori, general, global model, inside of which we may specify “sub-models”, including Harris’, and including those that may include quite foreign explanatory components. The “design principle” of a Feyerabendian global model is about specifying limits within which many traditional explanations can “operate”, so that we do not restrict what kind of sub-models we can compare simply because they do not make sense within a *particular* conception of how the behaviour of stem cells ought to be described.

#### 14.3.4 The Global Semiotic Agent Model

While the details of the global model employed in this study are advanced in Chapter 2, we should briefly consider the broad strokes of the picture so that we have an idea of what types of cellular activities we are interested in. Ultimately, we are looking for explanations of a phenomenon P (zebrafish retinal formation and development), produced by a mechanistic system S, which we take to be constituted (primarily) by cellular xs which  $\phi$  together to produce the overall phenomenon P. We want a tissue-level explanation that drills down to the relevant macromolecular underpinnings of our cellular behaviours  $\phi$ .

At the tissue level, we already have a good, general, a priori description of the sorts of cellular  $\phi$ -ing that give rise to tissue morphogenesis offered by our own evolutionary developmental biologist Elly Larsen:

All morphology results from only six cell behaviors and two tissue strategies [Fig. 1]. Morphological evolution occurs as a result of heritable modification of where and when the six cell behaviors are expressed. The six cell behaviors contribute to many different tissue structures; for example, cell death is important in embryonic processes, such as the removal of cells between human fingers to allow their separation. Matrix secretion is important in forming structures like bone, but is also important in more subtle aspects of development such as providing chemical signs which guide migrating cells. Migrating cells themselves can help form an astonishing array of structures including skeletal elements and muscles. However, the cell behaviors perform their roles according to one of two tissue strategies.

In the self-governing (autonomous) strategy, sheets of cells, known as epithelia, produce a variety of morphological structures, including the four general types: tubes, rods, spheres and sheets. The power of self-governing epithelial sheets to generate structures is well illustrated by imaginal discs in *D. melanogaster*. These are sacs of cells, which originate as out-pockets from embryonic and early larval epithelial sheets. During larval life, localized cell divisions produce disc-specific folds. At metamorphosis, each disc turns into a specific appendage, such as a leg or antenna, chiefly through changes in cell shape.

Unlike the self-governing strategy, a common tissue interaction or conditional strategy employs mesenchyme cell migration and utilizes interactions between mesenchyme and epithelia. Mesenchyme cells originate from epithelia through cell detachment and migrate through the epithelial basement membrane. Their migratory path and cell differentiation is directed by interaction with overlying epithelial cells. In vertebrate limb development, signals from the epithelium influence cell division and differentiation in mesenchyme, while the mesenchyme modulates cell division in the epithelium. Thus, the tissue interaction strategy adds tissue level controls to the development of the four morphological structures from which organisms are created. It is important to realize that the tissue- level controls produced by the tissue interaction strategy may act as developmental constraints to evolution. [Lar92]

Dr. Larsen's view is summarised in her useful figure, adapted here as Fig. 14.4. Larsen is here concerned mainly with explaining the evolution of *morphogenesis*, which is to say, the specialised functional roles that appear in the generated morphology are of less concern than the form-producing, spatial behaviours of the cells that produce them. While the SCBT is, of course, in conversation with the younger evolutionary developmental tradition (EDT), Larsen's perspective on tissue-level interactions is salutary. Harris' HJMEx are mainly about the dynamics of conceptually isolated cells, but what if he is wrong? In order to establish that, say, the inclusion of tissue-level effects improves the explanatory power of one of Harris' models, we have to be able to account for these effects in the first place. Moreover, Larsen has a clear perspective on tissue-level "intermeshing properties": sometimes these are involved in tissue formation (autonomous), sometimes they aren't (interacting). At this point we should scrutinise these "intermeshing properties" more carefully.

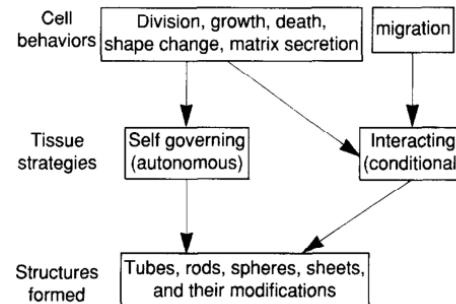


Figure 14.4: Relationship between processes and biological structures, excerpted from [Lar92]. The six cell behaviors and two tissue strategies from which all morphology results.

Since we know already that a tissue-level interactions are constituted by cellular interactions (which are in turn constituted by macromolecular zs  $\mu$ ing together in various ways). Let us first deal with the issue of the appropriate level of organisation for intermeshing properties specified by tissue-level HJMEx. We can imagine specifying all of these properties at the “basement” level of macromolecular organisation in physico-chemical terms, and in so doing, specifying the tissue-level system that this web of macromolecular zs  $\mu$ ing would constitute. This would be a sort of macromolecular reduction, where we would try to explain Larsen’s “morphogenetic alphabet” [LM87] from the ground up. From the discussion above, and the polysemy we expect to encounter in our HJMEx, we can infer that in general, this would be futile. Moreover, “Systems” approaches since Waddington have emphasized the organisation of the cell as a type of cybernetic bioreactor or factory. In Monod’s memorable phrase, “You have an exact logical equivalence between these two—the factory and the cell.” The various feedback loops at this level are taken to constitute the general entity which is said to “regulate” some cytoplasmic process. Although I think the cybernetic view is mistaken in significant ways, Waddington and Monod obviously understood perfectly well that molecular MEx would not make much sense as descriptions of isolated enzymes or chunks of signalling pathways-cum-transcription networks. It may be possible to conceptually, if not experimentally<sup>30</sup>, isolate these components and describe them in terms of their feedbacks on one another (as in the *lac* operon), but these components only *do* anything when installed in an appropriately regulated cybernetic cellular factory. Therefore, I take it that there is general agreement that a tissue-level HJMEx will focus on macromolecular phenomena at the level of the cell—where we cannot specify the macromolecular composition of some relevant system, we expect to encounter a statistical generalisation about what we take to be the activity of that macromolecular system operating *at the level of whole cells*.

Let us briefly consider an example: we might postulate that in differentiating, a cell may change its transmembrane protein distribution over time, such that nearby cells are given a differentiation-promoting signal. If we want to represent this in our agent model, will we be looking for a description of the conformational states of transmembrane proteins, and the various models describing their interactions with neighbouring extracellular transmembrane domains? Only to the extent that such a description can inform us of what is happening at the *overall* level of the cell, since we need to minimally hypothesize that such a protein is present in sufficient median concentration across the interacting membrane surface to have the pro-differentiation effect. As we may want to refer to a potentially polysemous macromolecular system (say, several separate pro-differentiation signalling pathways that separately contribute to this effect), the cell forms the “ground floor” above the “macromolecular basement”. The fact that SCBT traditional explanations (and Harris’ EHJMEx) are organised around the cellular “ground floor” is thus unsurprising—being HJMEx, parts of the “macromolecular basement” are not excavated, and other parts contain potentially chaotic polysemic macromolecular MEx.

We should further consider what this overall picture implies about higher levels of organisation before we move on. There are a few possible conceptual levels that are of less immediate interest: the lineage clearly matters, but in a retrospective, analytical sense, for instance. Tissue-level phenomena may meaningfully “feed back” on the behaving cells which constitute those phenomena. Tissue-level phenomena will generally arise from composite action of cells, such that we expect to see population-

---

<sup>30</sup>Obviously, enzyme biochemistry permits partial isolation of some kinds of macromolecular components *in vitro*. The basic idea of synthetic chemistry was that biochemistry would allow the assembly of “biological lego” in this biochemically reductionist sense. The fact that industrial bioprocess research has largely moved on from the idea of synthetic cells and is now focusing on “*in vitro* synthetic biology” is telling, given that this is nothing other than a description of the biochemical roots of the “synthetic biology tradition”.

level characteristics established at the level of the tissue, constituted by and feeding back upon the behaviour of the cells constituting those populations. If we bring in abstract “tissue-level” effects, like a morphogen gradient, we understand this as a description of a population of cells, either locally or, in the case of tissue-interaction strategies, a second tissue-level mechanism meshing with the first. In the case of two tissue-level HJMEx interacting, we may expect one of these to be a highly abstract generalisation of a paired tissue, like a generalisation of the morphogen gradient, for instance. In this case the intermeshing property may be located in the cells of a local tissue responding to distant production of a signal. We understand that all of these generalisations are “constituted” of cellular  $\text{xs } \phi\text{-ing}$ , which are in turn made of macromolecular  $\text{zs } \mu\text{-ing}$ . It therefore seems that there is nothing in our selected modelling approach that would contradict the foregoing discussion of what we understand by HJMEx in the MBT.

Having identified in my discussion of the SBE the agency or end-directedness of life as a good candidate for a counterinductive advance, and given my discussion of biosemiotics, I repeat the suggestion here that semiotic phenomena must be accounted for in a more sophisticated manner than allowed for by the cybernetic approach. In particular, it should allow for phenomena to be divided between what the eminent philosopher of science Charles Sanders Peirce identified as “dyadic”, or asemiotic, brute facts, and those identified as “triadic”, or legitimately semiotic, habitual phenomena. Although considerations of space prevent a recapitulation of this argument here, we may simply note that the NetLogo agent modelling approach does not restrict us from making this kind of representation. This implies that we should be able to compare relatively asemiotic logics like those of Harris’ EHJMEx to those that include explicit representation of semiotic phenomena, given a global model that includes semiosis, since semiotic phenomena can simply be removed to express Harris’ HJMEx as individual submodels within the global one. We therefore move to discuss the specifics of the statistical method involved in doing so.

## 14.4 The Systems Biology Encounter - The Molecular Biological Tradition Under Impact

### 14.4.1 Historical sketch of the biological encounter with complexity

### 14.4.2 Explanatory Models in the Stem Cell Biology Tradition

## 14.5 ”Metaphysical ingredients” encountered in MEx

### 14.5.1 Information theory

The modern field of information theory (Information Theoretical Tradition, ITT) was essentially founded by the American polymath Claude Shannon during his work at Bell labs in the 1940s, building on cryptographic research, but elaborating a novel fundamental theory of communication. Shannon used the thermodynamic concept of entropy to characterise the properties of entities involved in communication (sources, encoders, channels, etc).

It is important to immediately disabuse the reader of any notion that Shannon was at all concerned with meaning, or with the *contents* of communication<sup>31</sup>. Nowhere do the specific identities and contents of messages that habitually enter our minds when we think of biological communication enter Shannon’s

---

<sup>31</sup>Bioinformatically literate biologists are aware of this, but most of us do not routinely deal with formal definitions from information theory.

theory. Communication, for Shannon and for the electrical engineers for whom he was providing his theory, is an extremely broad concept. As Shannon's exegete Warren Weaver indicated in his classic introduction to Shannon's work, "it may be desireable to use a [broad] definition of communication, namely, one which would include the procedures by means of which one mechanism (say automatic equipment to track an airplane and to computer its probably future positions) affects another mechanism (say a guided missile chasing this airplane)" [SW63, p.10]. Therefore, Shannon's theory can be used to characterise some aspects of biological systems, but it must be clearly understood that "information" refers to an abstract physical quantity and not to the lay definition concerning the specific contents of some communication.

Broadly speaking, the ITT has been applied to biological systems in two ways: analysis of biological communication and statistical measures of biological order.

1. ITT analyses of biological communication Shannon's

#### 14.5.2 RPC Fate Specification as a "Stochastic" Process

Much of the complexity of this document is attributable to Harris' regular invocation of "stochastic" and related adjectives to describe the behaviour of RPCs. This is what lead me to characterise the theory primarily in those terms- the Intrinsic Stochastic Effects (ISE) EHJMEx. It may not be immediately obvious why this should be the case; most scientists assume that they know what words like "stochastic" and "random" mean well enough to use them in rigorous technical publications. We may not be aware that there has been a sprawling debate on the meaning of these terms since the earliest statistical formulations began to appear in the 19th century. However, even the simplest examples (as current today as they were in Laplace's time) reveal how difficult this topic can be.

If we consider the classic example of the coin flip, a process whose outcome we generally regard as being in some way "stochastic" or due to "chance", we immediately face the question of whether these descriptions refer to our inability to know the outcome of the process, or whether they refer to properties of the process itself. In other words, if we could specify the mechanics of the coin toss with sufficient precision, could we predict the outcome? This reflects two possible senses in which we may legitimately describe a process as "stochastic": referring to an epistemic dimension (we may describe some process as stochastic because we are unable to predict its outcome *a priori*), or referring to an ontological dimension (we describe the process as stochastic because this, in some way, describes how it *really is* independent of our knowledge of it).

Complicating matters is the sheer number of implications that we tend to associate with "stochasticity" and "randomness". We may be saying something about the causal structure of an event with deep metaphysical implications. It is common to distinguish between "deterministic" and "stochastic" processes, as though "stochastic" literally meant "indeterministic"- something like the Copenhagen interpretation of quantum physics. We may mean something about the apparent disorderliness of a series of outcomes of some process, with mathematical and information theoretical implications. What is an apparently simple observation- cellular fate distribution in RPC lineages is "stochastic", now seems to require at least a little clarification or interpretation.

Unfortunately, in Harris' case, a fulsome philosophically-inclined review of the ISE EHJMEx is yet to appear. We may attribute some of the difficulty encountered in interpreting the meaning of "stochasticity" to the cramped style necessary in scientific reports. In general, we may say that stochasticity, for

Harris, applies to at least the following entities:

1. The population-level phenomenal outcome of the RPC fate specification process (phenomenon P)
2. The overall behaviour of the macromolecular system whose operations produces these outcomes (mechanism S  $\Phi$ -ing)
3. The particular behaviour of some component of the macromolecular system, eg. stochastic expression of transcription factors (component x  $\phi$ -ing)
4. The statistical generalisations used to characterise relevant aspects of S  $\Phi$ -ing and x  $\phi$ -ing

It is, moreover, hardly fair to expect Harris to be advancing a coherent theory about the ontological, objective basis of randomness or probability. Indeed, there is no real agreement on these concepts in probability theory or amongst the philosophers. Still, this leaves us in the awkward position of not knowing quite what the leading EHJMEx for retinal formation is actually saying about its explanandum. The EHJMEx is thus at risk of circularity- the explanandum (unpredictable variability in clonal outcomes, P) has as explanans a MEx containing an abstract mathematical model tuned to produce this unpredictable variability. This may, in other words, turn out to be a convoluted case of model overfitting, if the “stochasticity” in question does not have a material biological referent. Before considering this, we need to define our terms more carefully to avoid the pervasive confusion mentioned above.

#### 14.5.2.1 Chance versus Randomness

A commonplace belief is that randomness refers to outcomes produced by chance events. In an extensive and useful discussion, Antony Eagle reviews the evidence for this Commonplace Thesis, or (CT)[Eag18], drawing on discussions in the PTT. Importantly, he notes that chance and randomness are not identical, and that one can conceivably exist without the other. This, in effect, disproves the (CT)- it is very difficult to imagine how the two concepts can be directly related in this productive fashion. I will attempt a brief summary of Eagle’s argument:

Chance is mainly used to refer to processes. Exemplars are coin flips and die rolls. We can think of these as “single-case” probabilities that we take to inhere in the process. For instance, we may say that an evenly weighted coin has a .5 probability of returning a value of heads on a flip, even if it is only flipped once. That is, probabilities can be taken to be objective properties of individual instances of processes, and not only descriptions of the frequencies of the process’ outcomes over many repetitions. This is closely related to the logical concept of “possibility”. If something is possible, it has a chance of occurring. However, possibility is a logical binary; something is either possible or impossible. A “single-case” probability is understood as something like an objective feature of a system as a whole given its actual configuration and the relevant natural laws.

Randomness, by contrast, mainly refers to process *outcomes*. That is, randomness is a property of a series of outcomes of multiple instances of some process. It turns out to be challenging merely to define what a “random” binary sequence might be (perhaps generated by a series of coin flips). However, in general, we may say that a random sequence of outcomes is one that cannot be generated by an description shorter than the sequence itself. That is, there is no set of rules that can generate a genuinely random sequence from a shorter sequence. In algorithmic information theory, the length of the ruleset required to produce some piece of information (like a sequence of measured outcomes) is

called the Kolmogorov complexity of that object; if the Kolmogorov complexity of the object is equal to the object's length, the object definitionally has the property of algorithmic or Kolmogorov randomness.

Eagle produces numerous examples of the dissociability of these concepts, from which I have selected two concise illustrations:

Chance Without Randomness

...

A fair coin, tossed 1000 times, has a positive chance of landing heads more than 700 times. But any outcome sequence of 1000 tosses which contains more than 700 heads will be compressible (long runs of heads are common enough to be exploited by an efficient coding algorithm, and 1000 outcomes is long enough to swamp the constants involved in defining the universal prefix-free Kolmogorov complexity). So any such outcome sequence will not be random, even though it quite easily could come about by chance.

...

Randomness Without Chance

...

Open or dissipative systems, those which are not confined to a state space region of constant energy, are one much studied class [of the objects of deterministic classical physics- notably, biological systems are dissipative], because such systems are paradigms of chaotic systems ... the behaviour of a chaotic system will be intuitively random ... [t]he sensitive dependence on initial conditions means that, no matter how accurate our finite discrimination of the initial state of a given chaotic system is, there will exist states indiscriminable from the initial state (and so consistent with our knowledge of the initial state), but which would diverge arbitrarily far from the actual evolution of the system. No matter, then, how well we know the initial condition (as long as we do not have infinite powers of discrimination)<sup>32</sup>., there is another state the system could be in for all we know that will evolve to a discriminably different future condition. Since this divergence happens relatively quickly, the system is unable to be predicted ... Just as before, the classical physical theory underlying the dynamics of these chaotic systems is one in which probability does not feature. [Eag18]

Therefore, the (CT) is untenable. Processes are “chancy”; collections of process outcomes, “trials”, or instantiations are “random”. It is tempting to say that Harris is explaining random fate outcomes with descriptions of chancy processes occurring internally to RPCs. Let us examine whether this is plausible.

#### 14.5.2.2 Chance in molecular mechanisms

I turn first to consider what it might mean to describe the behaviour of a biological macromolecular system as “chancy”. Let us again distinguish between the ontological and epistemic dimensions of this description. There is a sense in which mechanisms  $S \Psi\text{-ing}$  could be said to be objectively chancy, and one in which the “chancy” outcome reflects our ignorance of some source of variability in the process.

Eagle proffers two common lines of argument in favour of chanciness as an objective property of processes. The first is the notion of the “single-case” probability mentioned above. The examples given are single coin flips, and the decay of single radioactive atoms, which are commonly taken to have chancy outcomes irrespective of anyone's beliefs about them. As Eagle notes, this is closely related to statistician's ideas about stable processes, or trials:

---

<sup>32</sup>Note that this condition defines chaotic randomness as an epistemic, rather than ontological, feature of complex systems- a being with infinite powers of discrimination could predict the evolution of a complex classical system with perfect accuracy.

It is the stable trial principle that has the closest connection with single-case chance, however. For in requiring that duplicate trials should receive the same chances, it is natural to take the chance to be grounded in the properties of that trial, plus the laws of nature. It is quite conceivable that the same laws could obtain even if that kind of trial has only one instance, and the very same chances should be assigned in that situation. But then there are well-defined chances even though that type of event occurs only once.

...

The upshot of this discussion is that chance is a *process* notion, rather than being entirely determined by features of the outcome to which the surface grammar of chance ascriptions assigns the chance. For if there can be a single-case chance of  $\frac{1}{2}$  for a coin to land heads on a toss even if there is only one actual toss, and it lands tails, then surely the chance cannot be fixed by properties of the outcome ‘lands heads’, as that outcome does not exist. The chance must rather be grounded in features of the process that can produce the outcome: the coin-tossing trial, including the mass distribution of the coin and the details of how it is tossed, in this case, plus the background conditions and laws that govern the trial. Whether or not an event happens by chance is a feature of the process that produced it, not the event itself. The fact that a coin lands heads does not fix that the coin landed heads by chance, because if it was simply placed heads up, as opposed to tossed in a normal fashion, we have the same outcome not by chance. Sometimes features of the outcome event cannot be readily separated from the features of its causes that characterise the process by means of which it was produced.

[Eag18]

Examining the example of the coin toss, we find a fairly simple answer to the question posed earlier: if we knew enough about the mechanics of the toss, could we predict its outcome? The answer is yes, we can- the statistician Persi Diaconis has built a coin tossing machine that reliably produces heads or tails [Kes04]. We therefore know that tightly controlling the mechanics of a coin toss allows us to treat this system as entirely deterministic, without any significant element of chance in the outcome. A coin toss is only chancy when the human doing it does not have full control over the mechanical parameters of the process. Conceptually, there is no *a priori* reason why a coin-tosser should not be able to regularise the angular momentum of their thumb-flick by training with a strain gauge, place the coin on a stable surface allowing flicking, and achieve the same effect as the coin-tossing machine. In this case, Eagle’s suggestion that “[s]ometimes features of the outcome event cannot be readily separated from the features of its causes that characterise the process” seems obviously wrong- the “chancy” element of coin tossing is fully separable from the rest of the coin tossing process, and replaceable with a non-chancy component.

If the foregoing argument is correct, it seems that the coin toss is an example of the epistemic, rather than ontological, dimension of chance. The process appears to be chancy, or random, because the human tossing the coin is not able to precisely control the mechanical parameters of the process. Indeed, as Diaconis notes, these epistemically-limited tossers do not actually produce unbiased random outcomes- human coin tosses come up as they were started slightly more often than with the obverse face [DHM07].

Eagle’s second example of an “objective single-case chance”, is the decay of a radioactive atom. This is a common method of making covert appeals to the second line of argument for objective chance, which is the existence of orthodox quantum theory. There is no known physical process whose parameters are thought to define the lifetime of individual radioactive atoms, in the way that there is a well-specified physical process that produces a particular coin toss outcome. Rather, this is an appeal to the Copenhagen theoretical principle that it is *a priori* impossible to predict the lifetimes of individual atoms. As appeals to quantum theory to ground “objective chance” in biological processes are becoming

more common, let us consider whether a quantum theoretical explanation might plausibly underpin the “stochasticity” of RPC fate specification.

#### 14.5.2.3 Quantum indeterminacy - Is relevant to the ISE?

Eagle suggests that, because the Copenhagen interpretation of quantum physics has wide currency among physicists, the theory’s implied indeterminacy of physical phenomena at the quantum level could ground “objective chance”. While common, this argument downplays the fact that quantum theory is not a homogenous scientific tradition. Unfortunately, a significant misrepresentation of the QTT’s internal history has given rise to the impression that the Copenhagen theory is the unanimous or best articulation of quantum theory. We must briefly examine this misrepresentation before we can understand whether Eagle’s argument makes sense.

The conventional history of the mid 20th-century QTT holds that John Stewart Bell, in the demonstration of his famous inequality, conclusively proved that deterministic (so-called “hidden variable”) theories of quantum mechanics were incorrect. As demonstrated (strangely, without any acknowledgement) in the very pages Eagle’s argument appears in, this is a highly partisan and misleading view. Eminent Bohmian theorist Sheldon Goldstein conclusively demonstrates that Bell was an advocate of the deterministic Bohmian mechanical theory, and thought his famous inequality demonstrated that quantum phenomena could not be *local*, not that they could not be *deterministic*[Gol17].

Indeed, Bohmian quantum mechanics are fully deterministic, describe all of the same phenomena as Copenhagen, and in several cases resolve problems that orthodox quantum theory cannot [Gol17]. We are not, therefore, facing unanimous expert consensus that there is objective chance at the quantum level. We rather have a situation where physical phenomena are adequately described by two different traditional theories, one of which takes its statistical generalisations to be descriptions of ontological indeterminacy (Copenhagen), and the other to reflect epistemic uncertainty about a determinate universe (Bohm). Moreover, there is no reason to prefer the Copenhagen approach, given that the Bohmian theory explains Copenhagen’s paradoxical results “without further ado”[Gol17]<sup>33</sup>, deals with empirically verified phenomena of physical and biological interest that Copenhagen does not (eg. electron tunneling), and was the preferred approach of the man who understood better than anyone his own results, JS Bell.

Therefore, Eagle’s argument is incorrect. There is no reason to suppose that the existence of quantum theoretical models that posit objective chance is good evidence for the reality of objective chance. Moreover, there are good reasons to suppose that the converse is true. In sum, then, we may say that there is no reason to consider Harris’ argument to refer to *objective, ontological* chance, since the arguments for the existence of both single-case objective chance or quantum chance are weak and biologically irrelevant. Clearly, however, the *epistemological* dimension of chance is in play here.

#### 14.5.2.4 Randomness in RPC fate specification

Having dealt with how the concept of chance might apply to Harris’ ISE EHJMEx above, let us consider how the term “random” might relate to the process of RPC fate specification and differentiation. As introduced earlier, the technical meaning of “randomness” pertains to sequences of process outcomes.

---

<sup>33</sup>Bizarrely, Copenhagen partisans claim that Bohmian mechanics is formally equivalent to the Copenhagen approach. If this is the case, chance is *clearly* a function of model choices and not of any underlying ontological reality. However, Bohmian mechanics is, in fact, substantially more complete than its Copenhagen equivalent, which, by Eagle’s (defective) logic, suggests reality is more likely to reflect the deterministic rather than the chancy approach.

The process outcomes Harris is concerned with are the temporally-arranged fate outcomes of some particular RPC lineage. Therefore, we must ask whether these sequences meet any reasonable technical definition of “random”.

Harris’ own model proves that RPC fate outcomes are not algorithmically random. That is, the sequence of outcomes has a structure that can be meaningfully compressed by rules which produce typical RPC fate outcomes (Harris’ mathematical models are such rule sets). One might object that Harris’ meaning is that the particular rules which give rise to cellular fate “choice” in his models involve random number generation. In this case, the claim is trivially about the model and not about the sequence of outcomes that is actually observed in zebrafish eyes. Indeed, all of Harris’ later models *axiomatically assume* a tripartite temporal structure to the differentiation process<sup>34</sup>. This is precisely the type of sequential bias which allows efficient compression of a non-random sequence of outcomes by an algorithm. Therefore, Harris himself concedes that RPC fate specification is not an algorithmically random process<sup>35</sup>.

We should further note that the question of how ordered, which is to say, non-random processes like fate specification arise in biological systems is a fundamental question of the biological sciences. It has long been recognised that classical and quantum physical systems which have algorithmically random initial conditions do not spontaneously evolve to a state of order. In many ways, then, it is the extent to which variable sequences of outcomes like RPC fate specification *depart* from algorithmic randomness which is of interest when we are asking questions like “how does the ordered structure of a retina arise from RPC activity?”

#### 14.5.2.5 Summary: “Stochastic” or “variable”?

Above, I argue that the RPC fate specification process is not objectively chancy (since objective chance is an empirically unsupported concept), nor random (since the sequence of RPC lineage outcomes is structured and therefore non-random). What, then, should we make of the argument that this process is “stochastic”? Let us consider the forceful argument of the great Bayesian statistician Edwin Thompson Jaynes:

“Belief in the existence of ‘stochastic processes’ in the real world; i.e. that the property of being ‘stochastic’ rather than ‘deterministic’ is a real physical property of a process, that exists independently of human information, is [an] example of the mind projection fallacy: attributing one’s own ignorance to Nature instead. The current literature of probability theory is full of claims to the effect that a ‘Gaussian random process’ is fully determined by its first and second moments. If it were made clear that this is only the defining property for an abstract mathematical model, there could be no objection to this; but it is always presented in verbiage that implies that one is describing an objectively true property of a real physical process. To one who believes such a thing literally, there could be no motivation to investigate the causes more deeply than noting the first and second moments, and so the real processes at work might never be discovered. This is not only irrational because one is throwing away the very information that is essential to understand the physical process; if carried into practice it can have disastrous consequences. Indeed, there is no such thing as a ‘stochastic process’ in the sense that the individual events have no specific causes.” [JBE03]

---

<sup>34</sup>That is, an early bias in RPC production is produced in these models by the a priori commitment to a “rule” which results in early RPC production.

<sup>35</sup>Having debunked the lay sense of “random” being equivalent to “chancy” above, there is no reason to consider these other, confused, non-technical definitions of randomness.

It is important to emphasize that the utility of stochastic modelling techniques should not be taken to suggest that on some level the modelled phenomenon is actually, i.e. irreducibly, random and without causal structure. When speaking of biological “randomness” or “stochasticity”, biologists rarely precisely define what is meant by these terms. This vagueness sometimes arises from or results in a theoretical deficit where properties of statistical models are understood to directly reflect the system being modelled; the scientist has failed to heed Korbzysky’s dictum insisting that “a map *is not* the territory it represents, but, if correct, it has a *similar structure* to the territory, which accounts for its usefulness” [Kor05] (italics in original). The “structural similarity” here is between the model’s outcomes and the collection of actually-observed population outcomes, *not* the underlying biological process giving rise to measured outcomes.

However, it would be trivial and a violation of my commitment to the principle of charity to dismiss Harris’ argument on this basis. While I do believe that Harris has fallen into the “mind-projection fallacy” James excoriates above, we must ask how this occurred. What was the attraction of this fallacious explanation in the first place? I suggest that Harris appeals to stochasticity as an explanation for *variability* in fate outcomes. Harris describes his work as an attempt to explain how variable clonal outcomes could give rise to an “invariant” retina. We want to know how it is that variable clonal outcomes are produced and what this phenomenon’s significance is to overall retinal formation. The value of Harris’ ISE EHJMEx, I suggest, lies in its emphasis on the unpredictable variability of RPC lineage outcomes. I have suggested a scheme for identifying possible sources of biological variability in a separate section of this chapter.

### 14.5.3 Biological Variability

As I have argued in section 14.5.2, there are no good reasons to think that differential outcomes in biological systems are a result of some fundamental, ontological randomness. I share Jaynes’ objection to this argument- the result is that the so-described phenomenon is understood to be, on some level, causeless. We are nevertheless left with the fact of unpredictable biological variability. Of course, observations documenting these variable phenomena are perplexing in the context of the pervasive Monodian conventional view, to wit, that cells are logically equivalent to cybernetic chemical factories. In a factory, process control is fundamentally about managing variability. Normally, we want to produce some regular sequence of outcomes for a production line with minimal variability. The chemical factory metaphor requires modification when applied to the context of Harris’ work, however<sup>36</sup>. The products we are interested in are cellular fate outcomes themselves. Starting with an aggregate of similar (genetically, if not epigenetically, identical) RPCs, a specification hierarchy emerges. To re-fit the industrial metaphor to the context, the chemical-factory equivalent would be to explain how one can start with a factory that makes machine tools and end with a group of factories that make the diverse chemical products of a highly networked modern chemical industrial sector. In the absence of any central planner, how might the ontogenesis of such an industry take place? What are the factors that give rise to the “decision” to build one type of factory or another, and how do we explain the apparent increase in networked complexity and functional specialisation? What is the source of this apparent scaffolding novelty, and how can we explain the specific outcomes that we observe?<sup>37</sup>

---

<sup>36</sup>This modification is a logical consequence of the transposition of Monod’s cybernetic metaphor for bacterial operons to a multicellular system.

<sup>37</sup>The factory metaphor makes it obvious that this type of question has far more to do with developmental or historical economics than it does with cybernetic process control of petrochemical cracking vessels or fermenters. Aside from Markoff

Indeed, an explanation of biological variability is, alongside explaining the emergence of biological order, the fundamental task of the biological sciences. Harris has bravely advanced a theory that has, at times, referenced different possible sources for the variability observed in

#### **14.5.3.1 Noise**

#### **14.5.4 Time**

#### **14.5.5 Semiosis, Meaning**

#### **14.5.6 Emergence, self-organisation**

#### **14.5.7 Scale Hierarchies, Specification Hierarchies**

### **14.5.8 Waddington's topological model of development**

C. H. Waddington was one (and almost certainly the best remembered) of the embryologists who so profited. He wrote: “During the recent war, engineers attained some facility in designing machines to carry out tasks which earlier generations would have considered beyond the capacities of anything but an intelligent being . . . The ideas suggested by these self-regulating mechanisms are both very relevant to biology and rather novel.”<sup>28</sup> Waddington himself worked in the Operation Research Section of the Royal Air Force Coastal Command, and it was from his own experience and from that of his friends in self-steering gunnery that he learned to draw the analogy which was to become increasingly familiar in the cybernetics of the 1950s and 1960s: “The behaviours of an automatic pilot, of a target-tracking gun-sight, or of an embryo, all exhibit the characteristics of an activity guided by a purpose.”<sup>29</sup> Indeed, it was at this time, and in this context, that his work on canalization began.<sup>30</sup> In Waddington’s first introduction of the term, he wrote, “The main thesis is that developmental reactions as they occur in organization submitted to natural selection, are, in general, canalized. That is to say, they are adjusted so as to bring about one definite end result regardless of minor variations in conditions during the course of the reaction.” In his view, canalization is built into the organism by natural selection as a consequence of its obvious advantages: “It ensures the production of the normal, that is, optimal type in the face of the unavoidable hazards of existence.”<sup>31</sup> Canalization was a term Waddington had borrowed from his reading of Alfred North Whitehead, and the concept clearly accorded with much of his own prewar thinking about “epigenetic landscapes.”<sup>32</sup> But it was only after the war that he began to envision the possibility of a theoretical account of such characteristic features of biological organization. An explanation of “developmental canalization,” he wrote, requires supplementing conventional gene theory with an “epigenetic theory”—one in which discrete and separate entities of classical genetics would be displaced by collections of genes which could ‘lock in’ development through their interactions.”<sup>33</sup> In other words, an account of developmental stability needs to be sought in the complex system of reactions that make up the developmental process. The search for quantitative models displaying such behavior underlay much of Waddington’s theoretical efforts well into the 1970s. However, he soon concluded that the particular models developed by Ross Ashby and other cyberneticians on self-organizing systems were not really appropriate to biological development. Instead, he concentrated on feedback models of cross-reacting systems of metabolic reactions. Yet he did not have a great deal of success with these models. Indeed, it is not his theoretical

---

chain theory, biologists have generally avoided methods used to study systems where outcomes heavily depend on the history of the system and related contingencies, probably because of the lack of scientific prestige associated with many of these formalisms (“soft science”). Nonetheless, it is hard to see how one can explain tissue development exclusively in terms of control of cell-internal processes—attempt this is like trying to explain the historical evolution of the biotechnological industry in terms of the PID controllers it uses to run its fermenters. This will be, at best, a partial and highly misleading account.

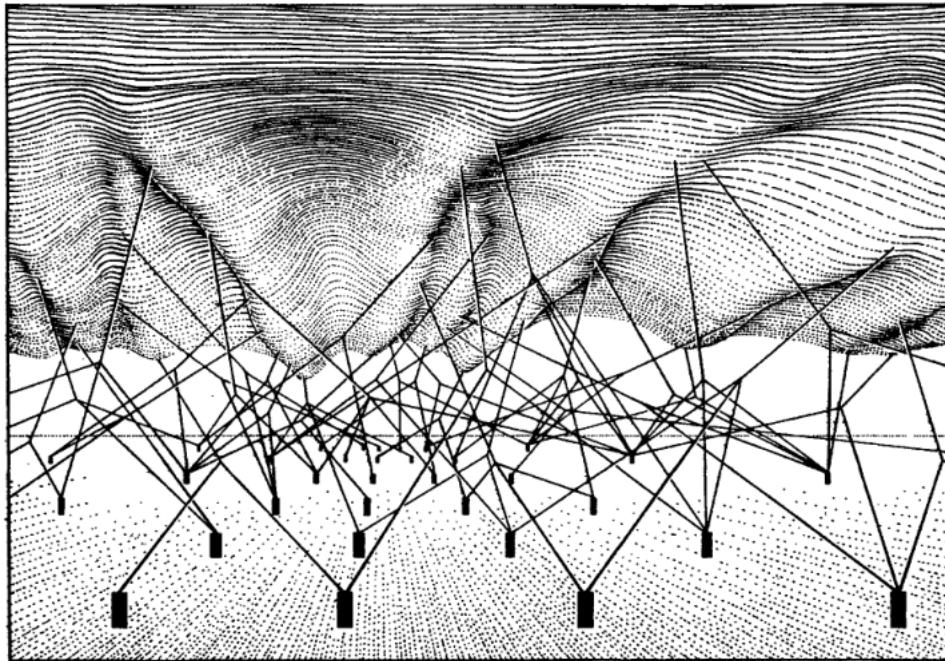


Figure 14.5: The complex system of interactions underlying the epigenetic landscape, excerpted from [Wad57]. The pegs in the ground represent genes; the strings leading from them the chemical tendencies which the genes produce. The modelling of the epigenetic landscape, which slopes down from above one's head towards the distance, is controlled by the pull of these numerous guy-ropes which are ultimately anchored to the genes.

work but the experimental work from his laboratory in the 1940s and '50s that now, with the benefit of hindsight, attracts the most interest. [Kel00, p.117-119]

#### 14.5.9 Morphogenetic fields

#### 14.5.10 Physical explanations

### 14.6 Implications of Limits in Theory and Practice For Model Selection

#### 14.6.1 Nicholas Rescher's Account of Scientific Progress

#### 14.6.2 Implications of Theoretical Limits of Science

#### 14.6.3 Implications of Thermo-economic Limits of Science

#### 14.6.4 Practical computational limits

# Bibliography

- [ABS<sup>+</sup>10] W. Ted Allison, Linda K. Barthel, Kristina M. Skebo, Masaki Takechi, Shoji Kawamura, and Pamela A. Raymond. Ontogeny of cone photoreceptor mosaics in zebrafish. *The Journal of Comparative Neurology*, 518(20):4182–4195, October 2010.
- [AH09] Michalis Agathocleous and William A. Harris. From progenitors to differentiated cells in the vertebrate retina. *Annual Review of Cell and Developmental Biology*, 25:45–69, 2009.
- [AHW<sup>+</sup>20] Afnan Azizi, Anne Herrmann, Yinan Wan, Salvador JRP Buse, Philipp J Keller, Raymond E Goldstein, and William A Harris. Nuclear crowding and nonlinear diffusion during interkinetic nuclear migration in the zebrafish retina. 9(58635):31, 2020.
- [ALHP07] Michalis Agathocleous, Morgane Locker, William A. Harris, and Muriel Perron. A General Role of Hedgehog in the Regulation of Proliferation. *Cell Cycle*, 6(2):156–159, January 2007.
- [AR08] Ruben Adler and Pamela A. Raymond. Have we achieved a unified model of photoreceptor cell fate specification in vertebrates? *Brain Research*, 1192:134–150, February 2008.
- [BA02] Kenneth P. Burnham and David Raymond Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, 2nd ed edition, 2002.
- [BA13] Rafael Bravo and David E Axelrod. A calibrated agent-based computer model of stochastic cell dynamics in normal human colon crypts useful for in silico experiments. *Theoretical Biology and Medical Modelling*, 10(1):66, 2013.
- [BB20] M.J. Brewer and A. Butler. Model selection and the cult of AIC. In *Departmental Seminar*, University of Wollongong, Australia, February 2020.
- [BGL<sup>+</sup>10] Klaus A. Becker, Prachi N. Ghule, Jane B. Lian, Janet L. Stein, Andre J. van Wijnen, and Gary S. Stein. Cyclin D2 and the CDK substrate p220<sup>NPAT</sup> are required for self-renewal of human embryonic stem cells. *Journal of Cellular Physiology*, 222(2):456–464, February 2010.
- [Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.
- [BJ79] D. H. Beach and Marcus Jacobson. Influences of thyroxine on cell proliferation in the retina of the clawed frog at different ages. *The Journal of Comparative Neurology*, 183(3):615–623, February 1979.

- [BNL<sup>+</sup>96] M. Burmeister, J. Novak, M. Y. Liang, S. Basu, L. Ploder, N. L. Hawes, D. Vidgen, F. Hoover, D. Goldman, V. I. Kalnins, T. H. Roderick, B. A. Taylor, M. H. Hankin, and R. R. McInnes. Ocular retardation mouse caused by Chx10 homeobox null allele: Impaired retinal progenitor proliferation and bipolar cell differentiation. *Nature Genetics*, 12(4):376–384, April 1996.
- [Bra99] Michael E. Bratman. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press, Cambridge, 1999.
- [BRD<sup>+</sup>15] Henrik Boije, Steffen Rulands, Stefanie Dudczig, Benjamin D. Simons, and William A. Harris. The Independent Probabilistic Firing of Transcription Factors: A Paradigm for Clonal Variability in the Zebrafish Retina. *Developmental Cell*, 34(5):532–543, September 2015.
- [CAH<sup>+</sup>14] L. Centanin, J.-J. Ander, B. Hoeckendorf, K. Lust, T. Kellner, I. Kraemer, C. Urbany, E. Hasel, W. A. Harris, B. D. Simons, and J. Wittbrodt. Exclusive multipotency and preferential asymmetric divisions in post-embryonic neural stem cells of the fish retina. *Development*, 141(18):3472–3482, September 2014.
- [CAI<sup>+</sup>04] Brenda LK Coles, Brigitte Angénieux, Tomoyuki Inoue, Katia Del Rio-Tsonis, Jason R. Spence, Roderick R. McInnes, Yvan Arsenijevic, and Derek van der Kooy. Facile isolation and the characterization of human retinal stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(44):15772–15777, 2004.
- [CAR<sup>+</sup>15] Renee W. Chow, Alexandra D. Almeida, Owen Randlett, Caren Norden, and William A. Harris. Inhibitory neuron migration and IPL formation in the developing zebrafish retina. *Development*, 142(15):2665–2677, August 2015.
- [Cav18] Florencia Cavodeassi. Dynamic Tissue Rearrangements during Vertebrate Eye Morphogenesis: Insights from Fish Models. *Journal of Developmental Biology*, 6(1):4, February 2018.
- [CAY<sup>+</sup>96] Constance L. Cepko, Christopher P. Austin, Xianjie Yang, Macrene Alexiades, and Diala Ezzeddine. Cell fate determination in the vertebrate retina. *Proceedings of the National Academy of Sciences*, 93(2):589–595, 1996.
- [CBR03] Michel Cayouette, Ben A. Barres, and Martin Raff. Importance of intrinsic mechanisms in cell fate decisions in the developing rat retina. *Neuron*, 40(5):897–904, December 2003.
- [CC07] Bo Chen and Constance L Cepko. Requirement of histone deacetylase activity for the expression of critical photoreceptor genes. *BMC Developmental Biology*, 7(1):78, 2007.
- [CCY<sup>+</sup>05] Florencia Cavodeassi, Filipa Carreira-Barbosa, Rodrigo M. Young, Miguel L. Concha, Miguel L. Allende, Corinne Houart, Masazumi Tada, and Stephen W. Wilson. Early Stages of Zebrafish Eye Formation Require the Coordinated Activity of Wnt11, Fz5, and the Wnt/β-Catenin Pathway. *Neuron*, 47(1):43–56, July 2005.
- [CHW11] Lázaro Centanin, Burkhard Hoeckendorf, and Joachim Wittbrodt. Fate Restriction and Multipotency in Retinal Stem Cells. *Cell Stem Cell*, 9(6):553–562, December 2011.

- [CYV<sup>+</sup>08] Anna M. Clark, Sanghee Yun, Eric S. Veien, Yuan Y. Wu, Robert L. Chow, Richard I. Dorsky, and Edward M. Levine. Negative regulation of Vsx1 by its paralog Chx10/Vsx2 is conserved in the vertebrate retina. *Brain Research*, 1192:99–113, February 2008.
- [Dar88] Charles Darwin. *The Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life*. J. Murray, London :, 1888.
- [DC00] Michael A. Dyer and Constance L. Cepko. Control of Müller glial cell proliferation and activation following retinal injury. *Nature Neuroscience*, 3(9):873–880, September 2000.
- [DCRH97] R. I. Dorsky, W. S. Chang, D. H. Rapaport, and W. A. Harris. Regulation of neuronal diversity in the Xenopus retina by Delta signalling. *Nature*, 385(6611):67–70, January 1997.
- [DFWV<sup>+</sup>97] MARCO Di Frusco, Hans Weiher, Barbara C. Vanderhyden, Takashi Imai, Tadahiro Shiomii, T. A. Hori, Rudolf Jaenisch, and Douglas A. Gray. Proviral inactivation of the Npat gene of Mpv 20 mice results in early embryonic arrest. *Molecular and cellular biology*, 17(7):4080–4086, 1997.
- [DH05] T. A. Down and Tim J.P. Hubbard. NestedMICA: Sensitive inference of over-represented motifs in nucleic acid sequence. *Nucleic Acids Research*, 33(5):1445–1453, March 2005.
- [DHM07] Persi Diaconis, Susan Holmes, and Richard Montgomery. Dynamical Bias in the Coin Toss. *SIAM Review*, 49(2):211–235, January 2007.
- [DPCH03] Tilak Das, Bernhard Payer, Michel Cayouette, and William A. Harris. In vivo time-lapse imaging of cell divisions during neurogenesis in the developing zebrafish retina. *Neuron*, 37(4):597–609, 2003.
- [DRH95] Richard I Dorsky, David H Rapaport, and William A Harris. Xotch inhibits cell differentiation in the xenopus retina. *Neuron*, 14(3):487–496, March 1995.
- [Eag18] Antony Eagle. Chance versus Randomness. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition, 2018.
- [EHML09] Ted Erclik, Volker Hartenstein, Roderick R. McInnes, and Howard D. Lipshitz. Eye evolution at high resolution: The neuron as a unit of homology. *Developmental Biology*, 332(1):70–79, August 2009.
- [ESY<sup>+</sup>17] Peter Engerer, Sachihiro C Suzuki, Takeshi Yoshimatsu, Prisca Chapouton, Nancy Obeng, Benjamin Odermatt, Philip R Williams, Thomas Misgeld, and Leanne Godinho. Uncoupling of neurogenesis and differentiation during retinal development. *The EMBO Journal*, 36(9):1134–1146, May 2017.
- [Fag12] Melinda Bonnie Fagan. Waddington redux: Models and explanation in stem cell and systems biology. *Biology & Philosophy*, 27(2):179–213, March 2012.
- [Fag13] Melinda Bonnie Fagan. *Philosophy of Stem Cell Biology: Knowledge in Flesh and Blood*. New Directions in the Philosophy of Science. Palgrave Macmillan, Hounds mills, Basingstoke, Hampshire, 2013.

- [Fag15a] Melinda Fagan. Crucial stem cell experiments? Stem cells, uncertainty, and single-cell experiments. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 30(2):183, June 2015.
- [Fag15b] Melinda Bonnie Fagan. Collaborative explanation and biological mechanisms. *Studies in History and Philosophy of Science Part A*, 52:67–78, August 2015.
- [FEF<sup>+</sup>09] Curtis R. French, Timothy Erickson, Danielle V. French, David B. Pilgrim, and Andrew J. Waskiewicz. Gdf6a is required for the initiation of dorsal–ventral retinal patterning and lens development. *Developmental Biology*, 333(1):37–47, September 2009.
- [Fey93] Paul Feyerabend. *Against Method*. Verso, London ; New York, 3rd ed edition, 1993.
- [FHB09] F. Feroz, M. P. Hobson, and M. Bridges. MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, October 2009.
- [FR00] Andy J. Fischer and Thomas A. Reh. Identification of a Proliferating Marginal Zone of Retinal Progenitors in Postnatal Chickens. *Developmental Biology*, 220(2):197–210, April 2000.
- [FR03] Andy J. Fischer and Thomas A. Reh. Potential of Müller glia to become neurogenic retinal progenitor cells: Retinal Müller Glia as a Source of Stem Cells. *Glia*, 43(1):70–76, July 2003.
- [GBB<sup>+</sup>03] G. Gao, A. P. Bracken, K. Burkard, D. Pasini, M. Classon, C. Attwooll, M. Sagara, T. Imai, K. Helin, and J. Zhao. NPAT Expression Is Regulated by E2F and Is Essential for Cell Cycle Progression. *Molecular and Cellular Biology*, 23(8):2821–2833, April 2003.
- [GDL<sup>+</sup>09] Prachi N. Ghule, Zbigniew Dominski, Jane B. Lian, Janet L. Stein, Andre J. van Wijnen, and Gary S. Stein. The subnuclear organization of histone gene regulatory proteins and 3' end processing factors of normal somatic and embryonic stem cells is compromised in selected human cancer cell types. *Journal of Cellular Physiology*, 220(1):129–135, July 2009.
- [Geh96] Walter J. Gehring. The master control gene for morphogenesis and evolution of the eye. *Genes to Cells*, 1(1):11–15, January 1996.
- [GNB08] Nathan J. Gosse, Linda M. Nevin, and Herwig Baier. Retinotopic order in the absence of axon competition. *Nature*, 452(7189):892–895, April 2008.
- [Gol17] Sheldon Goldstein. Bohmian Mechanics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition, 2017.
- [GS07] Frédéric Gaillard and Yves Sauvé. Cell-based therapy for retina degeneration: The promise of a cure. *Vision Research*, 47(22):2815–2824, October 2007.

- [GTR<sup>+</sup>10] D. M. Gilbert, S.-I. Takebayashi, T. Ryba, J. Lu, B. D. Pope, K. A. Wilson, and I. Hiratani. Space and Time in the Nucleus: Developmental Control of Replication Timing and Chromosome Architecture. *Cold Spring Harbor Symposia on Quantitative Biology*, 75(0):143–153, January 2010.
- [GWC<sup>+</sup>07] Leanne Godinho, Philip R. Williams, Yvonne Claassen, Elayne Provost, Steven D. Leach, Maarten Kamermans, and Rachel O.L. Wong. Nonapical Symmetric Divisions Underlie Horizontal Cell Layer Formation in the Developing Retina In Vivo. *Neuron*, 56(4):597–603, November 2007.
- [GZC<sup>+</sup>11] Francisco L. A. F. Gomes, Gen Zhang, Felix Carbonell, José A. Correa, William A. Harris, Benjamin D. Simons, and Michel Cayouette. Reconstruction of rat retinal progenitor cell lineages in vitro reveals a surprising degree of stochasticity in cell fate decisions. *Development (Cambridge, England)*, 138(2):227–235, January 2011.
- [HBEH88] Christine E. Holt, Thomas W. Bertsch, Hillary M. Ellis, and William A. Harris. Cellular Determination in the Xenopus Retina is Independent of Lineage and Birth Date. *Neuron*, 1(1):15–26, March 1988.
- [HCG95] G Halder, P Callaerts, and W. Gehring. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science*, 267(5205):1788–1792, March 1995.
- [HE99] Minjie Hu and Stephen S. Easter. Retinal Neurogenesis: The Formation of the Initial Central Patch of Postmitotic Cells. *Developmental Biology*, 207(2):309–321, March 1999.
- [Hea67] D.F. Heath. Normal or Log-normal: Appropriate Distributions. *Nature*, 213:1159–1160, March 1967.
- [HH91] William A. Harris and Volker Hartenstein. Neuronal determination without cell division in *xenopus* embryos. *Neuron*, 6:499–515, 1991.
- [HHHL19] Edward Higson, Will Handley, Michael Hobson, and Anthony Lasenby. Dynamic nested sampling: An improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, 29(5):891–913, September 2019.
- [HKB08] Herbert Hoijtink, Irene Klugkist, and Paul A. Boelen, editors. *Bayesian Evaluation of Informative Hypotheses*. Statistics for Social and Behavioral Sciences. Springer, New York, 2008.
- [Hor04] D. J. Horsford. Chx10 repression of Mitf is required for the maintenance of mammalian neuroretinal identity. *Development*, 132(1):177–187, December 2004.
- [HP98] William A. Harris and Muriel Perron. Molecular recapitulation: The growth of the vertebrate retina. *International Journal of Developmental Biology*, 42:299–304, 1998.
- [HSK<sup>+</sup>13] Steven A. Harvey, Ian Sealy, Ross Kettleborough, Fruzsina Fenyes, Richard White, Derek Stemple, and James C. Smith. Identification of the zebrafish maternal and paternal transcriptomes. *Development*, 140(13):2703–2710, July 2013.

- [HYSL11] Hongpeng He, Fa-Xing Yu, Chi Sun, and Yan Luo. CBP/p300 and SIRT1 Are Involved in Transcriptional Regulation of S-Phase Specific Histone Genes. *PLoS ONE*, 6(7):e22088, July 2011.
- [HZA<sup>+</sup>12] Jie He, Gen Zhang, Alexandra D. Almeida, Michel Cayouette, Benjamin D. Simons, and William A. Harris. How Variable Clones Build an Invariant Retina. *Neuron*, 75(5):786–798, September 2012.
- [ICW13] Kenzo Ivanovitch, Florencia Cavodeassi, and Stephen W. Wilson. Precocious Acquisition of Neuroepithelial Character in the Eye Field Underlies the Onset of Eye Morphogenesis. *Developmental Cell*, 27(3):293–305, November 2013.
- [IKRN16] Jaroslav Icha, Christiane Kunath, Mauricio Rocha-Martins, and Caren Norden. Independent modes of ganglion cell translocation ensure correct lamination of the zebrafish retina. *The Journal of Cell Biology*, 215(2):259–275, October 2016.
- [Ioa05] John P. A. Ioannidis. Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8):e124, August 2005.
- [IYS<sup>+</sup>96] T Imai, M Yamauchi, N Seki, T Sugawara, T Saito, Y Matsuda, H Ito, T Nagase, N Nomura, and T Hori. Identification and characterization of a new gene physically linked to the ATM gene. *Genome Research*, 6(5):439–447, May 1996.
- [JBE03] Edwin T Jaynes, G. Larry Bretthorst, and EBSCO Publishing. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.
- [Kan06] Kunihiko Kaneko. *Life: An Introduction to Complex Systems Biology*. Understanding Complex Systems. Springer, Berlin ; New York, 2006.
- [Kel00] Evelyn Fox Keller. *The Century of the Gene*. Harvard University Press, Cambridge, Mass, 2000.
- [Kes04] David Kestenbaum. The Not So Random Coin Toss. <https://www.npr.org/templates/story/story.php?storyId=1697475>, 2004.
- [Kit84] Philip Kitcher. 1953 and all That. A Tale of Two Sciences. *The Philosophical Review*, 93(3):335, July 1984.
- [KKT<sup>+</sup>13] Vijayalakshmi Kari, Oleksandra Karpiuk, Bettina Tieg, Malte Kriegs, Ekkehard Dikomey, Heike Krebber, Yvonne Begus-Nahrmann, and Steven A. Johnsen. A Subset of Histone H2B Genes Produces Polyadenylated mRNAs under a Variety of Cellular Conditions. *PLoS ONE*, 8(5):e63745, May 2013.
- [Kon06] Toru Kondo. Epigenetic alchemy for cell fate conversion. *Current Opinion in Genetics & Development*, 16(5):502–507, October 2006.
- [Kor05] Alfred Korzybski. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Inst. of General Semantics, Brooklyn, N.Y, 5. ed., 3. print edition, 2005.

- [Koz08] Zbynek Kozmik. The role of Pax genes in eye evolution. *Brain Research Bulletin*, 75(2-4):335–339, March 2008.
- [KSN99] Nathan L. Kleinman, James C. Spall, and Daniel Q. Naiman. Simulation-Based Optimization with Stochastic Approximation Using Common Random Numbers. *Management Science*, 45(11):1570–1578, 1999.
- [LAA<sup>+</sup>06] M. Locker, M. Agathocleous, M. A. Amato, K. Parain, W. A. Harris, and M. Perron. Hedgehog signaling and the retina: Insights into the mechanisms controlling the proliferative properties of neural precursors. *Genes & Development*, 20(21):3036–3048, November 2006.
- [Lar92] Ellen W. Larsen. Tissue strategies as developmental constraints: Implications for animal evolution. *Trends in Ecology & Evolution*, 7(12):414–417, December 1992.
- [LD09] Enhu Li and Eric H. Davidson. Building developmental gene regulatory networks. *Birth Defects Research Part C: Embryo Today: Reviews*, 87(2):123–130, June 2009.
- [LHKR08] Deepak A. Lamba, Susan Hayes, Mike O. Karl, and Thomas Reh. Baf60c is a component of the neural progenitor-specific BAF complex in developing retina. *Developmental Dynamics*, 237(10):3016–3023, September 2008.
- [LHO<sup>+</sup>00] Zheng Li, Minjie Hu, Małgorzata J. Ochocinska, Nancy M. Joseph, and Stephen S. Easter. Modulation of cell proliferation in the embryonic retina of zebrafish (*Danio rerio*). *Developmental Dynamics*, 219(3):391–401, 2000.
- [LM87] E. Larsen and H. M. G. McLaughlin. The morphogenetic alphabet. Lessons for simple-minded genes. *BioEssays*, 7(3):130–132, September 1987.
- [LWR<sup>+</sup>07] Julie Lessard, Jiang I. Wu, Jeffrey A. Ranish, Mimi Wan, Monte M. Winslow, Brett T. Staahl, Hai Wu, Ruedi Aebersold, Isabella A. Graef, and Gerald R. Crabtree. An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*, 55(2):201–215, July 2007.
- [Lyo13] Louis Lyons. Discovering the Significance of 5 sigma. *arXiv:1310.1284 [hep-ex, physics:hep-ph, physics:physics]*, October 2013.
- [Ma00] T. Ma. Cell cycle-regulated phosphorylation of p220NPAT by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. *Genes & Development*, 14(18):2298–2313, September 2000.
- [MAA<sup>+</sup>01] Till Marquardt, Ruth Ashery-Padan, Nicole Andrejewski, Raffaella Scardigli, Francois Guillemot, and Peter Gruss. Pax6 Is Required for the Multipotent State of Retinal Progenitor Cells. *Trends in Biochemical Sciences*, 30(3):i, 2001.
- [MAB<sup>+</sup>13] Gary R. Mirams, Christopher J. Arthurs, Miguel O. Bernabeu, Rafel Bordas, Jonathan Cooper, Alberto Corrias, Yohan Davit, Sara-Jane Dunn, Alexander G. Fletcher, Daniel G. Harvey, Megan E. Marsh, James M. Osborne, Pras Pathmanathan, Joe Pitt-Francis, James Southern, Nejib Zemzemi, and David J. Gavaghan. Chaste: An Open Source C++ Library

- for Computational Physiology and Biology. *PLoS Computational Biology*, 9(3):e1002970, March 2013.
- [MBE<sup>+</sup>07] Karine Massé, Surinder Bhamra, Robert Eason, Nicholas Dale, and Elizabeth A. Jones. Purine-mediated signalling triggers eye development. *Nature*, 449(7165):1058–1062, October 2007.
- [MDBN<sup>+</sup>05] Juan-Ramon Martinez-Morales, Filippo Del Bene, Gabriela Nica, Matthias Hammerschmidt, Paola Bovolenta, and Joachim Wittbrodt. Differentiation of the Vertebrate Retina Is Coordinated by an FGF Signaling Center. *Developmental Cell*, 8(4):565–574, April 2005.
- [MFKM12] Kiran Mahajan, Bin Fang, John M Koomen, and Nupam P Mahajan. H2B Tyr37 phosphorylation suppresses expression of replication-dependent core histone genes. *Nature Structural & Molecular Biology*, 19(9):930–937, August 2012.
- [MGv<sup>+</sup>09] Partha Mitra, Prachi N. Ghule, Margaretha van der Deen, Ricardo Medina, Rong-lin Xie, William F. Holmes, Xin Ye, Keiichi I. Nakayama, J. Wade Harper, Janet L. Stein, Gary S. Stein, and Andre J. van Wijnen. CDK inhibitors selectively diminish cell cycle controlled activation of the histone H4 gene promoter by p220<sup>NPAT</sup> and HiNF-P. *Journal of Cellular Physiology*, 219(2):438–448, May 2009.
- [MHR<sup>+</sup>99] Nicholas Marsh-Armstrong, Haochu Huang, Benjamin F Remo, Tong Tong Liu, and Donald D Brown. Asymmetric Growth and Development of the Xenopus laevis Retina during Metamorphosis Is Controlled by Type III Deiodinase. *Neuron*, 24(4):871–878, December 1999.
- [ML15] Nicolas Malagon and Ellen Larsen. Heredity and Self-Organization: Partners in the Generation and Evolution of Phenotypes. In *International Review of Cell and Molecular Biology*, volume 315, pages 153–181. Elsevier, 2015.
- [MMDM04] Kathryn B. Moore, Kathleen Mood, Ira O. Daar, and Sally A. Moody. Morphogenetic Movements Underlying Eye Field Formation Require Interactions between the FGF and ephrinB1 Signaling Pathways. *Developmental Cell*, 6(1):55–67, January 2004.
- [Mor09] Michel Morange. A new revolution? The place of systems biology and synthetic biology in the history of biology. *EMBO Reports*, 10(Suppl 1):S50–S53, August 2009.
- [Mor11] Michel Morange. Recent opportunities for an increasing role for physical explanations in biology. *Studies in History and Philosophy of Biol & Biomed Sci*, 42(2):139–144, 2011.
- [MT60] E. A. McCulloch and J. E. Till. The Radiation Sensitivity of Normal Mouse Bone Marrow Cells, Determined by Quantitative Marrow Transplantation into Irradiated Mice. *Radiation Research*, 13(1):115, July 1960.
- [MXM<sup>+</sup>03] Partha Mitra, Rong-Lin Xie, Ricardo Medina, Hayk Hovhannisyan, S. Kaleem Zaidi, Yue Wei, J. Wade Harper, Janet L. Stein, André J. van Wijnen, and Gary S. Stein. Identification of HiNF-P, a Key Activator of Cell Cycle-Controlled Histone H4 Genes at the Onset of S Phase. *Molecular and Cellular Biology*, 23(22):8110–8123, November 2003.

- [MZLF02] A Murciano, J Zamora, J Lopezsanchez, and J Fraude. Interkinetic Nuclear Movement May Provide Spatial Clues to the Regulation of Neurogenesis. *Molecular and Cellular Neuroscience*, 21(2):285–300, October 2002.
- [Neu00] C. J. Neumann. Patterning of the Zebrafish Retina by a Wave of Sonic Hedgehog Activity. *Science*, 289(5487):2137–2139, September 2000.
- [NLM89] R. S. Nowakowski, S. B. Lewin, and M. W. Miller. Bromodeoxyuridine immunohistochemical determination of the lengths of the cell cycle and the DNA-synthetic phase for an anatomically defined population. *Journal of Neurocytology*, 18(3):311–318, June 1989.
- [NYLH09] Caren Norden, Stephen Young, Brian A. Link, and William A. Harris. Actomyosin Is the Main Driver of Interkinetic Nuclear Migration in the Retina. *Cell*, 138(6):1195–1208, September 2009.
- [PBS<sup>+</sup>18] Martina Pesaresi, Sergi A. Bonilla-Pons, Giacoma Simonte, Daniela Sanges, Umberto Di Vicino, and Maria Pia Cosma. Endogenous Mobilization of Bone-Marrow Cells Into the Murine Retina Induces Fusion-Mediated Reprogramming of Müller Glia Cells. *EBioMedicine*, 0(0), February 2018.
- [PEM<sup>+</sup>09] David M. Parichy, Michael R. Elizondo, Margaret G. Mills, Tiffany N. Gordon, and Raymond E. Engeszer. Normal table of postembryonic zebrafish development: Staging by externally visible anatomy of the living fish. *Developmental Dynamics*, 238(12):2975–3015, December 2009.
- [PJ10] J. Pirngruber and S. A. Johnsen. Induced G1 cell-cycle arrest controls replication-dependent histone mRNA 3' end processing through p21, NPAT and CDK9. *Oncogene*, 29(19):2853–2863, 2010.
- [PKVH98] Muriel Perron, Shami Kanekar, Monica L. Vetter, and William A Harris. The genetic sequence of retinal development in the ciliary margin of the *xenopus* eye. *Developmental Biology*, (199):185–200, 1998.
- [PSS<sup>+</sup>09] Judith Pirngruber, Andrei Shchebet, Lisa Schreiber, Efrat Shema, Neri Minsky, Rob D Chapman, Dirk Eick, Yael Aylon, Moshe Oren, and Steven A Johnsen. CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO reports*, 10(8):894–900, August 2009.
- [RBBP06] Pamela A. Raymond, Linda K. Barthel, Rebecca L. Bernardos, and John J. Perkowski. Molecular characterization of retinal stem cells and their niches in adult zebrafish. *BMC developmental biology*, 6(1):36, 2006.
- [RDT<sup>+</sup>01] J. T. Rasmussen, M. A. Deardorff, C. Tan, M. S. Rao, P. S. Klein, and M. L. Vetter. Regulation of eye development by frizzled signaling in *Xenopus*. *Proceedings of the National Academy of Sciences*, 98(7):3861–3866, March 2001.
- [Res00] Nicholas Rescher. *Nature and Understanding*. Oxford University Press, New York, 2000.
- [Res05] Nicholas Rescher. *Cognitive Harmony*. University of Pittsburgh Press, Pittsburgh, PA, 2005.

- [RO04] Jonathan M. Raser and Erin K. O’Shea. Control of Stochasticity in Eukaryotic Gene Expression. *Science; Washington*, 304(5678):1811–4, June 2004.
- [Rv08] Arjun Raj and Alexander van Oudenaarden. Stochastic gene expression and its consequences. *Cell*, 135(2):216–226, October 2008.
- [Sad97] Payman Sadegh. Constrained Optimization via Stochastic Approximation with a Simultaneous Perturbation Gradient Approximation. *Automatica*, 33(5):889–892, May 1997.
- [Sch93] Kenneth F. Schaffner. *Discovery and Explanation in Biology and Medicine*. Science and Its Conceptual Foundations. University of Chicago Press, Chicago, 1993.
- [SF03] Deborah L Stenkamp and Ruth A Frey. Extraretinal and retinal hedgehog signaling sequentially regulate retinal differentiation in zebrafish. *Developmental Biology*, 258(2):349–363, June 2003.
- [SH14] Corinna Singleman and Nathalia G. Holtzman. Growth and Maturation in the Zebrafish, *Danio Rerio* : A Staging Tool for Teaching and Research. *Zebrafish*, 11(4):396–406, August 2014.
- [Ski06] John Skilling. Nested Sampling for Bayesian Computations. In *Proc. Valencia*, Benidorm (Alicante, Spain), June 2006. inference.org.uk.
- [Ski12] John Skilling. Bayesian computation in big spaces-nested sampling and Galilean Monte Carlo. In *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, pages 145–156, Waterloo, Ontario, Canada, 2012.
- [Spa98] J. C. Spall. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on Aerospace and Electronic Systems*, 34(3):817–823, July 1998.
- [SSV<sup>+</sup>16] Daniela Sanges, Giacoma Simonte, Umberto Di Vicino, Neus Romo, Isabel Pinilla, Marta Nicolás, and Maria Pia Cosma. Reprogramming Müller glia via in vivo cell fusion regenerates murine photoreceptors. *The Journal of Clinical Investigation*, 126(8):3104–3116, August 2016.
- [STN<sup>+</sup>02] Masashi Sagara, Eri Takeda, Akiyo Nishiyama, Syunsaku Utsumi, Yoshiroh Toyama, Shigeki Yuasa, Yasuharu Ninomiya, and Takashi Imai. Characterization of functional regions for nuclear localization of NPAT. *The Journal of Biochemistry*, 132(6):875–879, 2002.
- [SVT13] Lourdes Serrano, Berta N Vazquez, and Jay Tischfield. Chromatin structure, pluripotency and differentiation. *Experimental Biology and Medicine*, 238(3):259–270, March 2013.
- [SW63] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1963.

- [Swe97] Rod Swenson. Evolutionary Theory Developing: The Problem(s) With Darwin’s Dangerous Idea. *Ecological Psychology*, 9(1):47–96, March 1997.
- [TC87] David L. Turner and Constance L. Cepko. A common progenitor for neurons and glia persists in rat retina late in development. *Nature*, 328(9), July 1987.
- [TK06] Dmitry N. Tsigankov and Alexei A. Koulakov. A unifying model for activity-dependent and activity-independent mechanisms predicts complete structure of topographic maps in ephrin-A deficient mice. *Journal of Computational Neuroscience*, 21(1):101–114, August 2006.
- [TMS64] J. E. Till, E. A. Mcculloch, and L. Siminovitch. A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. *Proceedings of the National Academy of Sciences of the United States of America*, 51:29–36, January 1964.
- [TR86] Sally Temple and Martin C. Raff. Clonal analysis of oligodendrocyte development in culture: Evidence for a developmental clock that counts cell divisions. *Cell*, 44(5):773–779, March 1986.
- [TR14] W.-W. Tee and D. Reinberg. Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development*, 141(12):2376–2390, June 2014.
- [Tro00] V. Tropepe. Retinal Stem Cells in the Adult Mammalian Eye. *Science*, 287(5460):2032–2036, March 2000.
- [Tro08] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, March 2008.
- [TSC90] D. L. Turner, E. Y. Snyder, and C. L. Cepko. Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron*, 4(6):833–845, June 1990.
- [TSC08] Jeffrey M. Trimarchi, Michael B. Stadler, and Constance L. Cepko. Individual Retinal Progenitor Cells Display Extensive Heterogeneity of Gene Expression. *PLOS ONE*, 3(2):e1588, February 2008.
- [VGV<sup>+</sup>18] Jessie Van houcke, Emiel Geeraerts, Sophie Vanhunsel, An Beckers, Lut Noterdaeme, Marjike Christiaens, Ilse Bollaerts, Lies De Groef, and Lieve Moons. Extensive growth is followed by neurodegenerative pathology in the continuously expanding adult zebrafish retina. *Biogerontology*, October 2018.
- [VJM<sup>+</sup>09] Marta Vitorino, Patricia R Jusuf, Daniel Maurus, Yukiko Kimura, Shin-ichi Higashijima, and William A Harris. Vsx2 in the zebrafish retina: Restricted lineages through derepression. *Neural Development*, 4(1):14, 2009.
- [VL54] V. Vilter and L. Lewis. [Existence and distribution of mitoses in the retina of the deepsea fish Bathylagus benedicti]. - PubMed - NCBI. *C R Seances Soc Biol Fil.*, 148(21-22):1771–5, 1954.
- [vM77] L. v. Salvini-Plawen and Ernst Mayr. On the Evolution of Photoreceptors and Eyes. In Max K. Hecht, William C. Steere, and Bruce Wallace, editors, *Evolutionary Biology*, pages 207–263. Springer US, Boston, MA, 1977.

- [Wad57] C H Waddington. *The Strategy of the Genes : A Discussion of Some Aspects of Theoretical Biology*. Routledge Taylor & Francis Group, London New York, 1957.
- [Wag07] Günter P. Wagner. The developmental genetics of homology. *Nature Reviews Genetics*, 8(6):473–479, 2007.
- [WAR<sup>+</sup>16] Y. Wan, A. D. Almeida, S. Rulands, N. Chalour, L. Muresan, Y. Wu, B. D. Simons, J. He, and W. A. Harris. The ciliary marginal zone of the zebrafish retina: Clonal and time-lapse analysis of a continuously growing tissue. *Development*, 143(7):1099–1107, April 2016.
- [Wes00] M. Westerfield. The Zebrafish Book : A Guide for the Laboratory Use of Zebrafish. [http://zfin.org/zf\\_info/zfbook/zfbk.html](http://zfin.org/zf_info/zfbook/zfbk.html), 2000.
- [WF88] R. Wetts and S. E. Fraser. Multipotent precursors can give rise to all major cell types of the frog retina. *Science*, 239(4844):1142–1145, March 1988.
- [WG92] Robert W Williams and Dan Goldowitz. Lineage versus environment in embryonic retina: A revisionist perspective. *TINS*, 15(10):6, 1992.
- [WIEO04] Aiyan Wang, Tsuyoshi Ikura, Kazuhiro Eto, and Masato S. Ota. Dynamic interaction of p220NPAT and CBP/p300 promotes S-phase entry. *Biochemical and Biophysical Research Communications*, 325(4):1509–1516, December 2004.
- [Wik18] Wikipedia. Stochastic process. *Wikipedia*, July 2018.
- [Win15] R. Winklbauer. Cell adhesion strength from cortical tension - an integration of concepts. *Journal of Cell Science*, 128(20):3687–3693, October 2015.
- [WJH03] Y. Wei, J. Jin, and J. W. Harper. The Cyclin E/Cdk2 Substrate and Cajal Body Component p220NPAT Activates Histone Transcription through a Novel LisH-Like Domain. *Molecular and Cellular Biology*, 23(10):3669–3680, May 2003.
- [WR90] Takashi Watanabe and Martin C. Raff. Rod photoreceptor development in vitro: Intrinsic properties of proliferating neuroepithelial cells change as development proceeds in the rat retina. *Neuron*, 4(3):461–467, March 1990.
- [WR09] L. L. Wong and D. H. Rapaport. Defining retinal progenitor cell competence in *Xenopus laevis* by clonal analysis. *Development*, 136(10):1707–1715, May 2009.
- [WSM<sup>+</sup>05] Ann M. Wehman, Wendy Staub, Jason R. Meyers, Pamela A. Raymond, and Herwig Baier. Genetic dissection of the zebrafish retinal stem-cell compartment. *Developmental Biology*, 281(1):53–65, May 2005.
- [WSP<sup>+</sup>09] Minde I. Willardsen, Arminda Suli, Yi Pan, Nicholas Marsh-Armstrong, Chi-Bin Chien, Heithem El-Hodiri, Nadean L. Brown, Kathryn B. Moore, and Monica L. Vetter. Temporal regulation of Ath5 gene expression during eye development. *Developmental Biology*, 326(2):471–481, February 2009.
- [Yam05] M. Yamaguchi. Histone deacetylase 1 regulates retinal neurogenesis in zebrafish by suppressing Wnt and Notch signaling pathways. *Development*, 132(13):3027–3043, July 2005.

- [YDH<sup>+</sup>11] Jingye Yang, Huzefa Dungrawala, Hui Hua, Arkadi Manukyan, Lesley Abraham, Wesley Lane, Holly Mead, Jill Wright, and Brandt L. Schneider. Cell size and growth rate are major determinants of replicative lifespan. *Cell Cycle*, 10(1):144–155, January 2011.
- [You85] Richard W. Young. Cell differentiation in the retina of the mouse. *The Anatomical Record*, 212(2):199–205, June 1985.
- [YQW<sup>+</sup>18] Kai Yao, Suo Qiu, Yanbin V. Wang, Silvia J. H. Park, Ethan J. Mohns, Bhupesh Mehta, Xinran Liu, Bo Chang, David Zenisek, Michael C. Crair, Jonathan B. Demb, and Bo Chen. Restoration of vision after de novo genesis of rod photoreceptors in mammalian retinas. *Nature*, August 2018.
- [YWNH03] X. Ye, Y. Wei, G. Nalepa, and J. W. Harper. The Cyclin E/Cdk2 Substrate p220NPAT Is Required for S-Phase Entry, Histone Gene Expression, and Cajal Body Maintenance in Human Somatic Cells. *Molecular and Cellular Biology*, 23(23):8586–8600, December 2003.
- [ŽCC<sup>+</sup>05] Mihaela Žigman, Michel Cayouette, Christoforos Charalambous, Alexander Schleiffer, Oliver Hoeller, Dara Dunican, Christopher R. McCudden, Nicole Firnberg, Ben A. Barres, David P. Siderovski, and Juergen A. Knoblich. Mammalian Inscuteable Regulates Spindle Orientation and Cell Fate in the Developing Retina. *Neuron*, 48(4):539–545, November 2005.
- [ZDI<sup>+</sup>98] Jiyong Zhao, Brian Dynlacht, Takashi Imai, Tada-aki Hori, and Ed Harlow. Expression of NPAT, a novel substrate of cyclin E–CDK2, promotes S-phase entry. *Genes & development*, 12(4):456–461, 1998.
- [ZKL<sup>+</sup>00] Jiyong Zhao, Brian K. Kennedy, Brandon D. Lawrence, David A. Barbie, A. Gregory Matera, Jonathan A. Fletcher, and Ed Harlow. NPAT links cyclin E–Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes & development*, 14(18):2283–2297, 2000.
- [Zub03] M. E. Zuber. Specification of the vertebrate eye by a network of eye field transcription factors. *Development*, 130(21):5155–5167, August 2003.