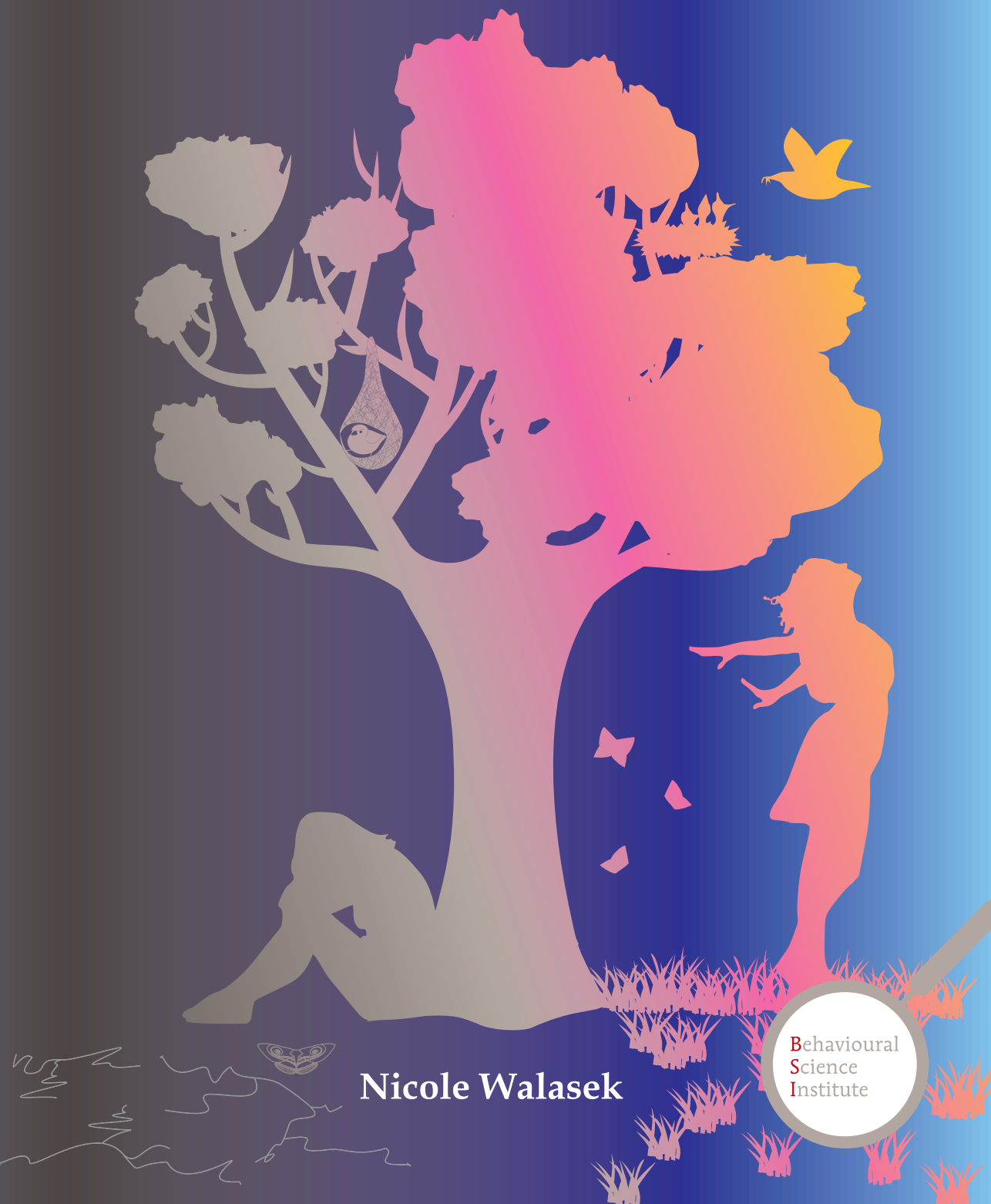
The evolution and development of sensitive periods: Theoretical and statistical approaches



Nicole Walasek

Behavioural Science Institute

The evolution and development of sensitive periods: Theoretical and statistical approaches

Nicole Walasek

The research presented in this dissertation was carried out at the Behavioural Science Institute of Radboud University Nijmegen, the Netherlands. This research was financially supported by grants from the Dutch Research Council (V1.Vidi.195.130) to WEF, the James S. McDonnell Foundation (https://doi.org/10.37717/220020502) to WEF, and the Jacobs Foundation (2017 1261 02) to WEF.

ISBN
978-94-6423-814-3

Cover design
Anne Morbach | www.schlaugemacht.net

Printing and lay-out
ProefschriftMaken | proefschriftmaken.nl

The evolution and development of sensitive periods: Theoretical and statistical approaches

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 12 september 2022
om 16.30 uur precies

door

Nicole Walasek
geboren op 18 februari 1992
te Nordhorn, Duitsland

Promotor:

Prof. dr. A.H.N. Cillessen

Copromotoren:
Dr. W.E. Frankenhuis (Universiteit Utrecht)
Dr. K. Panchanathan (University of Missouri, Verenigde Staten)

Manuscriptcommissie:
Prof. dr. R. Kievit
Prof. dr. M. Mangel (University of California, Santa Cruz, Verenigde Staten)
Dr. S. English (University of Bristol, Verenigd Koninkrijk)

The evolution and development of sensitive periods: Theoretical and statistical approaches

Dissertation

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Doctorate Board
to be defended in public on

Monday, September 12, 2022
at 16:30

by

Nicole Walasek
born on February 18, 1992
in Nordhorn, Germany

PhD supervisor:

Prof. dr. A.H.N. Cillessen

PhD co-supervisors:
Dr. W.E. Frankenhuis (Utrecht University)
Dr. K. Panchanathan (University of Missouri, USA)

Manuscript Committee:
Prof. dr. R. Kievit
Prof. dr. M. Mangel (University of California, Santa Cruz, USA)
Dr. S. English (University of Bristol, UK)

Table of contents

Chapter 1

Introduction

1.1 Sensitive periods: windows of opportunity during development?

Sensitive periods are times (or life stages) in which an organism's phenotypic development is more affected by experience than at other times (or stages). They exist across the entire tree of life and play a crucial role in development (Bradshaw, 1965; DeWitt & Scheiner, 2004; Nijhout, 2015; Schlichting et al., 1998). Sensitive periods result from developmental plasticity, i.e., the capacity of a genotype to produce different phenotypes depending on experience (Nettle & Bateson, 2015; Stearns, 1989). Invertebrates, for instance, may develop armored defensive morphologies in the presence of predators (e.g., water fleas develop protective helmets in predator-dense environments), but not when few predators are present (Agrawal et al., 1999). In humans, early life adversity and stress are known to have long lasting and often adverse effects on the development of children (Belsky & Pluess, 2009; Ellis & Boyce, 2008; Lin et al., 2020). Research on sensitive periods tends to assume that plasticity is costly, for example, due to costs of building and maintaining the required neural-cognitive machinery (DeWitt et al., 1998). These costs may result in periods of enhanced plasticity as opposed to lifelong plasticity. Without costs, all organisms should be "Darwinian demons," which always perfectly tailor their phenotypes to local conditions to increase survival and reproductive success (i.e. fitness) (Law, 1979). Such organisms do not exist, even if some species exhibit lifelong plasticity in certain traits.

Rather certain characteristics, such as the onset, timing and duration of sensitive periods, are not fixed (Bornstein, 1989; Fawcett & Frankenhuis, 2015). Instead, they vary between species, between individuals within a species, and between traits within a single individual. For example, zebra finches learn songs only early in life, while European starlings are life-long learners (Kelly et al., 2018; Mountjoy & Lemon, 1995). Zooming in on a species, we see individual differences. For example, human adolescents vary in the extent to which exposure to environmental unpredictability affects their rate of maturation (Belsky & Pluess, 2009; Del Giudice et al., 2011). A close-up of a single individual reveals variation between different traits, for instance, in the plasticity of cognitive versus emotional systems following adoption (e.g., from a harsh orphanage into a supportive family) (Tottenham et al., 2010; Zeanah et al., 2011). What explains this variability in sensitive periods between species, individuals, and traits?

1.2 Lessons learned from empirical work

We have learned much about the biological and neurophysiological mechanisms underlying variation in sensitive periods (Creanza et al., 2016; Knudsen, 2004). For decades the predominant view held among researchers was that the course of our development is set early in life. While early life is still considered as a sensitive period for many developmental traits, research has also identified sensitive periods during later developmental stages. Across different traits and species adolescence has emerged as such a 'later' sensitive period. For example, in some mammalian species experiences during adolescence shape adult social behavior to a larger extent than childhood experiences (Buwalda et al., 2011; Mutwill et al., 2020; Sachser et al., 2020). In humans, adolescence functions as a sensitive period for various neural and cognitive traits, such as the ability to cope with social stress,

memory formation, and developing learning strategies (Blakemore & Mills, 2014; Dahl, 2004; Fuhrmann et al., 2015; Knoll et al., 2016; Raab & Hartley, 2019).

Understanding when during development plasticity is enhanced is only the first step. The next step is to capitalize on this understanding to manipulate when and how individuals are shaped by experiences. An extreme case of such manipulation would be to reopen sensitive periods for specific experiences during development. Hypothetically, this could allow adults to learn languages with the same ease as children. While this is not possible to date, neuroscientists have developed experimental interventions to reopen sensitive periods for some traits and species (e.g., visual development in rats) (Hensch & Bilimoria, 2012; Reh et al., 2020). Such interventions range from invasive biochemical and surgical interventions, over less invasive electrical stimulation, to non-invasive environmental changes. The ability to manipulate the onset, timing, and duration of sensitive periods in this way has vast implications for developmental research. Research on sensitive periods outlines the possibility to reverse the effects of psychological and physiological trauma in human and non-human animals.

1.3 The evolution of sensitive periods

Despite progress in understanding the physiological mechanisms of sensitive periods, we know little about their evolution. Under what environmental conditions should we expect sensitive periods to evolve early in life as opposed to later during ontogeny (i.e., the stage that is relevant for an organism's development)? Or, more generally, how do environmental conditions shape levels of plasticity across an organism's development?

Historically, evolution and development have often been viewed as opposing forces in a zero-sum relationship (Frankenhuis & Fraley, 2017). Nowadays, we acknowledge that these processes are nested, albeit operating on different time scales (S. P. Wilson & Prescott, 2022). Organisms adapt to ecologies across generations through natural selection (evolutionary timescale) and within their individual lifetimes through development and learning (developmental timescale) (Frankenhuis & Fraley, 2017; Frankenhuis & Walasek, 2020). Natural selection shapes the developmental and learning systems, which tailor individuals to local conditions based on experience. In subsequent generations natural selection acts on the resulting variation in phenotypes (Frankenhuis et al., 2013). Considering evolution and development together can provide valuable insights into the study of sensitive periods. It can help us understand how environmental conditions across evolutionary timescales result in developmental systems that produce sensitive periods within developmental timescales.

Such insights may help us understand why some environmental changes, such as enrichment (i.e., increasing variation in environmental stimuli), can reverse abnormal visual development in rats (Hensch & Bilimoria, 2012). For example, recent modelling has identified conditions that may favor the evolution of enhanced plasticity at later developmental stages. Natural selection may favor such a pattern of plasticity, when experiences later during development are more informative about environmental conditions than early experiences (Walasek et al., 2021). In that sense, an enriched environment may provide

better information about the state of the environment. Enhanced plasticity might be a physiological reaction to process this 'valuable' information. Evolutionary insights can also help us identify other environmental conditions (e.g., changes in the environmental state), unrelated to enrichment, that select for enhanced plasticity and are thus 'candidates' for reopening sensitive periods. Subsequent empirical work could then design experimental studies (e.g., in rats) to test whether these conditions actually reopen sensitive periods.

1.4 Modelling as a tool to study evolution and development

Models lie on a continuum. They can range from highly specific models that make quantitative predictions to abstract models of processes that generate qualitative patterns (Servedio et al., 2014). Here, I am focusing on mathematical and computational models that are used to study evolutionary and developmental processes. Such models can provide insight into species-typical development and individual differences within species (Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017; Frankenhuis & Walasek, 2020). Modelling allows us to test existing theories and to generate new predictions and theories to fuel future empirical work (Otto & Rosales, 2020; Servedio et al., 2014). By design, these models do not include all possible variables. Rather, they are simplified versions of reality that capture only some essential components of a process or system (Frankenhuis et al., 2013; Frankenhuis & Tiokhin, 2018; Houston et al., 1988; Levins, 1966; Reimer et al., 2019; Smaldino, 2017). Each model makes particular assumptions that may be criticized. The models developed as part of this dissertation, for instance, assume that organisms can develop phenotypes that perfectly match their environment. This phenotypic gambit allows researchers to ignore matters of genetic and physiological realization (Fawcett et al., 2013; Frankenhuis et al., 2013; Frankenhuis & Tiokhin, 2018). If observations contradict model predictions, we need to refine our model. Statistician George Box famously remarked that all models are wrong but some are useful (Box, 1976). Simple models are useful because they make assumptions clear and explicit, remove ambiguities from natural language, and ensure logical consistency in argumentation (Borsboom et al., 2021; Frankenhuis & Tiokhin, 2018; Smaldino, 2017).

1.5 Lessons learned from modelling

A growing body of theoretical work studies the conditions in which plasticity is favored by natural selection over non-plastic developmental strategies. From this work we have learned that plasticity depends, among other things, on the stability of environmental conditions. Environmental conditions that are stable across and within generations select for fixed, non-plastic developmental strategies. Plasticity can be adaptive in environments that vary across and within generations (DeWitt & Scheiner, 2004; Schlichting et al., 1998). If, relative to the species' lifespan, conditions fluctuate rapidly between generations, but slowly within lifetimes, natural selection might favor plasticity (Botero et al., 2015; Snell-Rood & Steck, 2019; Stephens, 1991). If, however, the environment fluctuates rapidly within generations, it may be too costly for organisms to continuously adjust their phenotypes based on experience, favoring non-plastic strategies (Leung et al., 2020; Pfab et al., 2016). Alternatively, plasticity may be prolonged if the costs of being mismatched to the environment

outweigh costs of being plastic (English et al., 2016; Panchanathan & Frankenhuis, 2016; Pascalis et al., 2020).

In recent years, there has also been a surge in models used to explore when during development enhanced plasticity should be favored. These models do not start out assuming sensitive periods. Rather, sensitive periods might be favored by natural selection in response to the environmental conditions explored. Existing models found that patterns of sensitive periods depend on uncertainty about environmental conditions. Plasticity tends to be higher early in life when an organism's uncertainty about her environment at birth is higher and the better experiences ('cues') during development help to reduce this uncertainty (English et al., 2016; Fischer et al., 2014; Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Stamps & Krishnan, 2014a, 2014b, 2017). The extent to which cues reflect environmental conditions is called the 'cue reliability'. The cue reliability determines the cue's scope to reduce an organism's uncertainty. The higher the cue reliability, the more likely it is that plasticity 'closes' before the end of development, resulting in zero plasticity at the end of ontogeny. Such periods of heightened plasticity are called 'critical' periods. When cues are noisy and do not help to reduce uncertainty, plasticity declines more slowly across ontogeny (Frankenhuis et al., 2016). These patterns resonate with empirical findings. For example, studies in humans, birds, and rodents have shown that noisy inconsistent inputs tend to prolong sensitive periods (Chen et al., 2020; Freund et al., 2013; Tooley et al., 2021).

Insights gained from theoretical work has enriched empirical approaches to studying sensitive periods. For example, models by Frankenhuis and Panchanathan predicted that individual differences in phenotypes become more stable across development (Frankenhuis & Panchanathan, 2011a, 2011b; Panchanathan & Frankenhuis, 2016). This inspired studies of fish, which support this prediction (Bierbach et al., 2017; Kok et al., 2019). In addition to generating new predictions, such models can also be used to provide explanations of existing empirical findings about the development of sensitive periods. For example, the well-known observation that in most traits plasticity tends to decline over the lifespan is consistent with the results of current models (Frankenhuis & Fraley, 2017).

1.6 Limitations of existing models

As noted earlier, sensitive periods do not only occur early in life and sometimes may even be prolonged until the end of development. Existing models of sensitive period evolution offer little insight into conditions favoring enhanced plasticity at later developmental stages. However, researchers have speculated that natural selection might favor sensitive periods later during development when organisms experience variation in two main factors: the extent to which cues can reduce uncertainty about the environment ('cue reliability') and the environmental state itself (Fawcett & Frankenhuis, 2015). All but one of the existing models assume that the reliability of cues and the environmental state are constant within the lifetime of an organism. However, the reliability of cues may change across development when organisms' sensory systems mature, when the frequency of cues changes across development, or when the availability of some cues is restricted to specific developmental stages. For example, prenatal cues provide estimates about nutritional

conditions expected outside the womb (Kuzawa, 2005). Information through this channel is not available to the newborn and thus exclusive to the fetal stage. Also, most long-lived species do experience changes in environmental conditions within their lifetimes. For example, organisms experience environmental changes when they migrate across different habitats or due to variation in weather, season, and climate.

To our knowledge, there exists only one model of sensitive period evolution which assumes that the environmental state itself fluctuates within an organism's lifetime (Fischer et al., 2014). However, this model assumes unbounded phenotypic development, such that organisms can develop any phenotype at any age. This assumption does not apply to traits that develop incrementally, i.e., one step at a time, or irreversibly. Incremental and irreversible development constrains the range of phenotypes available at different developmental stages. As the end of development approaches, the organism has less time left to make phenotypic adjustments, limiting the range of realizable phenotypes.

Incremental and irreversible development is widespread in nature. For instance, plants gradually adjust the shape of leaves (e.g. area, thickness, and dissection) in response to environmental conditions, such as light intensity, humidity, and temperature (Callahan et al., 1997; Maugarny-Calès & Laufs, 2018; Schlichting, 1986). Animals, incrementally and often irreversibly, develop morphological defenses, such as protective armor, or increased body size in response to predator cues (Agrawal et al., 1999). In humans, the development of motor skills, such as sitting, standing, climbing, and walking, appears stepwise if measures are taken across weeks or months. However, this pattern reflects smaller incremental changes, which are visible once measures are taken frequently on shorter time scales (Adolph et al., 2008). Despite the ubiquity of incremental and irreversible development, only a few models have explored their consequences for the evolution and development of sensitive periods (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016). No models have simultaneously explored varying cue reliabilities or fluctuating environments. My dissertation fills this gap.

1.7 Bridging theoretical and empirical studies of sensitive periods

In recent years the value of modeling is increasingly being recognized across different disciplines (Borsboom et al., 2021; Frankenhuis & Tiokhin, 2018; Fried, 2020; Grainger et al., 2021; Guest & Martin, 2021; Muthukrishna & Henrich, 2019; Smaldino, 2017, 2020; van Rooij & Baggio, 2021). However, the links between theory and data are still relatively weak and there exists resistance among empiricists to engage more with theoretical work. This reduces the scope for synergies between theoretical and empirical studies. One reason for resistance towards incorporating existing theoretical work, is that some theoretical papers use technical language and jargon, that is not accessible for an empirical audience (Martínez & Mammola, 2021). One reason for resistance towards engaging in hands-on theory development is that modeling can seem intimidating because it usually requires computational and mathematical skills that social scientists are rarely trained in (Borsboom et al., 2021; Grainger et al., 2021; Smaldino, 2020). In other cases, empiricists might discard models altogether as too simplistic to be valuable for the problems they study. Many

researchers have engaged with this argument, emphasizing the value of simple models for research (e.g., Frankenhuis & Tiokhin, 2018; Smaldino, 2017). Still, there exist various ways in which modelers can do more to better integrate their work with empirical studies.

One way to achieve better integration is to develop models that are tailored towards specific phenomena and incorporate known mechanisms of those phenomena. For example, empirical research on sensitive periods distinguishes between experience-expectant and experience-dependent plasticity. This distinction is rarely adopted by evolutionary biologists (Frankenhuis & Nettle, 2020). Experience-expectant plasticity uses neural mechanisms prepared to process information that is available to all members of a species around the same developmental stage (Frankenhuis & Nettle, 2020; Gabard-Durnam & McLaughlin, 2019; Gabard-durnam & Mclaughlin, 2020; Greenough et al., 1987). Visual and auditory development, for example, are shaped by experience-expectant plasticity. Experience-dependent mechanisms process information that is unique to an individual, potentially resulting in brain changes in response to those experiences. However, not all traits neatly fall into either of these two categories (Frankenhuis & Walasek, 2020). An evolutionary model that incorporates these known mechanisms could help answer under what conditions we expect natural selection to favor traits that are experience-expectant, experience-dependent, or both.

Similarly, models could incorporate experimental features of empirical studies. For example, plasticity is quantified in various ways across empirical studies. Plasticity in humans is often quantified by comparing individuals who have been adopted (or have migrated) at different ages with each other (Mascie-Taylor & Little, 2004; Pallier, 2003; Zeanah et al., 2011). Plasticity in non-human animals (e.g., rodents, birds) allows for a greater degree of experimental control. These studies often use paradigms, such as cross-fostering, in which animals are transferred at different ages between different caregivers or patches (Breed & Moore, 2015), or dose-dependent experience paradigms that systematically vary the duration and intensity of exposure to particular stimuli (Groothuis & Taborsky, 2015). Some existing models of sensitive period evolution use paradigms, like adoption or dose-dependent experience studies, to achieve a closer correspondence between their models and empirical studies (Panchanathan & Frankenhuis, 2016; Stamps & Krishnan, 2014a, 2014b, 2017).

A different approach to forming stronger connection between models and data does not involve changing aspects of the model but the input to the model. Consider a model of sensitive periods which assumes that the environment fluctuates within an organism's lifetime (e.g., Fischer et al., 2014). Rather than exploring all possible rates of environmental fluctuations, such a model could focus on those rates that are known to be relevant for a specific species or trait. In this way, the results of the model would be more relevant to researchers studying these species or traits. For example, higher rates of environmental fluctuations may make it more challenging for organisms to predict their environment. Current levels of environmental unpredictability have been linked to how populations respond to novel environmental conditions (Bitter et al., 2021). Drawing on evolutionary theory, Bitter et al. (2021) have identified the extent to which animals can predict environmental fluctuations as a potential driver of phenotypic variation in novel

environments. Such variation may be key in ensuring survival in the face of climate change. Knowing the values of unpredictability statistics for different species can thus help to predict their response to novel conditions, such as those caused by climate change.

Although statistical properties of the environment are commonly measured in non-human animals and plants (Burgess & Marshall, 2011; Vasseur & Yodzis, 2004), we currently know little about the statistics of environments relevant for human development (Frankenhuis, Nettle, et al., 2019). Ideally, there would exist a database of environmental statistics relevant to human development. Such a database would contain the values of different statistics for different environmental variables (e.g., rate of environmental fluctuations), different timescales (e.g., years, months), and for diverse populations. Empiricists can use the database to identify environments that show specific characteristics (e.g., highly unpredictable) to test existing hypotheses or explore new ones. Modelers can use it to set model parameters based on their corresponding empirical values (e.g., autocorrelations above 0.8). In this way a database of environmental statistics could increase synergies between empirical, developmental research and models of developmental processes.

1.8 Current objectives

This dissertation has two goals. The first is to provide a deeper understanding of the evolution and development of sensitive periods. Specifically, my models provide insights into conditions that favor sensitive periods halfway through ontogeny, as well as residual plasticity at the end of ontogeny. These models contribute to an integrative theoretical framework of the evolution and development of sensitive periods. They thus enrich existing theory of the evolution of phenotypic plasticity. The second goal is to increase synergies between models like mine and empirical work. To this end, my models incorporate a range of study paradigms for quantifying plasticity commonly used in empirical studies (e.g., adoption paradigms). In addition, I present a computational framework for studying environmental statistics in developmental science. My framework can facilitate the development of a database of environmental statistics, creating common ground for empiricists and theoreticians.

1.9 Thesis outline

Goal 1: evolutionary models of incremental development

First, I will present central tenets, insights, and predictions of existing models of sensitive periods, discuss how they relate to empirical work, and how future models may improve the bridge between theory and data (**Chapter 2**). Second, I will present two evolutionary models of incremental development that fill existing gaps in the literature. Both models extend previous work on sensitive period evolution by Panchanathan & Frankenhuis (2016). Their model explores the evolution of sensitive periods when development is incremental and irreversible. It also assumes that the cue reliability and the environmental state are constant within an organism's lifetime. I relax these assumptions. The first model presented here introduces variation in the reliability of cues across an

organism's lifespan and presents novel study paradigms to quantify plasticity (**Chapter 3**). The second model assumes a stable cue reliability but allows the environmental state to fluctuate across ontogeny (**Chapter 4**). This incremental approach to studying the evolution of sensitive periods has helped me to identify and understand the unique contributions of the environmental conditions explored.

Goal 2: a framework for studying environmental statistics in developmental science

At the core of developmental science is the interplay between individuals and their environments over time. All developmental research involves assumptions, claims, or questions about environmental stability and change. Statistical definitions serve as the building blocks we use to explore and test questions related to stability and change. Yet, we often overlook their importance, focusing instead on constructs and measurement instruments. As a result, statistical definitions are only loosely connected to the constructs they represent and are inconsistent across studies. This puts developmental studies in disarray. Focusing on stability and change, I present a computational framework that organizes environmental statistics across development (**Chapter 5**). The framework highlights different statistical definitions of stability and change and provides tools to realize them. Using environmental unpredictability as a case study, I apply the framework to a dataset of crime rates in New York City across 15 years. Computing environmental statistics in this way may be a start for developing a database of environmental statistics. Such a database could grow into a public platform that is shared across disciplines interested in how the environment shapes development. I conclude the dissertation with a general discussion of my findings and suggestions for future directions (**Chapter 6**).

1.10 Methods used in this dissertation

Both models presented here assume the following life-history of the organism: organisms are born, randomly disperse into a new patch (i.e., discrete area which they occupy), develop to maturity in the new patch, reproduce, and die. At birth organisms are uncertain about the state of their patch and only equipped with a prior estimate over the distribution of states. During each time period in ontogeny organisms sample a cost-free, imperfect cue to the environmental state of their patch to reduce uncertainty. Based on those cues organisms make phenotypic decisions.

Stochastic dynamic programming

I use stochastic dynamic programming (SDP) to compute optimal developmental trajectories. SDP is often used by behavioral ecologists to predict animal behavior in sequential, state-dependent decision problems (e.g., foraging) (Frankenhuis et al., 2013, 2018; Frankenhuis, Panchanathan, et al., 2019; Houston et al., 1988; Reimer et al., 2019). Dynamic optimization aims to find the choice or action that maximizes expected fitness (e.g., reproductive success) for every possible state of the developing organism. In my models, an organism's state consists of its current phenotype and cues sampled. The result of this optimization procedure is called an optimal policy. Stochastic dynamic programming is suitable for problems that involve uncertainty and interdependent decisions across time.

In the models presented here, organisms are uncertain about environmental conditions and current decisions affect future outcomes.

As is common in behavioral ecology, I use backward induction to solve the stochastic dynamic programming equations. Backwards induction determines all possible states an organism can be in at the end of development and computes fitness associated with each state. Working its way backwards in time, the algorithm then considers all possible states prior to the final state. For each penultimate state it then determines the decision that maximizes expected fitness in the final time period. The procedure is repeated until the first time period is reached. The outcome of this procedure is the optimal decision for each possible state an organism can be in across development.

Bayesian inference

Natural selection is an optimizing process, and so it might favor information use consistent with Bayesian learning – the optimal way of information updating (McNamara et al., 2006; Stamps & Frankenhuis, 2016). Bayes' theorem allows me to combine an innate prior estimate, such as the inherited estimate about being born into a dangerous world, with information received due to observing a cue (e.g., observing a street fight). Organisms then arrive at an updated estimate about the current environmental state. Observing a street fight, which is a more likely event in a dangerous than a safe environment, increases the organism's 'belief' about the current environment being dangerous. Organisms act as if they have a belief. In reality, however, organisms are adapted to a probability distribution over possible environmental states. A conscious mental representation of this distribution is not implied.

Simulating experimental adoption studies: quantifying plasticity

After computing optimal policies, I simulate whole populations following the optimal policies and the resulting distributions of mature phenotypes. To quantify plasticity, I use an experimental twin study. The following basic setup is common to both models that I developed. I simulate identical individuals (clones) following the optimal policy, who are separated at a certain time period during ontogeny. From then on, one individual (the focal) is kept in its original patch while the other individual (the clone) is adopted to a mirror patch. The clone receives reciprocal, opposite cues from the focal individual until the end of ontogeny. Plasticity is quantified by comparing the phenotypic difference between pairs of clones at the end of ontogeny. Large phenotypic differences at the end of ontogeny indicate that clones have been shaped to a large extent by experience, implying a high level of plasticity. This procedure is repeated for 10,000 pairs of clones for every possible time period during ontogeny.

Time series analysis

The last study (**Chapter 5**) borrows methods from time series analysis to compute stability and change statistics from repeated measures data (see Jebb et al. (2015) for a beginner-friendly tutorial). Any time series can be decomposed into its individual components: trend, season, and random component. The trend describes how the level of a time series changes with time. Season refers to the presence of reoccurring patterns within a calendar year. The random component is what is left of the data after subtracting

the trend and seasonal patterns. The presence of a trend or changes in variance across time indicate that a participant's time series is likely non-stationary; that is, its statistical properties change across time.

1.11 Transparency

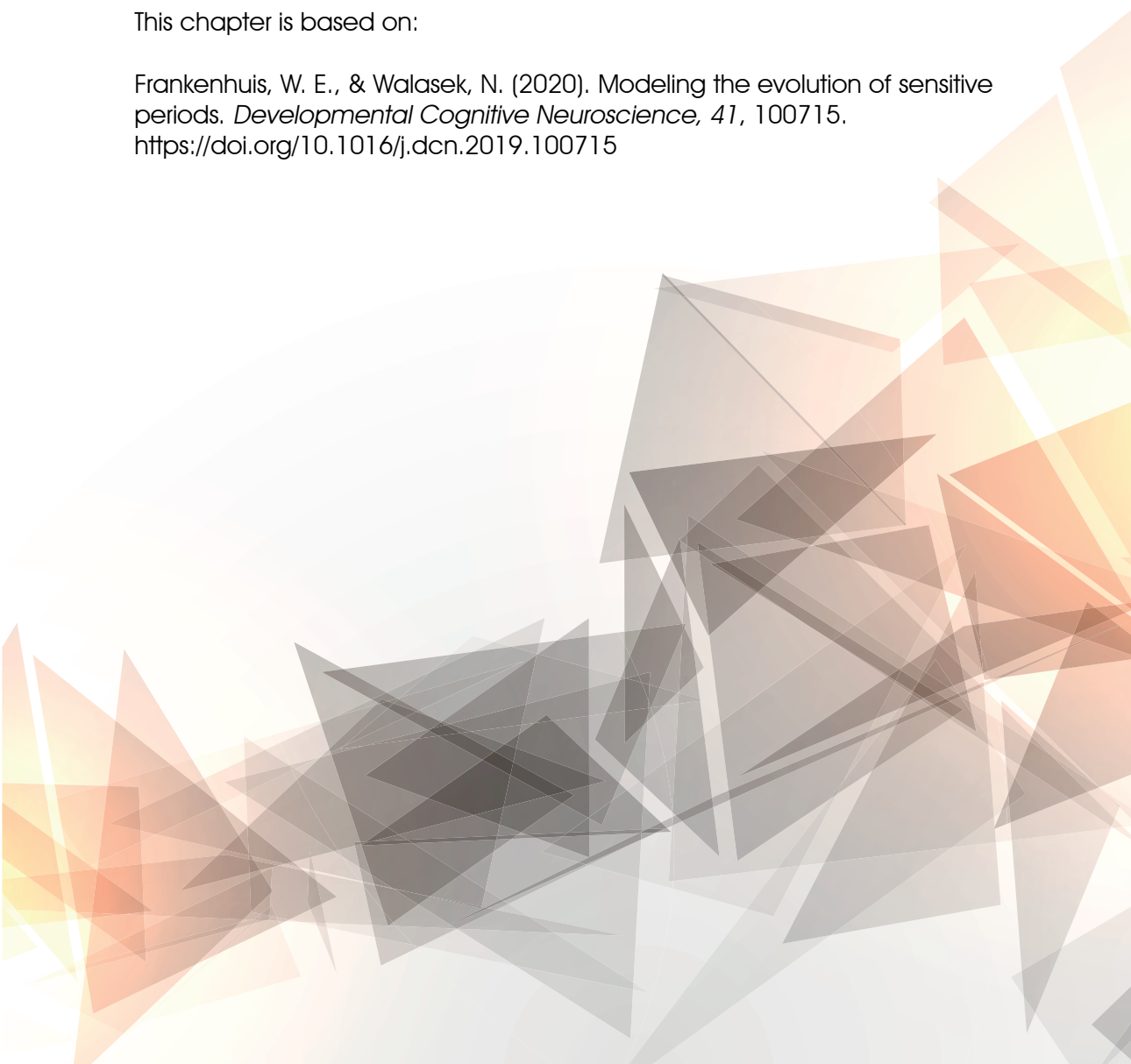
All chapters of this dissertation are based on articles that are either published in peer-reviewed journals or in preparation to being submitted to a journal. Chapters can be read independently from one another and in any order. In this dissertation, I focus on the development of theories and statistical definitions which is quite rare in a field as hyper-empirical as psychology. I hope that my work resonates with calls for more formal theory (Borsboom et al., 2021; Fried, 2020; Muthukrishna & Henrich, 2019; Smaldino, 2017), and more refined psychological constructs (Flake & Fried, 2020; Young et al., 2020). Increased integration of theory paired with open science research practices like preregistration of empirical studies, as well as the transparency of code and data, will hopefully increase the replicability of psychological studies in the years to come. All code and data that is part of this dissertation is openly available on GitHub. I provide links in each respective chapter. Additionally, I provide supplementary materials containing formulas and their derivations for each model.

Chapter 2

Modeling the Evolution of Sensitive Periods

This chapter is based on:

Frankenhuis, W. E., & Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience, 41*, 100715. https://doi.org/10.1016/j.dcn.2019.100715



2.0 Abstract

In the past decade, there has been monumental progress in our understanding of the neurobiological basis of sensitive periods. Little is known, however, about the evolution of sensitive periods. Recent studies have started to address this gap. Biologists have built mathematical models exploring the environmental conditions in which sensitive periods are likely to evolve. These models investigate how mechanisms of plasticity can respond optimally to experiences during an individual's lifetime. This paper discusses the central tenets, insights, and predictions of these models, in relation to empirical work on humans and other animals. We also discuss which future models are needed to improve the bridge between theory and data, advancing their synergy. We hope this work will contribute to recently emerging connections between the fields of developmental neuroscience and evolutionary biology.

2.1 Introduction

Finding common ground

Sensitive periods are widely studied across the social and biological sciences. The term is used in different ways in different disciplines. In this paper, we define a 'sensitive period' as a time period (or life stage) in which experience shapes a trait to a larger extent than the same experience does in other time periods (Fawcett & Frankenhuis, 2015). This definition is intentionally broad: the effects of experience are not necessarily limited to a sensitive period, and phenotypes developed during a sensitive period might be modifiable by later experience. Further, the definition is agnostic about mechanism; it concerns only the impact of experience on phenotype, not the mechanisms that implement this relation. Such generality has pros and cons. Pros are: this definition can be applied across species (from plants, to animals without brains, to homo sapiens) and to diverse phenomena (e.g., early programming, plasticity declining with age, adolescence offering new opportunities for adaptation). Cons are: when a general definition is applied to any particular species, it ignores mechanisms known to implement sensitive periods in this particular species (e.g., experience-*expectant* and experience-*dependent* plasticity in humans; Gabard-Durnam & McLaughlin, 2019; Galván, 2010; Greenough et al., 1987; Johnson, 2005).

Developmental neuroscientists often study sensitive periods using the framework of experience-*expectant* and experience-*dependent* plasticity. The former is plasticity that integrates "environmental information that is ubiquitous in the environment and common to all species members" (Greenough et al., 1987, p. 539); it involves neural mechanisms that come prepared for incorporating specific information (e.g., invariants in perceptual input). The latter, in contrast, is plasticity that integrates "environmental information that is idiosyncratic, or unique to the individual" (idem); it involves active formation of new synaptic connections in response to specific situations, which differ between individuals. This distinction captures an impressive array of processes in a variety of species, and has enabled tremendous progress in our understanding of the neurobiological mechanisms of plasticity. Moreover, subsequent elaborations of the distinction – which incorporate more refined descriptions of, for instance, the effects of timing and dose of experience (Dunn et al., 2019; Gabard-Durnam & McLaughlin, 2019) – are even better able to accommodate heterogeneity in contemporary data. We fully acknowledge the merit of this framework. Yet, we also agree with the scholars who originally developed this framework that it paints "a much more straightforward picture (…) than probably exists" (p. 551). In particular, the framework has limitations when applied across the full tree of life.

The framework does not capture all classes of plasticity particularly well. For instance, polyphenic traits are traits where multiple, discrete phenotypes emerge from a single genotype, depending on environmental conditions. In many reptiles, variations in nest temperature determine sex (whether an organism becomes male or female). In many insects, temperature, photoperiod, or nutrition determines the caste of an individual (e.g., a larvae can develop into a queen, worker, or soldier). In many crustaceans, exposure to chemicals released by predators induces the development of defensive armor (Gilbert, 2003). Such polyphenisms respond to environmental information that is not common to all species members, nor unique to an individual. Rather, all members of a species have

evolved to 'expect' different states of the environment, and are 'prepared' to develop a range of different phenotypes, depending on environmental or somatic conditions. The same logic applies to developmental mechanisms that are evolutionarily prepared to produce continuous phenotypic variation (e.g., adaptive calibration of rate of maturation in response to nutritional condition).

Polyphenic traits have some characteristics of experience-*expectant* plasticity, some of experience-*dependent* plasticity, and some that fit neither class well. All species members 'expect' particular experiences and are 'prepared' to respond to them. However, these experiences differ between individuals, frequently in non-idiosyncratic ways (e.g., in each generation, a predictable proportion of the population is exposed to each type of experience). Moreover, in some cases, the impact of experience is restricted to a single time window; in others, there are several time windows. In some cases, these windows are well delineated; in others, their onset and offset are more gradual. In some cases, time windows are neatly ordered; in others, their ordering is more variable. In some species, the effects of experience are irreversible; in others, they can be reversed (e.g., some fish can switch sex multiple times, including sex-specific behavioral repertoires, depending on social hierarchy), albeit perhaps more easily in some life stages than others. In terms of mechanism, some of these responses depend on neural overproduction and pruning, but many do not (e.g., the development of armor in crustaceans).

Nature rarely comes only in two kinds. More often, it presents a smorgasbord. In such cases, dichotomous frameworks can be extremely powerful, but not necessarily for all purposes. Here, we need a broad definition that describes changes in plasticity across ontogeny in a wide range of species. In addition, this definition should exclude cases where plasticity does not change across ontogeny. As noted earlier that our definition also has limitations; it is agnostic about mechanism. When studying species that fit experience-*expectant* and experience-*dependent* plasticity, researchers may prefer to use those terms.

Why do sensitive periods exist?

Why are organisms not *Darwinian demons*, capable of adjusting their phenotypes with equal ease to new conditions throughout their lifetimes (Law, 1979)? In reality there is variation in plasticity (i.e., the ability to tailor development based on experience) between different species, between individuals, and even between different brain systems within a single individual. For instance, some bird species are only able to learn new songs in their first weeks, while others retain this ability throughout their lives (Beecher & Brenowitz, 2005). After being adopted, some children adjust better than others to the new conditions (Ellis et al., 2011). And, different brain systems within a person may adjust to new conditions at different rates (Zeanah et al., 2011).

In the past decade, there has been formidable progress in our understanding of the mechanisms that determine changes in plasticity over the life course (Takesian & Hensch, 2013). It is now possible to modify aspects of sensitive periods (such as their timing and duration) for a range of neurobiological systems in different species, through experiential or pharmacological manipulation. This research truly has applied potential. For instance, it can inform interventions that erase neural signatures of trauma (Hensch & Bilimoria, 2012). It

also raises important ethical questions; for instance, whether it is ever ethical to apply such techniques to humans, and if so at what age and on which grounds. Despite great progress in our understanding of *how* sensitive periods work, we know little about *why* sensitive periods exist. Specifically, we know little about the conditions that favor the evolution of sensitive periods, about which sensitive periods are adaptive and which ones are not; and if adaptive, what a sensitive period's function is. There are plausible hypotheses about specific observations, but there is no unifying theory.

Fortunately, a unifying framework is starting to emerge in biology. In the past decade, a set of formal (i.e., mathematical) models has emerged exploring the evolution of sensitive periods. These models explore how mechanisms of plasticity can respond in fitness-enhancing ways to experiences during an individual's lifetime (see Section 2.2 'Evolutionary modeling of sensitive periods'); and, as a consequence, produce variation in plasticity between species, individuals, and systems within the brain. Biologists acknowledge that not all plastic responses, nor all variations in plasticity, are adaptive. Yet, they explore 'optimal' developmental responses in different conditions. The reason is epistemological: in order to know which variation is adaptive, we need theory predicting what animals ought to do 'if' they are responding adaptively (Frankenhuis et al., 2013, 2018). If predictions match observations, we find some support for our model and gain new insight. If predictions and observations do not match, we should modify our model, for instance, by incorporating constraints on animals' abilities to sample and use available information in their environment in an optimal way (Marcus, 2009; Todd & Gigerenzer, 2000).This optimality approach is already used widely in different subfields of biology (e.g., functional morphology, behavioral ecology) and in cognitive science as well (see Section 2.2, 'The value of modeling').

We do not discuss formal models in detail (see Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017). Rather, we will describe their central tenets, insights they provide, their predictions, and a selective review of empirical research on humans and other animals. In addition, we discuss which future models are needed to improve the bridge between theory and data, advancing their synergy.

2.2 Evolutionary models of sensitive periods

The value of modeling

Formal models have several advantages over natural language. Natural language is often more ambiguous than mathematics, and inferences – for instance, from premises to predictions – based on human reasoning are more fallible (e.g., subject to confirmation bias) than a mathematical analysis. Almost 200 years ago, Darwin recognized the value of mathematics. He wrote in his autobiography: "I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense" (1828–1831). As with natural language theories, the utility of any particular model will depend on the validity of its assumptions (do these match the phenomenon of interest), the rigor of its analysis (exploring all of the relevant conditions), and the interpretation of results (drawing reasonable conclusions about the world).

Models, like maps, focus our attention on some factors and processes, while leaving out others (Epstein, 2008; Smaldino, 2017). In cognitive neuroscience, models focus on neurocognitive processes and their outcomes (van den Bos & Eppinger, 2016). In evolutionary biology, models focus on evolutionary processes (e.g., natural selection, mutation) and their outcomes. Models in cognitive neuroscience are often fitted to data (such as decisions, reaction times, and brain activity) with the goal to better understand the processes generating these data. By contrast, evolutionary models often start with general axioms (e.g., natural selection favors adaptive mechanisms), make additional assumptions about a phenomenon (e.g., plastic mechanisms tailor phenotypes to local conditions), and explore which mechanisms are favored, depending on environmental conditions (Frankenhuis et al., 2013; Frankenhuis, Panchanathan, et al., 2019).

Cognitive neuroscientists use formal models to explore questions about causal mechanisms at a *proximate* level. Evolutionary biologists use models to explore questions about evolutionary selection pressures at an *ultimate* level. For instance: if environmental conditions fluctuate at a particular rate, should natural selection favor plasticity or not? If plasticity is favored, should its level be uniform or variable across ontogeny? If variable, should we expect the onset and offset of enhanced plasticity to be punctuated or gradual? Is there one peak or multiple ones? Although evolutionary models do not directly provide insight into neurobiological mechanisms, they do offer hypotheses about the factors and processes that influence levels of plasticity (see Section 2.3). However, prediction is not the only goal of models. Models also help to organize existing observations, for instance, by explaining the adaptive function (or lack thereof) of known mechanisms. Proximate and ultimate explanations are not in opposition to each other, but mutually compatible; they exist at different levels. A biologist who has achieved a complete understanding of the neurobiological mechanisms of plasticity in, say, a soapberry bug, might still wonder: 'What evolutionary selection pressures have favored plasticity in this species?'

A study of soapberry bugs – a half-inch-long, seed-eating insect – illustrates how plasticity can evolve. In Oklahoma, harsh weather conditions (e.g., storms) randomly kill subsets of individuals, and so soapberry bugs are exposed to a variable sex ratio (the ratio of females to males). There, males have evolved the ability to adjust their levels of mate guarding. When there are many rivals, they guard their current mate. When there are few rivals, they search for new mates. In Florida, by contrast, the weather is not harsh and so the sex ratio is stable over time. There, males have not evolved plasticity in their level of mate guarding; that is, males in Florida always guard the same amount. If these males are artificially exposed to varying sex ratios in the lab, they do not adjust their levels of mate guarding (Carroll & Corneli, 1995). This study shows that plasticity is a target of natural selection. Although this study did not examine whether there are sensitive periods in the development of mate guarding, formal models could be developed that explore at what life stages soapberry bugs should sample the local sex ratio, for how long they should sample, and how observations of the local sex ratio should affect their behavior.

Evolutionary models thus help to explain, at an ultimate level, 'why' different mechanisms have evolved in different species and traits; but not, at a proximate level, 'how' these mechanisms work. These models do, however, offer predictions about 'how'

animals should respond to experiences at a behavioral level. Such predictions offer insight into proximate-level, developmental processes, even when these predictions do not tell us 'how' behavior is accomplished (e.g., prolong plasticity if experience is too noisy to infer the statistical structure of the environment). Until biologists and neuroscientists achieve a complete understanding of mechanisms, therefore, evolutionary models can do more than just explain variation between species and traits; they can also help to uncover the factors and processes that influence the onset, duration, and offset of plasticity across ontogeny.

Evolutionary modeling of sensitive periods

All phenotypes, even those shared among all members of a species, result from developmental processes. It follows that natural selection can only influence phenotypes by shaping developmental systems; that is, the array of causal factors and processes that construct phenotypes (Barrett, 2014; Frankenhuis et al., 2013). The modeler wants to understand what evolutionary pressures, across generations, result in mechanisms that produce sensitive periods, within generations, based on experience.

Evolutionary models of sensitive periods do not start out assuming a sensitive period. Rather, such a period might emerge as the outcome favored by natural selection; that is, the outcome that maximizes biological 'fitness'. Psychologists often use the term fitness to denote individual survival and reproduction. Biologists, however, typically use the term to refer to the reproductive success of developmental systems (or mechanisms, strategies, genotypes). These systems generate distributions of phenotypes (individuals), which might pass on the developmental system by reproducing. Thus, offspring inherit developmental systems from their parents. The fitness of a developmental system, then, depends on the extent to which the individuals it generates produce more offspring than those produced by other systems. From this viewpoint, the adaptive value of a sensitive period depends not on whether any particular individual benefits from it. What matters, rather, is whether the system achieves high fitness relative to other systems, *because* it generates phenotypes that are more affected by experience at certain times of life than others (e.g., early in ontogeny). Section 2.4, 'Bridging evolutionary modeling and empirical paradigms' discusses different ways to quantify the impact of experience on phenotype.

All developmental systems include both genotypic and environmental factors and processes. However, the roles of these components in the production of phenotypes differ between developmental systems (Barrett, 2014; Bjorklund & Ellis, 2014; Frankenhuis et al., 2013; Gottlieb, 1991; Lickliter & Honeycutt, 2003; Tooby et al., 2003). Some systems use aspects of their environments that are shared among all species members (e.g., invariants in the visual environment used to construct perceptual abilities). Others use aspects of their environments that vary between species members (e.g., polyphenisms). Both types of systems are common in nature, and both may exhibit sensitive periods. However, all evolutionary models of sensitive periods (that we are aware of) have explored systems that are exposed to environmental variation between generations, within generations between individuals, or both. There is, therefore, clearly a need for models of the evolution of sensitive periods in neural systems that are adapted to environmental invariants. Here, we restrict our discussion to existing models. The question addressed by these models is

whether experience should differently affect development, depending on its timing, dose, the information it provides, and so on.

To be able to explore these kinds of effects, a model needs to include two or more time periods in which organisms are able to access 'cues' that can shape their phenotypes (Frankenhuis, Panchanathan, et al., 2019). A cue is an observation that provides information (i.e., reduces uncertainty), either about the environment (e.g., safe or dangerous) or about the organism itself (e.g., its somatic condition). Cues are often imperfect (e.g., there may be smoke but no fire). The reliability of a cue depends on the extent to which it discriminates between different states of the environment or states of the organism. A cue has high reliability if it is much more likely to occur in certain states of the world than others (e.g., violence is more likely to occur in poor than in rich neighborhoods). A cue has low reliability if it is almost equally likely to occur in different states of the world (e.g., seeing a person lock their house may be about equally likely in poor and rich neighborhoods). Many species use cues to infer their current conditions (e.g., the level of danger), and some species also use cues to predict their (likely) future conditions. As we discuss below, the reliability of cues might affect the optimal level of plasticity, because this optimal level might depend on the extent to which an organism 'knows' (has information about) what the current conditions are and how likely these conditions are to change or remain the same.

Evolutionary models usually conceptualize development as a sequential decision-making process. These models often describe the 'state' of an organism, which determines the decisions it makes, in terms of two components: estimates of the environment (e.g., safe or dangerous) and phenotypic condition (e.g., nutritional reserves). The model then computes, for every possible state of an organism, which decision maximizes the fitness of its mechanisms. We distinguish between 'mechanism' (or strategies, genotypes) and 'organism' here too because, as noted, decisions that are optimal for a mechanism might produce outcomes that are actually detrimental for a subset of individuals (Frankenhuis & Del Giudice, 2012). For instance, in winner-takes-all mating systems (e.g., the alpha has many more babies than other group members, as in elephant seals), it may be adaptive for developmental mechanisms to produce aggressive animals that vie for the top rank in the social hierarchy. Such mechanisms may have higher fitness than alternative mechanisms that produce less aggressive individuals. Fighting is stressful and some animals will die. Yet, fighters may be maximizing the fitness of the mechanisms that created them.

In cognitive modeling, a prior is the estimate of an individual at the beginning of a decision problem (e.g., whether or not to wait for a reward). This estimate usually has its source within the individual's lifetime (Dunlap & Stephens, 2016). It is based on personal experience (e.g., past promises were broken) or learned socially (e.g., people say future rewards are unreliable). In evolutionary models, in contrast, the prior does not necessarily represent psychological knowledge. Rather, it is an adaptation of a developmental system to the distribution of environments experienced by a lineage over evolutionary time. For instance, if a species was consistently exposed to high levels of predation, it may embody this statistical regularity by building anti-predator defenses by default, unless it receives strong evidence (contradicting the prior) that the current environment is actually safe. In evolutionary models, organisms may inherit their priors from their distant ancestors (e.g.,

via genes), from their immediate ancestors (e.g., via parental effects, epigenetic factors), or a combination of both (Dall et al., 2005, 2015; Mangel, 1990; McNamara et al., 2006; Pfab et al., 2016; Stamps & Frankenhuis, 2016; Stamps & Krishnan, 2014a, 2014b; Trimmer et al., 2011; Uller et al., 2015). Organisms update their priors based on the cues they sample during their lifetimes – often in a Bayesian fashion, the optimal way of updating – while making decisions that affect their phenotypes. These decisions and their phenotypic consequences illuminate the evolution of sensitive periods.

2.3 Plasticity depends on information about the current environment

Evolutionary models of sensitive periods have produced a variety of insights. We do not provide an exhaustive discussion of these insights here. Rather, we focus on one insight that is particularly relevant to developmental cognitive neuroscience. Readers who wish to read more may consider the following resources (Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017; Stamps & Krishnan, 2017).

Plasticity often depends on the extent to which the prior and cue reliability, that an animal is adapted to, provide information (reduce uncertainty) about the current state of the environment. Animals that are adapted to more uncertainty about current or future conditions – for instance, because their lineage evolved in diverse environments, or because cues have low reliability – might benefit from having greater plasticity. In many cases, the amount of information an animal is adapted to increases over its lifetime, because the animal learns about its environment; as a consequence, its plasticity might decline. This finding emerges in many models, and within models across a broad range of parameter combinations.

This finding fits with empirical research showing that sensitive periods in *experience-expectant plasticity* might be prolonged if: (i) animals are deprived of cues (Hensch, 2004; Knudsen, 2004; Michel & Tyler, 2005); (ii) animals process noisy cues (Chang & Merzenich, 2003); (iii) cues are gradually changing (Bateson & Martin, 1999; Bolhuis, 1991); or (iv) perceptual systems offer the brain unstable inputs, possibly because they are developing or are disrupted (Thomas & Johnson, 2008). For instance, in zebra finches, the absence of tutors extends the sensitive period for song learning (Kelly et al., 2018), with greater numbers of new neurons being added to the high vocal center (an avian brain region) (Wilbrecht, 2006). The lack of exposure to faces prevents perceptual narrowing in Japanese macaques (Sugita, 2008). Exposing rat pups to a stream of white noise delays their auditory specialization (Chang & Merzenich, 2003). However, deprivation does not *merely* prolong sensitive periods. In humans, for instance, it can accelerate synaptic pruning and limit myelination, reducing cortical thickness and white matter integrity (McLaughlin et al., 2017); and related, in non-human primates, it can result in neural disuse and inefficient processing (Scott et al., 2007). In general, the effects of adverse experience on the brain are complex and diverse, because they result from a multitude of processes (Gabard-Durnam & McLaughlin, 2019; Galván, 2010). We argue that one such process is the rate at which the brain is able to infer the statistical structure of the environment.

The common denominator is that the animal lacks access to reliable cues about its environment. As Bateson and Martin (1999) noted: "processes that bring the sensitive period to an end are related to the gathering of crucial information and, except in extreme cases, do not shut down until that information has been gathered" (p. 162). However, as we discussed in Section 2, the fit an animal achieves with its environment depends not only on the cues it collects, but also on the distribution of environments its lineage has adapted to (its evolved prior). The extent to which a given cue shifts this prior depends on its variance as well as the extent of agreement between prior and cue (Stamps & Frankenhuis, 2016). If a prior has more variance (more uncertainty about the state of the environment), a given cue shifts an estimate more than when it has less variance. And, the more a cue disagrees with the prior, the more the prior shifts.

Evolutionary modeling accordingly predicts, all else being equal, that if animals who have different priors are exposed to the same cue, those whose priors and cue are in agreement change their phenotypes less than those whose prior and cue disagree (Stamps & Frankenhuis, 2016). Biologists have recently tested this novel prediction in fruit flies (Stamps et al., 2018). They first showed that larvae vary in the extent to which they are attracted to the odor of ethyl acetate (a fruity smelling liquid); some flies had positive priors about ethyl acetate, others had negative priors. Then they showed that flies that had positive priors changed their behaviors more after an aversive training regime (experimental exposure to a negative cue) than flies with negative priors. So, the extent of phenotypic change depended on the convergence between prior and cue.

Similarly, individual variation in the duration of sensitive periods in humans may also depend on the agreement between priors and cues. If two individuals who have different priors are exposed to the same cues, the individual whose prior and cues agree more might lose their plasticity earlier. As priors are inherited from parents (e.g., via the genome or epigenome), we might expect children whose environment matches that of their parents to have shorter sensitive periods than children who develop in a different environment than their parents. Similarly, we might expect individuals who have more consistent experiences (e.g., all safe cues versus some safe cues and some danger cues), or more reliable cues (e.g., extreme experiences that occur only in extreme conditions) to reduce their uncertainty faster, and hence lose their plasticity earlier, than individuals who have less consistent experiences or who sample less reliable cues (Frankenhuis & Panchanathan, 2011a, 2011b; Panchanathan & Frankenhuis, 2016). To our knowledge, these hypotheses have not been tested in humans. The results of such tests would be of great interest to researchers in the field of 'differential susceptibility,' who study the developmental emergence of individual differences in plasticity (Ellis et al., 2011).

2.4 Sensitive periods in adolescence and other life stages

Evolutionary models have shed light on the conditions favoring sensitive periods early in life. However, there may be sensitive periods in other life stages as well, such as middle childhood (Del Giudice, 2014; Del Giudice & Belsky, 2010) and adolescence (Blakemore & Mills, 2014; Dahl, 2004; DePasquale et al., 2018; Fuhrmann et al., 2015; Sachser et al.,

2018). Some of these sensitive periods might be adaptive. Few formal models, however, have explored the evolution of 'mid-ontogeny' sensitive periods. In this section, we discuss initial steps towards such models.

Adaptive reasons for sensitive periods

In a theoretical paper, Fawcett and Frankenhuis (2015) proposed that sensitive periods evolve when there is variation across ontogeny in (a) the availability of cues, (b) the informativeness of cues, (c) the fitness benefit of information, and (d) the fitness cost of plasticity. First we briefly describe each of these arguments. Then we discuss a model of mid-ontogeny sensitive periods, which we recently developed (Walasek et al., 2021).

(a) A cue might only be available in some life stages and not in others, or be more likely to occur in some life stages than others. For instance, Kuzawa (2005) hypothesized that pregnant women transmit physiological signals to their fetus, which provide a summary of her lifetime nutritional experience, and which the fetus uses to predict its own postnatal nutritional environment. As this putative cue is only present inside the womb, people may only be sensitive to this cue during the fetal life stage. In a similar way, sensitivity to other cues may be limited to later life stages. For instance, courtship cues – e.g., being approached with sexual intent – are extremely rare early in life and increase in frequency closer to puberty. Therefore, people's sensitivity to these cues, and their use of such cues in guiding their reproductive strategies, might increase over the course of childhood. People apparently use the quality of courtship cues to estimate their own desirability as a mate – unromantically referred to as 'mate value' in the biological sciences – and then use this estimate to determine what attributes they expect in future mates (e.g., which value such a mate should have, which level of commitment to the relationship, which level of investment in shared offspring) (Conroy-Beam et al., 2016).

A cue might also become available, or increase in frequency, at the life stage in which animals first have to navigate the environment on their own, independently of their parents. A test case exists when individuals differ in the timing of this species-typical developmental milestone. For instance, consistent with the stress acceleration hypothesis, rodent pups that receive low levels of maternal care leave the nest at a younger age, and such fledging is accompanied by accelerated development of emotion circuits that enable learning about dangers that become relevant after fledging (Bath et al., 2016; Callaghan & Tottenham, 2016; Gee et al., 2013; Sullivan & Holman, 2010; for research showing parental modulation of learning in humans, see Tottenham et al., 2019; for a formal model exploring how a person's attachment style in adulthood may be shaped by relationships early in life, see Chumbley & Steinhoff, 2019). In this case, individual variation in leaving the nest, a developmental milestone, is associated with individual variation in increased levels of plasticity. Applying this idea to human development, we may speculate that people experience a temporary increase in plasticity when they move from one environment to another (e.g., moving to a new school, or into a new neighborhood). We may also speculate that an increase in prediction error could be a mediating mechanism. The old environment was predictable, reducing the need for plasticity. After moving the individual might benefit from elevated levels of plasticity to learn the statistical structure of the new environment. Thus, by hypothesis, changes in the availability and frequency of cues – such as those occurring with

a new set of experiences – might increase plasticity in cognitive functions that need to be adapted to aspects of the environment that are likely to have changed.

(b) Even if a cue is present throughout ontogeny, it may be more reliable in some life stages than others (Fawcett & Frankenhuis, 2015). For instance, in addition to courtship cues, people might use non-sexual social cues – such as receiving positive social attention – to estimate their mate value. Such social cues are present from birth, but their reliabilities as indicators of mate value might increase from infancy to adolescence; that is, social attention received by an infant (based on their 'cuteness') presumably conveys less information about mate value at puberty than social attention received by a prepubescent teen. A cue's reliability sets an upper bound to how much can be learned from a cue. However, the amount of information an individual extracts from a cue also depends on her perceptual and cognitive abilities. This amount can increase over ontogeny if perceptual systems become more accurate as they mature, or if understanding a cue depends on acquired knowledge (e.g., a child may not understand a subtle form of social rejection used by adults). In such cases, mechanisms using the cue may increase their sensitivity to the cue over the course of ontogeny.

(c) Even if a cue provides the same amount of information throughout ontogeny, its potential to affect fitness might be higher at some life stages than others (Fawcett & Frankenhuis, 2015). For instance, an individual who receives a cue indicative of her mate value (e.g., courtship cues) will have more to gain from using this cue around puberty – when her future reproductive potential is high – than following menopause. Therefore, we may expect individuals to be more sensitive to such cues during adolescence compared with old age. In sum: adolescents might be particularly sensitive to social feedback for (at least) three different reasons: such feedback might be more available, more reliable, and have more scope to affect fitness.

(d) The costs of plasticity, like its benefits, might vary across ontogeny. These costs may include the energy invested in building, maintaining, and running the neural systems to perceive and use cues (Auld et al., 2010; DeWitt et al., 1998; Relyea, 2002). Although it has been challenging to document costs of plasticity empirically, there are convincing examples. For instance, fruit flies bred for enhanced (associative) learning ability evolve shorter life spans as a consequence. Their investment in plasticity thus trades off with somatic maintenance (Mery & Kawecki, 2003, 2004, 2005). When fruit flies are energetically starved (experimentally), their brain shuts down the formation of aversive long-term memories, which are costly to produce (Plaçais & Preat, 2013). Re-feeding starved flies, however, facilitates memory formation, showing that plasticity can be regained (Hirano et al., 2013). This flexibility suggests that plasticity (the ability to adjust development based on experience) does not change across ontogeny. Hence this example does not qualify as a sensitive period, according to our definition. However, the example does illustrate that plasticity may trade off with other energetically expensive activities. Accordingly, if all members of a species are low on resources at a particular life stage (e.g., salmon after having swum upstream to reproduce and die in their natal patch), we may speculate that natural selection favors a species-typical decline in plasticity at this life stage. The idea that plasticity is costly might initially seem at odds with the empirical finding that putting breaks

on plasticity (e.g., perineuronal nets) is metabolically costly (Werker & Hensch, 2015), but it is not. If costly molecular mechanisms exist in order to regulate plasticity, they are a cost of plasticity; non-plastic organisms would not need such mechanisms.

Bridging evolutionary modeling and empirical paradigms

Formal models to date have assumed that cues are equally reliable in all time periods. We have recently developed a model in which the cue reliability varies across ontogeny (Walasek et al., 2021). In our model, individuals sample cues to the current conditions, while gradually – step-by-step, in each time period – tailoring their phenotypes to these conditions. We vary the cue reliability in three ways: cues may become more reliable over ontogeny (increasing), less reliable (decreasing), or first more reliable and then less reliable (triangular). To find out whether natural selection favors sensitive periods in mid-ontogeny, we evolve (optimal) developmental strategies in different environments (combinations of priors and cue reliability patterns). Then we expose organisms following these strategies to experiences (cues) in order to observe optimal decisions and resulting developmental trajectories.

We use 'study paradigms' that resemble those used in empirical research on sensitive periods to uncover plasticity. Studies of humans typically compare people who have been adopted (or have migrated) at different ages with each other, and with people who have not been adopted (or not migrated) (Mascie-Taylor & Little, 2004; Pallier, 2003; Zeanah et al., 2011). Studies of non-human animals (e.g., rodents, birds) often use controlled experimental setups, such as *cross-fostering paradigms* in which animals are experimentally transferred at different times between different caregivers or patches (Breed & Moore, 2015), or *dose-dependent experience paradigms* that systematically vary the duration and amount of exposure to particular experiences (Groothuis & Taborsky, 2015).

We have explored similar types of paradigms in order to foster a bridge between theoretical and empirical studies of sensitive periods. We instantiate these paradigms by creating identical twins (clones) that are separated at different times during ontogeny and then exposed to different experiences. Next we measure the resulting differences in their phenotypes. If these differences are small, the developmental system had little plasticity at the time of separation; if it is large, it had much plasticity. We also vary the timing of separation. That is, we create and separate twins in each developmental time period. If twins separated early in life diverge more than twins separated later, there is a sensitive period early in life. If twins diverge most when separated mid-ontogeny, plasticity is highest in mid-ontogeny.

We also vary 'how' experiences differ between the twins during their separation (Figure 2.1). Extreme divergence in experience occurs with *yoked, opposite cues*; here, if one twin samples a danger cue, the other samples a safe cue. This treatment is artificial; in empirical studies, it only occurs in controlled lab conditions. However, a milder form of divergence, *opposite patch cues*, more closely matches a situation in which twins are separated (through adoption or migration) into different conditions. Third, we explore *deprivation*, receiving cues that are too noisy to extract information from, which corresponds to sensory deprivation in lab conditions (Hubel & Wiesel, 1970) or exceptionally traumatizing

real-world circumstances. For instance, children might have spent extended time in a dark enclosed space while in hiding during wartime (Wolf, 2007) or have grown up in very deprived orphanages (Kaler & Freeman, 1994).

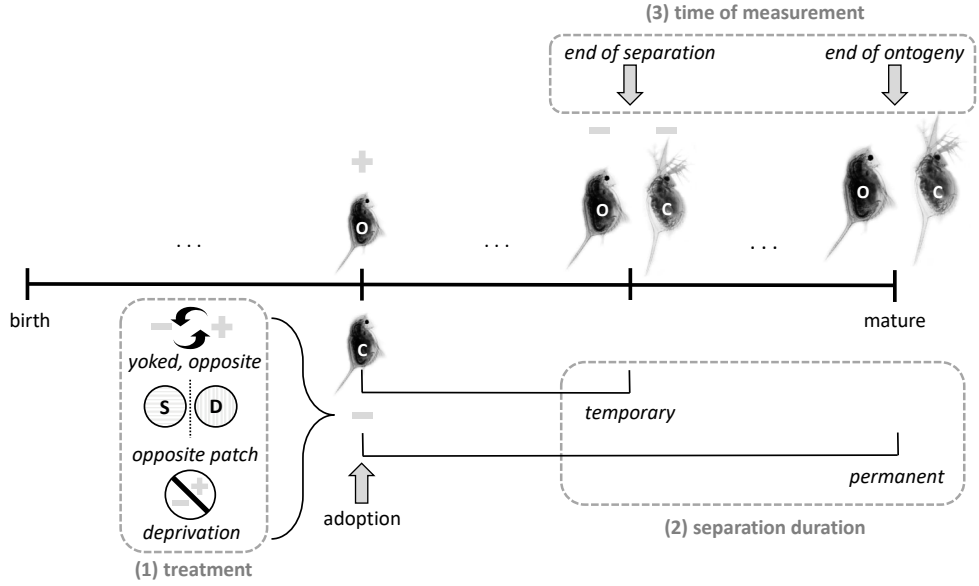


Figure 2.1 Measuring changes in plasticity across ontogeny. We separate twins (original, denoted as O, and clone, denoted as C) at different ages. We vary three dimensions: treatment, separation duration, and time of measurement. (1) *Treatment* refers to how the experiences of the original and clone differ during their separation. The clone might experience yoked, opposite cues; cues from the opposite patch; or deprivation. With yoked opposite cues, the clone always samples the opposite cue of the original: if the original samples a minus cue [-], the clone samples a plus cue [+]. With cues from the opposite patch, the clone samples a sequence of cues typical of the opposite patch: if the original tends to sample more minus cues, the clone tends to sample more plus cues. In our Figure, the original and the clone are both in the dangerous patch (denoted as D), but the clone receives cues typical of the safe patch (denoted as S). With deprivation, the clone receives cues that are too noisy to extract information; thus preventing learning about the environment. (2) *Separation duration* refers to whether the separation of twins is permanent or temporary. Permanent separation occurs if twins experience different conditions from their separation until the end of ontogeny (maturity). Temporary separation occurs if twins are reunited before the end of ontogeny. (3) *Time of measurement* refers to when differences in the phenotypes of twins are measured. We measure differences in phenotypes of twins at two different time points: at the end of their separation and at the end of ontogeny. Our results show that different treatments tend to produce (qualitatively) similar patterns of plasticity. Our predictions are therefore similar for different treatments and for different measurement times used in empirical research. Copyright: we have used the images of Daphnia with permission from Dr. Linda Weiss (2019).

We also compare 'permanent' versus 'temporary' separation. Permanent separation occurs when individuals experience different conditions from their separation until the end of ontogeny (maturity). In the real world, permanent separation might occur if one child is adopted and another is not, or if children are adopted into different homes. In lab conditions, permanent separation occurs in cross-fostering studies. Temporary separation

might occur when siblings are separated (e.g., during a war or a natural disaster), and are later reunited within one home. In the lab, temporary separation occurs in dose-dependent experience studies, in which the experiences of individuals differ to a specific degree at a particular time. Finally, we measure differences in phenotypes of twins at two different time points: at the end of their separation and at the end of ontogeny.

In sum: we vary the timing of separation (age), the extent to which experiences differ (yoked, opposite cues; opposite patch; and deprivation), and whether separation is permanent or temporary. We then measure differences in phenotypes both at the end of their separation and at the end of ontogeny. Jointly, these treatments cover many paradigms used in empirical research.

Our results show that natural selection favors sensitive periods in mid-ontogeny in two conditions: if cues become more reliable over time, or if cues first become more reliable and then less reliable. If cues start out more reliable and become less reliable, sensitive periods are never favored. These results are strikingly general across prior probabilities (under the assumption that cues are highly reliable in at least some time periods). Sensitive periods also look remarkably similar for increasing and triangular cue reliabilities; perhaps in both cases organisms already tend to have good estimates of the state of the environment in mid-ontogeny (when cue reliabilities start declining in the triangular case and keep going up in the increasing case). Previous work has shown that if the cues are equally reliable and the environmental state is stable across ontogeny, sensitive periods evolve early in life (see Section 2.3). Integrating across models, we conclude that if cue reliabilities are either constant or decrease over ontogeny, sensitive periods evolve early in life; but if they increase over ontogeny (or are triangular), sensitive periods evolve mid-ontogeny. Our results depend on the study paradigm, but only as a matter of degree, not kind. This is good news: it means that the predictions of our model should hold across different treatments and times of measurement in empirical research.

'Belief-and-phenotype' and 'belief-only' models

In some models, optimal decisions depend both on an organism's phenotypic state (e.g., traits already developed) and on the information available to an organism about its environment (i.e., as a function of its evolved prior and the cues it has sampled during its lifetime). Note: such estimates are often referred to as "beliefs", even though conscious deliberation, or even psychological representation, is not necessarily involved. In these models, organisms that have identical beliefs might make different decisions because their phenotypes differ. Our model is of this kind. We call this a 'belief-and-phenotype' model. Other models assume a one-to-one mapping between beliefs and phenotypes. In these models, organisms do not have phenotypes, only beliefs. Individuals with the same beliefs thus always make the same decisions. We call such models 'belief-only' models.

Table 2.1 presents for a set of models of the evolution of sensitive periods whether each model includes only beliefs or beliefs and phenotypes, how plasticity is quantified, and information about the study paradigm. Some models have used an explicit paradigm (like the ones we just discussed) to expose changes in plasticity over time; others have not and require readers to infer the degree of plasticity. To compare findings and predictions across

different models, it would be helpful if, when possible, researchers would use the same paradigms or explicitly describe their own paradigm.

Model	Phenotype (P) and/or Belief (B)	How is plasticity measured?	Study paradigm (if applicable)
Frankenhuis & Panchanathan, 2011	P&B	Number of cues sampled	
Fischer et al., 2014	P&B	Phenotypic adjustment after each time period in response to a sampled cue, current phenotype, and current belief about the environmental state	Phenotypic adjustment is measured after each cue; a range of cue reliabilities is explored
Stamps & Krishnan, 2014a	B	Difference in beliefs after repeated exposure to the same cue	Each individual is exposed to the same cue four times; differences in beliefs are measured after each time period
Stamps & Krishnan, 2014b	B	Within-individual design: absolute difference in beliefs before and after exposure to a cue at different ages Sequential design: one individual is exposed to two different cues for an extended period of time	Difference in beliefs is measured after each cue Three cue reliabilities: high, low, and moderate levels of danger
English et al., 2016	P&B	Within-individual effects of temporary food supplementation or deprivation during different time periods on phenotypes (age and size at maturity, reproductive success)	Extreme divergence (supplementation or deprivation); temporary treatment; plasticity is measured at the end of ontogeny
McNamara et al., 2016	P&B	Phenotypic variance of a genotype is attributed to different sources of cues	
Panchanathan & Frankenhuis, 2016	P&B	Phenotypic divergence between simulated twins as a function of separation time	Extreme divergence between experiences (yoked opposite cues); permanent separation; plasticity is measured at the end of ontogeny
Stamps & Krishnan, 2017	B	Within-individual design: absolute difference between beliefs before and after exposure to a cue at different ages Replicate-individual design: absolute differences between beliefs after exposure to different cues at the same age (measured at different ages)	Difference in belief is measured after each cue; various patterns of cue reliabilities are explored

Table 2.1 Comparison of formal models of sensitive periods. The first column describes the paper in which a model was published. The second column describes whether in this model an organism's decisions depend only on its beliefs, or also on its phenotype. Note: the term "belief", in this context, refers to the information available to an organism about its environment as a function of its prior and the cues it has sampled during its lifetime. It does not necessarily imply conscious deliberation or even psychological representation. The third column describes how plasticity is measured. The fourth column provides additional detail about the testing paradigm (e.g., when plasticity is measured).

Stamps and Krishnan (2014a, 2014b, 2017) have usefully studied the effects of within-individual and replicate-individual designs. Within-individual designs measure plasticity as the difference in belief both before and after exposure to a cue. Replicate-individual designs measure the difference between beliefs of two organisms after each is exposed to a different cue. If such designs would be used in all future models, this could accelerate the development of an integrative theoretical framework of sensitive periods.

2.5 Gaps and future directions

We have described the tenets of evolutionary models of sensitive periods, insights they provide, predictions they make, and a selection of empirical research on humans and other animals. We now turn to gaps in the literature as well as future directions. We have already discussed the need for more formal models of the evolution of sensitive periods in mid-ontogeny. We have also stressed the need for formal modelers to use a consistent set of methods for exposing changes in plasticity across the life course, ideally matching commonly-used empirical paradigms, such as studies of adoption and migration, cross-fostering, and dose-dependent experience. We discuss four other future directions.

First, the field needs more models that explore environmental variation occurring within the lifetime of individuals. For instance, in our model of varying cue reliabilities (see Section 2.4), experience varies quantitatively (i.e., more or less reliable cues), but not qualitatively (i.e., different kinds of experience). However, real animals often face different kinds of adaptive challenges and different types of information at different life stages (Bjorklund, 1997; Turkewitz & Kenny, 1982). Consider, for instance, the human development in the first year of life: "the training sets for statistical learning develop as the sensorimotor abilities of the infant develop, yielding a series of ordered datasets for visual learning that differ in content and structure between time-points but are highly selective at each time-point. These changing environments may constitute a developmentally ordered curriculum that optimizes learning across many domains" (Adolph & Hoch, 2019, p. 325; see also Smith et al., 2018). Moreover, starting at birth, infants are active agents that scan their environments and select the objects and events they attend to (Gibson, 1988). Future models could explore the evolution of sensitive periods when the experiences of organisms vary qualitatively over ontogeny, either because individuals create a curriculum for learning (and how organisms do this might itself be under selection), or because the environmental state varies over ontogeny (e.g., seasonality). Which inputs inform experience at different times of life will depend on the statistical structure of the environment. Models of such inputs should therefore be informed by empirical measures of environmental statistics – in particular, cue reliability and autocorrelation of environmental states – known to be critical dimensions in the evolution of sensitive periods (Frankenhuis, Nettle, et al., 2019).

Second, few models of the evolution of sensitive periods have explored the evolution of sequences of sensitive periods. Neuroscience suggests that such sequences help to build a well-structured brain. For instance: "Each succeeding large-scale region of cortex may […] be thought of as processing increasing orders of invariants from the stimulus stream, and passing either the invariant information extracted from the stream, or the residual

information once the invariant is extracted, forward to other regions of the brain" (Shrager & Johnson, 1996, p. 1119). In this way, the organism incrementally learns about higher-order invariants and adapts to them. Perceptual 'constraints,' previously thought to be limitations, might actually facilitate this process, i.e., be adaptations. For instance, infants' vision starting out blurry may help with basic-level category learning (French et al., 2002).

There are models of sensitive periods in neural development, which are often based on neural network architectures (Bullinaria, 2003; Ellefsen, 2013; Seidenberg & Zevin, 2006). These models typically explore the effects of proximate factors and processes, such as neuromodulation, on sensitive period development. These models are able to generate, for instance, changes in plasticity that resemble those produced by neural processes (e.g., sequences of sensitive periods). The goal is typically to evaluate existing hypotheses, or to generate novel hypotheses, about proximate factors and processes involved in sensitive periods (e.g., how heterogeneity in experience might affect the ability to learn language; Seidenberg & Zevin, 2006). These models are not designed, however, to provide insight into the evolution of development. Future modeling could integrate both adaptive function and mechanism (e.g., exploring the evolution of mechanisms that produce progressive sequences of sensitive periods). Such modeling would fit the current agenda in evolutionary biology to better integrate mechanisms into optimality models of behavior (Frankenhuis, Panchanathan, et al., 2019; Kacelnik, 2012; McNamara & Houston, 2009; Trimmer et al., 2012).

Third, evolutionary models of sensitive periods to date assume that natural selection has equipped organisms with instructions for adaptive behavior; that is, organisms are born knowing which decisions are adaptive given their current phenotypic and/or belief states. This assumption is appropriate when modeling certain traits, such as defensive armor that Daphnia (crustaceans) grow to protect against predation (Agrawal et al., 1999), which do not depend on feedback during an individual's lifetime. There are many cases, however, where the development of phenotypes depends on learning from past behaviors (Snell-Rood, 2012). When organisms do not come equipped with instructions for adaptive behavior, but need to learn such instructions, a division of labor arises: natural selection shapes the learning mechanisms and the developing organism learns how to behave adaptively (Frankenhuis, Panchanathan, et al., 2019). Recent models have started to examine how natural selection might shape the reinforcement learning mechanisms that enable organisms to learn adaptive behaviors (e.g., Dridi & Lehmann, 2016; Enquist et al., 2016; Singh et al., 2010; Yeh et al., 2018). These models however have, to our knowledge, yet to examine the evolution of sensitive periods in the learning of adaptive behavior. This would be a very exciting direction for future research. Moreover, organisms do not only learn specific adaptive behaviors (e.g., how to crack a nut), but also learn and select among broader decision-making strategies. For instance, depending on the level of control that an organism can exert over its environment (agency), it might calibrate its behavioral responses along a continuum ranging from proactive ('What can I do in this environment?') to reactive ('What can this environment do to me?') (Moscarello & Hartley, 2017). Future modeling should therefore explore the evolution of sensitive periods in the development of specific behaviors as well as broader decision-making strategies. This work may connect with recent

models of meta-reinforcement learning, which examine when agents should compose meta-policies that can switch among a set of previously learned policies.

Fourth, our future directions so far have focused on ways in which evolutionary models can be made more relevant to research in cognitive developmental neuroscience. On the other side of the bridge, neuroscientists can help to foster synergies by including high-quality measurements, ideally longitudinally, of the physical and social environment experienced by individuals in studies on neural development. Evolutionary modelers can use such measurements to estimate environmental statistics, such as cue reliability and environmental autocorrelation (Frankenhuis, Nettle, & Dall, 2019). Adaptation is essentially about the fit between individuals and their environments. Understanding adaptation, therefore, requires high-quality measurements of both. Large-scale longitudinal research projects that include detailed measurements of individuals' environments (objective measures) and lived experience (subjective measures), as well as measurement of cognitive and neural development, hold particularly great promise for synergies with evolutionary modeling.

To end: evolutionary biologists have been interested in adaptive behavior and development ever since Charles Darwin proposed the process of natural selection. Only recently, however, evolutionary biologists have developed an explicit interest in adaptive changes in levels of plasticity over the life course. This trend fits perfectly with the long-lasting focus on sensitive periods in developmental cognitive neuroscience. The time is ripe, therefore, for stronger ties and novel synergies between these two exciting fields.

Chapter 3

An evolutionary model of sensitive periods when the reliability of cues varies across ontogeny

This chapter is based on:

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2021). An evolutionary model of sensitive periods when the reliability of cues varies across ontogeny. *Behavioral Ecology, 33*(1), 1–14. https://doi.org/10.1093/beheco/arab113

Online supplements can be found alongside the published article.

Featured in:

Gee, D. G. (2022). When do sensitive periods emerge later in development? *Trends in Cognitive Sciences, 26*(2), 97–98. https://doi.org/10.1016/j.tics.2021.12.001

3.0 Abstract

Sensitive periods are widespread in nature, but their evolution is not well understood. Recent mathematical modeling has illuminated the conditions favoring the evolution of sensitive periods early in ontogeny. However, sensitive periods also exist at later stages of ontogeny, such as adolescence. Here, we present a mathematical model that explores the conditions that favor sensitive periods at later developmental stages. In our model, organisms use environmental cues to incrementally construct a phenotype that matches their environment. Unlike in previous models, the reliability of cues varies across ontogeny. We use stochastic dynamic programming to compute optimal policies for a range of evolutionary ecologies and then simulate developmental trajectories to obtain mature phenotypes. We measure changes in plasticity across ontogeny using study paradigms inspired by empirical research: adoption and cross-fostering. Our results show that sensitive periods only evolve later in ontogeny if the reliability of cues increases across ontogeny. The onset, duration, and offset of sensitive periods — and the magnitude of plasticity — depend on the specific parameter settings. If the reliability of cues decreases across ontogeny, sensitive periods are favored only early in ontogeny. These results are robust across different paradigms suggesting that empirical findings might be comparable despite different experimental designs.

3.1 Introduction

Sensitive periods are life stages during which experiences shape an organism's phenotypic development to a greater extent than other stages (Bateson, 1979; Fawcett & Frankenhuis, 2015). While heightened phenotypic plasticity early in life appears to be the norm, it is by no means the rule. As with everything else in nature, the timing of sensitive periods varies. Sensitive periods may vary in their onset, duration, and offset across species, within species, and even among different traits within a single individual. Zebra finches learn their songs early in life, while European starlings are lifelong learners. Human children vary in the extent to which exposure to adversity affects maturation rate (Belsky & Pluess, 2009; Del Giudice et al., 2011). And, for children adopted from harsh conditions into supportive ones, cognitive and emotional systems adjust at different rates (Tottenham et al., 2010; Zeanah et al., 2011). Decades of empirical research have advanced our understanding about the neurobiological bases of such variation in sensitive periods (Creanza et al., 2016; Knudsen, 2004), so much so that it is possible in some cases to experimentally modify the timing and duration of sensitive periods, and even to "reopen" sensitive periods that had already closed, through physiological intervention (Reh et al., 2020; Takesian & Hensch, 2013).

Existing models of sensitive period evolution

The theory exploring the conditions in which natural selection favors the evolution of phenotypic plasticity is well developed and understood (Chevin & Lande, 2011; Lande, 2014, 2019; Pigliucci, 2005; Via et al., 1995). Recently, formal modeling has focused on the timing of plasticity over the life course. These models explore the selection pressures that shape sensitive periods (reviewed in Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017). More specifically, they explore how the impact of experience on phenotypic development varies across ontogeny.

A general result of models to date is that sensitive periods are typically only favored early in ontogeny. This result has been observed in a variety of scenarios, including when organisms integrate information inherited through genes (or epigenes) with individual experience (Stamps & Krishnan, 2014a, 2014b, 2017), when organisms develop social behaviors such as helping (Kuijper et al., 2019), when experiences simultaneously impact the phenotype (e.g. a non-lethal predator attack reducing somatic quality) and allow learning about the environment (e.g. updating estimates of predator density) (English et al., 2016), and when organisms build phenotypes incrementally rather than instantaneously (e.g. predator defenses in Daphnia; (Whitman et al., 2009)) while sampling imperfect cues to the environmental state (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016). The duration of plasticity typically depends on the degree to which uncertainty about environmental conditions persists across ontogeny. Organisms that are able to reduce their uncertainty faster often lose plasticity earlier than organisms that remain uncertain.

To our knowledge, only two models have documented the evolution of sensitive periods later in ontogeny (Fischer et al., 2014; Stamps & Krishnan, 2017). In these models the highest levels of plasticity occur halfway through ontogeny. Both models assume that organisms start ontogeny with an induced phenotype and that development is fully reversible and unconstrainted, such that organisms can express any phenotype at any time

during ontogeny. Although these models find that age-dependent declines in plasticity are favored across the majority of explored conditions, they also find that plasticity may first increase early in ontogeny before decreasing when there is a discrepancy between organisms' inherited information and early-life experiences; that is, when these two sources of information indicate different states of the world. More generally, this discrepancy rule is said to cause small increases in plasticity early in ontogeny in Bayesian models of development, of which the Stamps and Krishnan (2017) model is one example (Fawcett & Frankenhuis, 2015; Stamps & Frankenhuis, 2016).

All models to date – i.e., those that find sensitive periods early in ontogeny as well as those that find sensitive periods halfway through ontogeny – have assumed that the cue reliability is constant within the lifetime of an organism. It is unknown how this assumption affects their shared finding that sensitive periods are typically favored early in ontogeny, rarely halfway through ontogeny, and never at the end of ontogeny. In this paper, we present a mathematical model that explores the timing of sensitive periods favored by natural selection when the cue reliability varies across ontogeny.

When are mid-ontogeny sensitive periods adaptive?

Though less common, sensitive periods in later developmental stages are widespread. In mammals, experiences during adolescence typically influence adult social behavior to a greater degree than experiences during childhood (Buwalda et al., 2011; Mutwill et al., 2020; Sachser et al., 2020). For example, adolescent guinea pigs housed in large colonies respond to being transferred to a new colony by developing lower levels of stress and aggression as adults, more so than juvenile guinea pigs do (Sachser et al., 2018). In humans, adolescence seems to be a period of enhanced plasticity in several neural and cognitive traits (Blakemore & Mills, 2014; Dahl, 2004; Fuhrmann et al., 2015; Knoll et al., 2016; Larsen & Luna, 2018). For example, adolescents are more sensitive to the effects of social stress, such as social isolation, on mental health, and are more capable of recovering from those same social stressors compared to children and adults (Fuhrmann et al., 2015). Recent work suggests that adolescents, more so than children or adults, rely on learning strategies that are specifically suited to exploring novel opportunities and challenges in the environment (Raab & Hartley, 2019).

Some researchers have speculated that natural selection might favor later sensitive periods when the reliability of cues varies across ontogeny (Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017). Variation in cue reliability may arise when the information available to an organism systematically changes across ontogeny. Such a change may happen if organisms use the same cue across ontogeny, but its reliability changes across different developmental stages. Another possibility is that organisms receive cues more frequently at some developmental stages compared to others, and combining cues increases reliability (Fawcett & Johnstone, 2003; Mariette, 2020). A third possibility is that organisms use different cues, with different reliabilities, at different developmental stages. In all three scenarios, natural selection might have adapted organisms to anticipate changes in cue reliabilities across developmental stages.

We explore three patterns of cue reliability across ontogeny: 'increasing', 'decreasing', and 'first increasing and then decreasing' (or 'triangular'). Cue reliability might increase when an animal estimates its competitive ability in adulthood based on its interactions with conspecifics in the juvenile period. For example, during male-male combat animals often use an opponent's relative body size to predict combat outcome and to adjust their behavior accordingly, such as whether to fight or not (Li et al., 2018; Matsumura et al., 2020; McCullough & Simmons, 2016). As the animal and its conspecifics approach their adult form, relative body size becomes an increasingly reliable indicator of competitive ability in adulthood. Cue reliability might decrease when cues are more frequent, or only available, earlier in life. Prenatal cues, for example, may provide an integrative summary of the experiences of recent matrilineal ancestors, which predicts future nutritional conditions more reliably than early postnatal observations (Kuzawa, 2005). Theoretically, it is also conceivable that cue reliability first increases and later decreases. Although examples of this pattern may be rarer in nature, we speculate that early adolescent social bonds in humans follow such a pattern. Adolescents form strong bonds with peers (Forbes & Dahl, 2010). The feedback adolescents receive from these relationships might be more informative about their social status or mate value in adulthood than feedback received in early childhood or right before the onset of adulthood (Allen et al., 2014; Forbes & Dahl, 2010). We do not explore the cue reliability 'first decreasing and later increasing'. This pattern has not been proposed in the literature nor are we aware of empirical examples in nature.

Our contribution

Here, we develop a model in which organisms sample environmental cues and tailor their phenotypes to the environmental state. Phenotypic development is both incremental and irreversible, in the sense that organisms gradually adjust phenotypes and that developed adjustments cannot be undone. Extending previous work (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016), we introduce variation in cue reliability across ontogeny. We use stochastic dynamic programming to compute optimal developmental policies across a range of evolutionary ecologies. Such a 'policy' prescribes the optimal developmental decision given the organism's state, which comprises the current phenotype and the environmental cues sampled thus far. The optimal policy maximizes expected fitness at the end of ontogeny. We then examine these optimal developmental policies to extract information about the patterning of phenotypic plasticity across ontogeny. In particular, we hope to better understand when natural selection favors the later emergence of sensitive periods.

We also examine how phenotypic variation develops among organisms who follow the same optimal policy. Previous models have shown that individual differences in phenotypes tend to stabilize across ontogeny (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016), but have not quantified this process. To this end, we develop a measure of trait repeatability. Repeatability is widely used in studies of animal personality to quantify consistency in individual differences over time (Fisher et al., 2018; Kok et al., 2019; Polverino et al., 2019; Roberts & DelVecchio, 2000; Trillmich et al., 2018).

Finally, we examine the robustness of our findings by conducting two kinds of sensitivity analyses. First, we quantify patterns of plasticity across ontogeny using paradigms

commonly used in empirical research. This approach links theoretical and empirical research: it allows us to compare qualitative predictions from different empirical paradigms. Second, we investigate differences in patterns of plasticity as a result of simplifying the model. Some models have incorporated phenotype as a fitness determinant (e.g., Panchanathan & Frankenhuis, 2016), and others only the information state of an organism (e.g., Stamps & Krishnan, 2014b, 2017). By comparing these models, we can explore to what extent our qualitative results generalize across models; whether a complex model structure that includes phenotypes alongside information states offers any insights that cannot be obtained from an information-only model; and how information state, both on its own and in combination with phenotype, affects the evolution of mid-ontogeny sensitive periods.

3.2 Model

The organism and the environment

Organisms are born and randomly disperse into discrete and non-overlapping patches which can be in one of two states: E_0 or E_1 (e.g., dangerous or safe). The state of a patch does not change over ontogeny. Organisms sample environmental cues and develop phenotypes, reproduce proportional to fitness, and die. We assume that organisms have adapted to the fixed distribution of patches in the environment (McNamara et al., 2006), and use this distribution at the onset of ontogeny as a prior estimate about the probability of being in one state or the other.

Ontogeny consists of $T = 20$ discrete time periods. Organisms can develop towards two phenotypic targets, P_0 and P_1 which correspond to the optimal, fully specialized phenotypes for E_0 and E_1. Increments toward each of these two phenotypic targets occur on independent dimensions; these two phenotypes are not endpoints of a single and continuous trait (for similar models, see Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016). For example, we might imagine that an organism can invest in a heavily armored phenotype to avoid predation or, instead, invest in a heavily adorned phenotype to attract mates. We track increments towards these targets with two numbers: the number of time periods specialized towards P_0 (denoted by y_0) and towards P_1 (denoted by y_1). At the onset of ontogeny organisms start with 0 specializations towards either phenotypic target ($y_0 = y_1 = 0$). In each time period, organisms receive an environmental cue and then either increment y_0 by 1, increment y_1 by 1, or wait and forgo specialization in this time period (leaving y_0 and y_1 unchanged). We denote the number of time periods waited by y_w.

Development is irreversible in the sense that once a phenotypic increment has developed, it cannot be undone. However, organisms can switch developmental trajectories and specialize towards the other phenotypic target, for instance, because they have revised their estimates. At the end of ontogeny, the number of increments towards P_0 and P_1 and the number of time periods waited sum to the total number of time periods ($y_0 + y_1 + y_w = T$). In this way, phenotypic development is constrained by the duration of ontogeny. The later organisms start specializing towards one of the phenotypic targets, the fewer increments they can make towards it.

Environmental cues provide informative but imperfect guidance. The reliability of a cue indicates the probability of receiving the current cue (C_0 or C_1) conditioned on being in the corresponding environmental state (E_0 or E_1). We assume that organisms 'know' the reliability of a cue because they have adapted to the association between cues and environmental states over evolutionary time. However, because the reliability of cues varies across ontogeny, we denote the cue reliabilities of C_0 and C_1 at time t as $P(C_{0,t}|E_0)$ and $P(C_{1,t}|E_1)$. The probabilities of observing an incorrect cue then correspond to: $P(C_{1,t}|E_0)$ = 1 - $P(C_{0,t}|E_0)$ and $P(C_{0,t}|E_1)$ = 1 - $P(C_{1,t}|E_1)$. We assume that the cue reliability is the same in both environmental states, i.e., $P(C_{0,t}|E_0) = P(C_{1,t}|E_1)$.

Over time, organisms build up a dataset comprising the cues that they have sampled. We denote the sequence of cues until time period t by $D_t = \{x_1, x_2, \ldots x_t\}$, where x_1, x_2, etc. until x_t denote the kind of cue (C_0 or C_1) received in each time period. At any given time t, the state of an organism comprises the developmental decisions it has made and the environmental cues it has received, denoted by the tuple (D_t, y_0, y_1, y_w, t).

We consider three patterns of cue reliability across ontogeny: (1) linearly increasing, (2) linearly decreasing, and (3) first linearly increasing and then linearly decreasing (triangular). All three patterns range between a minimum cue reliability of 0.55 and a maximum cue reliability of 0.95. We ensure that the average cue reliability across ontogeny is the same across cue reliability patterns. This controls for the total information available to organisms across all of ontogeny. To explore whether results are driven by the maximally attainable cue reliability, we also computed results for patterns ranging between 0.55 and 0.75 (see Appendix 1, Figure A1.1). Results from both ranges were qualitatively similar, so we report only the range 0.55 to 0.95 in the main text.

We assume that organisms are Bayesian learners (Dall et al., 2015; Mangel, 1990; McNamara et al., 2006; Mcnamara & Houston, 1980; Stamps & Frankenhuis, 2016; Trimmer et al., 2011; Tufto, 2000), using the fixed distribution of patches as the prior estimate of the environmental state and the time-dependent cue reliabilities to update these estimates (Stamps & Frankenhuis, 2016). To see how this works, suppose an organism has sampled a specific sequence of cues $D_{t=3} = \{C_0, C_1, C_0\}$. According to Bayes' theorem, its posterior estimate after the first cue is:

$$P(E_0|C_0) = \frac{P(C_0|E_0) \bullet P(E_0)}{P(C_0|E_0) \bullet P(E_0) + P(C_0|E_1) \bullet P(E_1)} \qquad (1)$$
$$P(E_1|C_0) = 1 - P(E_0|C_0)$$

To compute the posteriors $P(E_0|D_t)$ and $P(E_1|D_t)$ after the whole sequence of cues, we have to reapply Bayes' theorem for each cue using the previous posterior as the new prior. We provide an overview of our variables and the Bayesian inference in Appendix 1 ('Dynamic programming equations' sections a and b). Additionally, we depict which posteriors result from different cue reliability patterns and priors in the online supplements.

Mapping from phenotypes to fitness

We assume that fitness is accrued at the end of ontogeny (e.g., adulthood). A mature organism accrues fitness depending on how well its phenotype matches the environmental state. The better the match, the higher the fitness. Therefore, the earlier an organism specializes, the more it can improve its fit with the environment (Panchanathan & Frankenhuis, 2016). In this way, there is an opportunity cost to delaying phenotypic specialization (Dunlap & Stephens, 2016). In addition, we assume that developing a phenotype that does not match the environmental state reduces fitness, and the penalty magnitude depends on the degree of mismatch (Innes-Gold et al., 2019). We do not, however, assume a constitutive cost of plasticity in the sense that there is no explicit cost for building, running, and maintaining the physiological mechanisms enabling plasticity (Auld et al., 2010; DeWitt et al., 1998; Relyea, 2002). Nor do we assume a 'switch cost' if organisms switch from specializing from one phenotypic target to another.

Equations (2) – (4) show the mapping of phenotypic increments to fitness rewards and penalties at the end of ontogeny (see also Appendix 1 'Dynamic programming equations', section c). We denote the mature phenotype at the end of ontogeny by $Y_{mat} = (y_0, y_1, T)$. The parameter π_0 corresponds to the baseline fitness of an organism that waited throughout ontogeny, never specializing toward either phenotypic target. The expression $\phi(Y_{mat})$ corresponds to the fitness reward for correct phenotypic specializations. The expression $\psi(Y_{mat})$ corresponds to the fitness penalty for incorrect specializations. Thus, total fitness, $\pi(Y_{mat})$, is:

$$\pi(Y_{mat}) = \pi_0 + \phi(Y_{mat}) + \psi(Y_{mat}) \quad (2)$$

We explore three mappings between phenotypic increments and fitness effects. With 'linear' fitness effects, each correct (or incorrect) increment results in a constant marginal fitness gain (or loss). With 'decreasing' fitness effects, the marginal fitness gain (or loss) of each correct (or incorrect) increment decreases. And with 'increasing' fitness effects, the marginal fitness gain (or loss) of each correct (or incorrect) increment increases. The formulas for these mappings can be found in Appendix 1 'Dynamic programming equations', section c. The attainable fitness payoff for a perfectly matched organism is the same for each fitness mapping and for each environmental state.

To see how these mappings work, suppose that an organism has sampled a specific sequence of cues, D_t, throughout ontogeny. Its posterior estimates $P(E_0|D_{t=T})$ and $P(E_1|D_{t=T})$ reflect the probabilities of being in either environmental state at the end of ontogeny. Thus, to compute rewards and penalties, we need to compute the expectation across both environmental states, weighted by how likely each state is, as indicated by the posterior estimates at the end of ontogeny. We denote the mapping from phenotypic increments to rewards and penalties by $f(y)$, where y can refer to both y_0 and y_1, and derive the following expressions for expected rewards and penalties:

$$\phi(Y_{mat}) = P(E_0|D_{t=T}) \bullet f(y_0) + P(E_1|D_{t=T}) \bullet f(y_1)$$
$$\psi(Y_{mat}) = -\big(P(E_0|D_{t=T}) \bullet f(y_1) + P(E_1|D_{t=T}) \bullet f(y_0)\big) \quad (3)$$

Inserting this into equation (2) results in the final formula for total fitness at the end of ontogeny:

$$\pi(Y_{mat}) = \pi_0 + P(E_0|D_{t=T}) \bullet f(y_0) + P(E_1|D_{t=T}) \bullet f(y_1) - \big(P(E_0|D_{t=T}) \bullet f(y_1) + P(E_1|D_{t=T}) \bullet f(y_0)\big) \quad (4)$$

Optimal developmental policies

We use stochastic dynamic programming to compute optimal developmental policies for different evolutionary ecologies (Mangel & Clark, 2019; Mcnamara & Houston, 1980). We explore three prior distributions of environmental states: $P(E_0) = P(E_1) = 0.5$, $P(E_0) = 0.3$ and $P(E_1) = 0.7$, and $P(E_0) = 0.1$ and $P(E_1) = 0.9$; and three cue reliability patterns: increasing, decreasing, and triangular. For each possible state of an organism (D_t, y_0, y_1, y_w, t), stochastic dynamic programming identifies the developmental decision that will result in the highest expected fitness at the end of ontogeny. In the event of a tie between two or more options in a particular state, the organism chooses amongst the current alternatives with equal probability. $F(D_t, y_0, y_1, y_w, t, T)$ denotes the maximum expected fitness that can be attained as a result of decisions made between t and T. The organism chooses option a to maximize expected fitness:

$$F(D_t, y_0, y_1, y_w, t, T) = \max_{a \in \{0,1,w\}} F_a \text{, where}$$
$$F_0 = F(D_t, y_0 + 1, y_1, y_w, t + 1, T),$$
$$F_1 = F(D_t, y_0, y_1 + 1, y_w, t + 1, T),$$
$$F_w = F(D_t, y_0, y_1, y_w + 1, t + 1, T). \quad (5)$$

For each possible state of an organism we initialize $F(D_t, y_0, y_1, y_w, T, T)$, which represents the fitness at the end of ontogeny, with $\pi(Y_{mat})$ as defined in equation (4). Using this as a starting point, we solve equation (5) via backwards induction. We also describe our approach to computing optimal policies in Appendix 1 ('Dynamic programming equations', section d). Our code, written in Python 2.7, is available on GitHub (https://github.com/Nicole-Walasek/sensitive-periods-with-varying-cue-reliabilities).

Quantifying plasticity

We use a simulated 'twin' study to quantify trajectories of plasticity across ontogeny. We first simulate 10,000 pairs of twins with identical phenotypes and posteriors following the optimal policy up to time period t. From each pair we keep one twin in its natal patch (the focal individual) and move the other one into a "mirror" patch (the clone). From the time of separation until the end of ontogeny, the focal individual and the clone receive opposite environmental cues. That is, whenever the focal individual receives a cue indicating E_0, the clone receives a cue indicating E_1, and vice versa.

We then compare the mature phenotypes of twin pairs at the end of ontogeny. We define plasticity as the Euclidean distance between the two twins along the two phenotypic dimensions (y_0 and y_1). The larger the difference between mature twins, the more cues have shaped their phenotypes since their separation; thus, the more developmentally plastic these twins were at the time of their separation. Our paradigm resembles twin studies that compare similarities and differences between adult twins who were separated at different points in ontogeny to assess the impact of genetic and environmental factors on phenotypic development.

We distinguish between two measures of plasticity, 'absolute' and 'proportional'. Absolute plasticity is the average Euclidean distance across simulated twin pairs normalized to range between 0 and 1. Proportional plasticity is the average Euclidean distance between simulated twins, divided by the maximally achievable Euclidean distance from the moment of separation until the end of ontogeny. In contrast to absolute distance, proportional distance accounts for potential phenotypic distance, which is smaller the later separation occurs, as there is less time for twins to diverge. We offer both measures to facilitate comparison across different theoretical models and/or empirical studies that potentially use one or the other (or both) measures to quantify plasticity.

Quantifying rank-order stability

We developed a measure of trait repeatability to quantify the process by which individual differences in phenotypes develop and might stabilize over ontogeny. We assume that organisms within a population that show high trait repeatability also show stable phenotype ranks across ontogeny. Specifically, we assume that the higher trait repeatability is, the lower the likelihood of rank-switches between two time periods. We simulate a population of 10,000 organisms and rank them at each time point during ontogeny based on their phenotypic values. At each time period we compute the proportion of individuals that experiences a rank-switch (relative to the population size, which is constant) from one time period to the next. Organisms that have the same trait value share a rank. This paradigm allows us to compute not only the proportion of rank-switches between consecutive time periods, but also that between periods farther apart.

Sensitivity analyses

We conduct two different kinds of sensitivity analyses. First, we explore whether results are robust to variations in the basic twin study paradigm. Specifically, we vary (1) the degree to which cues sampled by the separated clone differ from those sampled by its identical twin; (2) whether separation is temporary, lasting only for a fixed number of time periods, or permanent until the end of ontogeny; and (3) whether twins are compared directly after separation rather than at the end of ontogeny. These paradigms resemble those used in empirical research with humans in developmental psychology or epidemiology and with non-human animals in behavioral ecology (Frankenhuis & Walasek, 2020). By varying the degree to which cues between separated twins differ we capture 'dose-dependent experience studies' which, for example, study how matched individuals from the same litter respond to different dosages of the same treatment. We vary separation duration and measurement time to capture 'cross-fostering studies', in which a subset of individuals is removed from their natal environment and raised in a different environment for some time

to disentangle the effects of rearing environment and subsequent differences in experience on phenotypic development. Differences between separated individuals and control individuals can be measured at the end of the separation duration, or at some later time after the separated individuals have been reintroduced to their original environment. By comparing these different paradigms, we are able to explore to what extent developmental trajectories of phenotypic plasticity uncovered in empirical studies may vary as a function of study paradigm. We depict the paradigms in Figure 3.1. We show the trajectories of phenotypic plasticity that result from these different paradigms in Appendix 1, Figures A1.6-A1.9.

Second, we explore the extent to which optimal decisions depend on phenotypic states versus posterior estimates. In our model, optimal decisions depend both on an organism's phenotypic state and on its posterior estimate. Accordingly, organisms with identical posteriors might make different decisions because their previously constructed phenotypes differ. Other kinds of models, however, have assumed a one-to-one mapping between posteriors and phenotypes (e.g., Stamps & Krishnan, 2014b, 2017). To explore to what extent the inclusion of phenotypic states yields qualitatively different outcomes than a posterior-only model, we compare patterns of plasticity derived from both models when the reliability of cues varies across ontogeny. In line with our basic twin study paradigm, we compare the average proportional phenotypic distance and average difference in posterior estimates across 10,000 simulated pairs of twins at the end of ontogeny, following permanent separation.

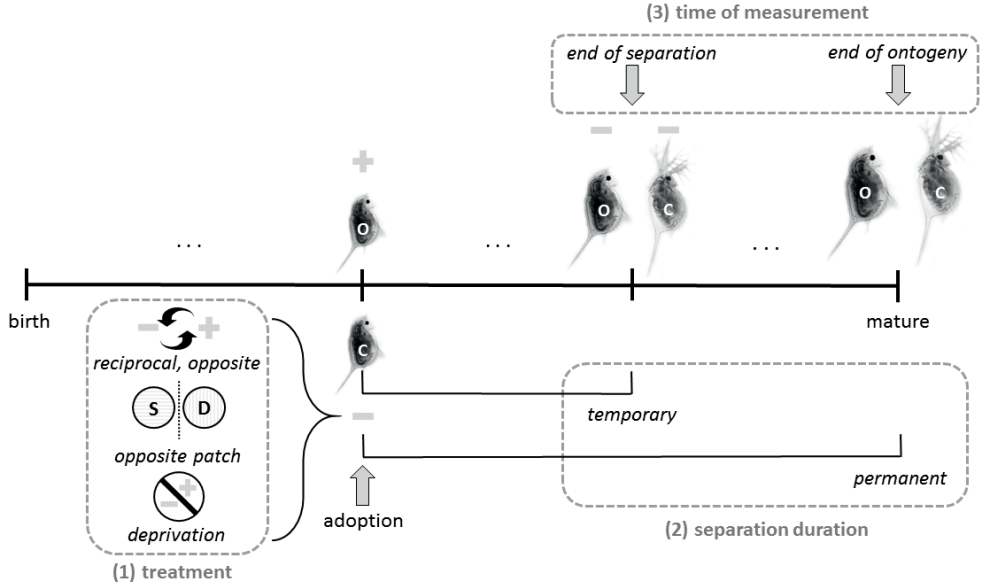


Figure 3.1 Measuring changes in plasticity across ontogeny. We separate twins (original, denoted with O, and clone, denoted with C) at different ages. We vary three dimensions: treatment, separation duration, and time of measurement. (1) Treatment refers to how the experiences of the original and clone differ during their separation. The clone might experience reciprocal opposite cues; cues from the opposite patch; or deprivation. With reciprocal opposite cues, the clone always samples the opposite cue of the original: if the original samples a minus cue [-], the clone samples a plus cue [+]. With cues from the opposite patch, the clone samples a sequence of cues typical of the opposite patch: if the original tends to sample more minus cues, the clone tends to sample more plus cues. In our Figure, the original and the clone are both in the dangerous patch (denoted with D), but the clone receives cues typical of the safe patch (denoted with S). With deprivation, the clone is equally likely to sample a plus or a minus cue; thus preventing learning about the environment. (2) Separation duration refers to whether the separation of twins is permanent or temporary. Permanent separation occurs if twins experience different conditions from their separation until the end of ontogeny (maturity). Temporary separation occurs if twins are reunited before the end of ontogeny. (3) Time of measurement refers to when differences in the phenotypes of twins are measured. We measure differences in phenotypes of twins at two different time points: at the end of their separation and at the end of ontogeny. Our results show that different treatments tend to produce (qualitatively) similar patterns of plasticity. Our predictions are therefore similar for different treatments and for different measurement times used in empirical research. Copyright: this Figure has been adapted from Frankenhuis and Walasek (2020) and we have used the images of Daphnia with permission from Dr. Linda Weiss (2019).

3.3 Results

In the main text, we describe results for linear rewards and linear penalties. We present results from other combinations of reward and penalty functions in SM 5 of the online supplements. We also provide additional analyses allowing comparison of results from this model with results of a previously published model of incremental development exploring fixed cue reliabilities (Panchanathan & Frankenhuis, 2016), in SM 5 of the online supplements and Appendix 1 (Figures A1.2-A1.3).

Sensitive periods may occur halfway through ontogeny

With absolute plasticity, sensitive periods are only favored early in ontogeny (Figure 3.2, grey lines and squares). With proportional plasticity, natural selection might favor sensitive periods in mid-ontogeny, but only if the cue reliability increases across ontogeny or first increases and then decreases, resulting in a triangular pattern (Figure 3.2, black lines and circles). Peaks are higher for the triangular cue reliability pattern, because the reliability of cues increases more rapidly during the first half of ontogeny. With decreasing cue reliabilities, sensitive periods evolve only early in ontogeny.

In most conditions, optimal policies track the cue reliabilities across time, meaning plasticity is highest when the cue reliability is highest. However, this is not always the case. When the cue reliability increases, plasticity peaks halfway through ontogeny, while cues are moderately reliable. By then, some organisms—those who have sampled consistent cue sets (see below)—have achieved a high level of confidence and their plasticity starts to decline.

Prior distributions only have a quantitative but not a qualitative impact on these patterns: the more uniform the prior distribution is, the lower the level of overall plasticity across ontogeny, as measured by the area under the curve. This small effect of prior is moderated by the cue reliability. Prior distributions have the strongest effect when cue reliabilities peak only at the end of ontogeny (increasing cue reliability). This makes sense. When information quality is low and one environmental state is much more likely than the other, organisms eschew plasticity and pick the more likely option.

Early in ontogeny prior distributions shape posterior estimates and thereby affect phenotypic development. However, as ontogeny proceeds and organisms sample more cues, the adjustment in posteriors and phenotypes in response to cues converges and becomes independent of the initial prior and cues take over in shaping both posteriors and phenotypes (Appendix 1, Figure A1.4). Eventually, phenotypic plasticity declines regardless of the prior distribution and cue reliability pattern. Plasticity declines more steeply if organisms have access to reliable cues earlier in ontogeny, as is the case for the decreasing and triangular pattern. More reliable cues imply more consistency in cue sequences. Thus, the optimal policy instructs organisms to lose plasticity early in ontogeny and to invest into phenotypic specialization to reap fitness benefits (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016).

Previous models of stable environments also find that plasticity is higher early in ontogeny and then rapidly declines when cue reliability is high and constant across ontogeny (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Stamps & Krishnan, 2017). Organisms use highly reliable cues at the onset of ontogeny to drastically reduce uncertainty about their environment, eliminating the need for continued plasticity. For the same reason, we find early-ontogeny sensitive periods with the decreasing cue reliability pattern. Combining our findings and those from previous models, we speculate that in environments that are stable across ontogeny, any pattern in which cues are highly reliable at the onset of ontogeny will lead to sensitive periods early in ontogeny. However, when the state of the environment fluctuates across ontogeny, we speculate that highly

reliable cues early in ontogeny are often not sufficient to reduce uncertainty about future conditions. Under these conditions, natural selection may favor prolonged plasticity if the reliability of cues decreases across ontogeny, or even multiple sensitive periods if the cue reliability first decreases and then increases.

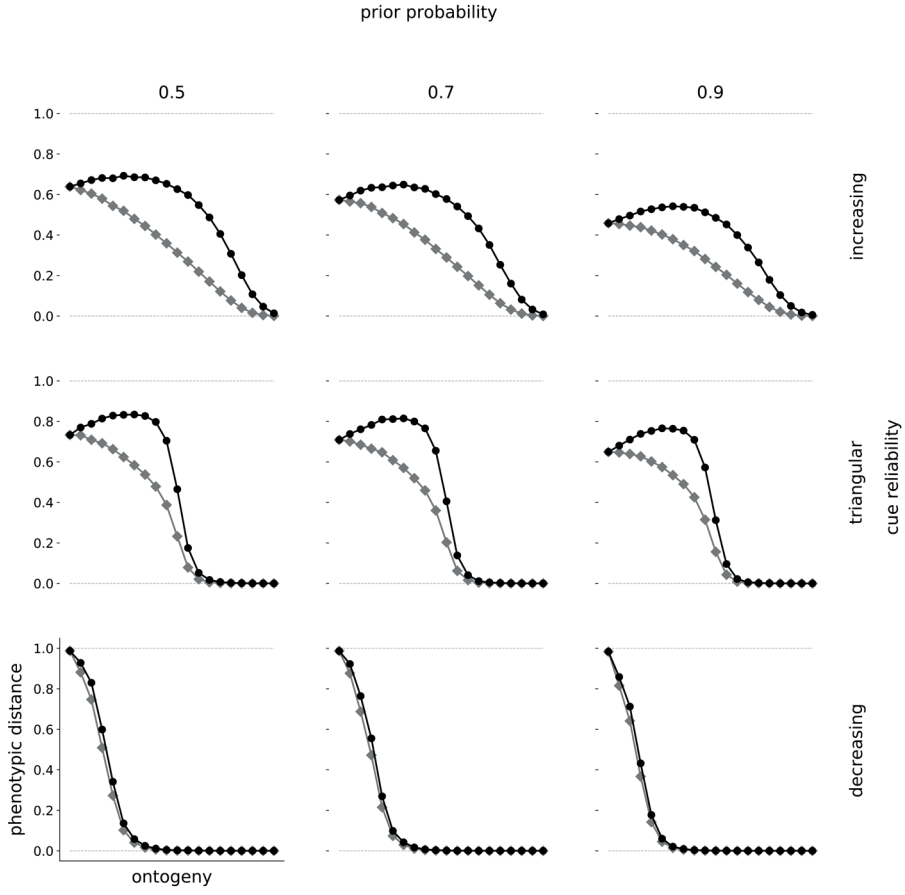


Figure 3.2 Plasticity across ontogeny. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels (see SM 5 in the online supplements for other combinations of rewards and penalties). The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance at the end of ontogeny between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance following separation.

Individual differences in sensitive periods

Across the entire range of parameter values, natural selection favors early plasticity that declines across ontogeny and tends toward zero by the end (Figure 3.3). This gradual decline in plasticity, however, masks substantial individual variation in the onset, duration, and offset in sensitive periods. Organisms that receive consistent cue sets early in ontogeny become insensitive to cues earlier in ontogeny, whereas organisms that receive inconsistent cue sets prolong plasticity. Consistent cue sets are those in which a large fraction of cues indicate one environmental state over the other.

The consistency of cue sets is related to the cue reliability pattern. When the reliability of cues decreases, cue sets are relatively consistent early in ontogeny and inconsistent later in ontogeny. In this case, natural selection favors early sensitivity and a rapid decline in plasticity across ontogeny. When the reliability of cues first increases and then decreases (triangular pattern), cue set consistency at first increases and peaks at mid-ontogeny before turning and becoming increasingly inconsistent. Here, plasticity declines rapidly after mid-ontogeny. When the reliability of cues increases, early cue sets are inconsistent and become increasingly more consistent over time. Organisms in this case prolong plasticity well beyond mid-ontogeny.

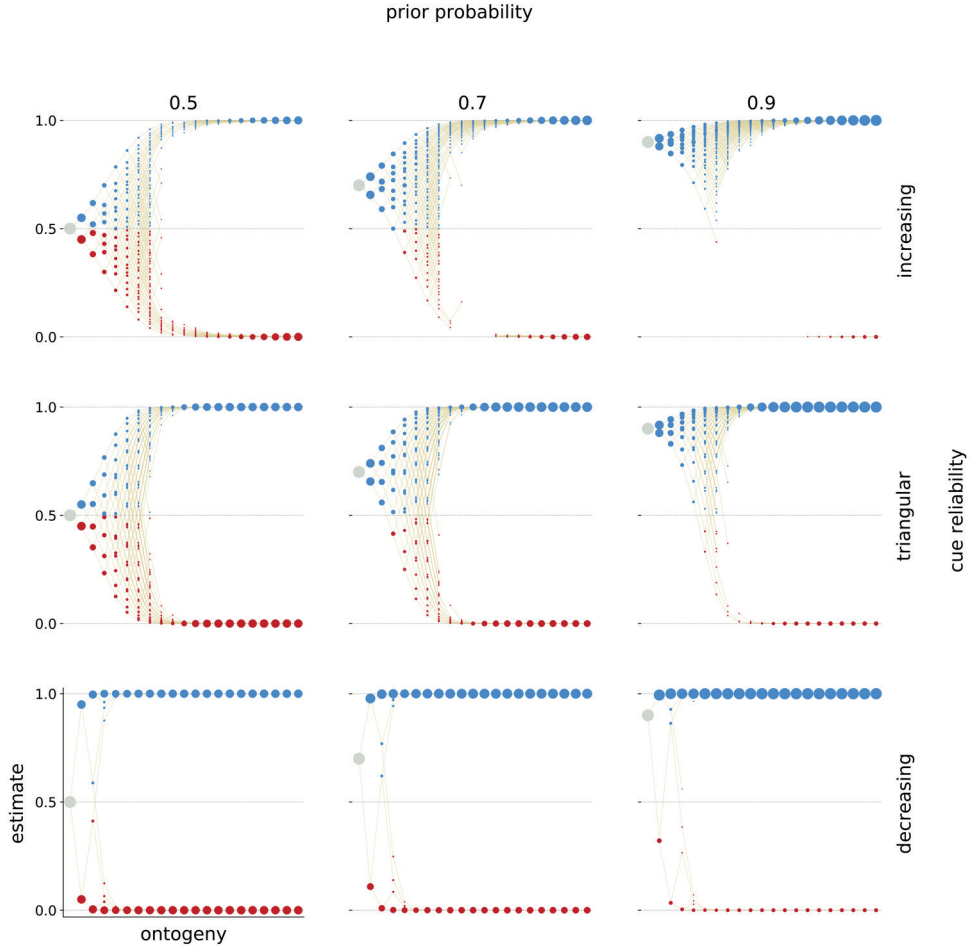


Figure 3.3 Optimal developmental policies. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels (see SM 5 in the online supplements for other combinations of rewards and penalties). The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel depicts the optimal developmental policy for the corresponding parameter values as well as information about the probability of reaching each possible state. The horizontal axis shows developmental time and the vertical axis shows an organism's estimate of being in E_1. Each organism begins ontogeny with the same prior (large grey circle) and then, in each time period, samples a cue, updates its posterior, and makes a phenotypic decision. Beige lines represent possible changes in posteriors across development, tracking possible developmental trajectories. Colored circles represent phenotypic decisions: black indicates waiting, red specializing towards P_0, and blue specializing towards P_1. The area of a circle is proportional to the probability of reaching the corresponding state. These probabilities sum to one within a time period. We only show states that have a probability of more than 0.5% of being reached.

Repeatability depends on the environment

We track the proportion of rank-switches in a population of developing organisms across ontogeny to infer trait repeatability. This allows us to quantify and visualize how individual phenotypic differences develop and stabilize over time.

Across prior distributions and cue reliability patterns, the proportion of rank-switches decreases as ontogeny proceeds indicating an increase in trait repeatability (bar charts in Figure 3.4). Over time organisms become more certain of their environmental state and consistently specialize towards it. When the prior distribution is uniform (0.5; left column, Figure 3.4), the decrease in rank-switches across ontogeny is accelerated when organisms have access to highly reliable cues early in ontogeny (decreasing cue reliability; bottom row, Figure 3.4). When cue reliability is low early in ontogeny (increasing or triangular cue reliability; top and middle rows, Figure 3.4), organisms 'drift' early on, resulting in more rank-switches, and only settle on specialization trajectories later on, after sufficiently reducing uncertainty about the environmental state.

When the prior distribution is informative (0.7 or 0.9; middle and right columns, Figure 3.4), the proportion of rank-switches might increase during mid-ontogeny when the cue reliability starts out low and increases over time, as a consequence of increasing or triangular cue reliabilities; top and middle rows, Figure 3.4). Under these conditions, the majority of organisms within a population starts specializing towards the same environmental state based on their priors, which keeps the proportion of rank-switches low. As the cue reliability begins to increase, organisms' posteriors are more likely to shift, leading to more phenotypic 'drifting'. This 'drifting' increases the proportion of rank-switches and thus temporarily lowers trait repeatability.

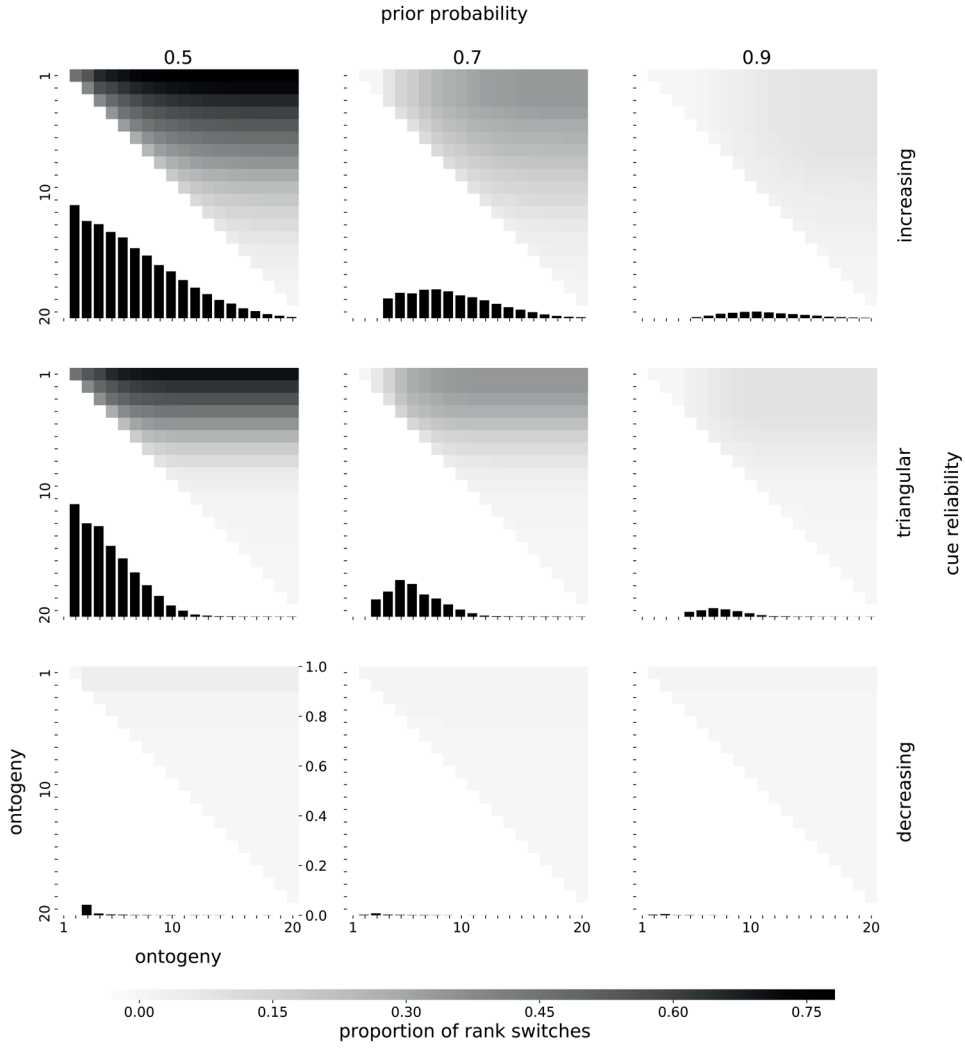


Figure 3.4 Rank-order stability. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels (see SM 5 in the online supplements for other combinations of rewards and penalties). The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel depicts a simulation of 10,000 organisms following the optimal policy across ontogeny. The environmental state is fixed to E_1. In each time period, organisms are ranked according to the number of specializations towards P_0. Organisms with the same number of specializations share a rank. Each square panel depicts two sets of results, one in the upper right triangle and another in the lower left triangle. For the upper right triangle, the relevant axes are the horizontal and the left vertical, each depicting the full range of ontogenetic time periods. Each cell in this triangle indicates the proportion of rank-switches occurring from the time period on the horizontal axis to the time period on the (left) vertical axis in gray scale, with lighter cells indicating fewer rank-switches and darker cells more rank-switches. The lower left triangle within each panel zooms in on the diagonal of the upper right triangle, depicting the proportion of rank-switches between consecutive time periods in a bar chart. We highlight this scenario as it is the most relevant for empirical research on animal personality where repeatability is typically measured across consecutive years. For this portion of the panel, the horizontal axis depicts ontogenetic time periods and the right vertical axis depicts the proportion of rank-switches in that time period.

Results are robust to study paradigm

We conducted the simulated twin study (depicted in Figure 3.2) under different study paradigms, resembling those used in empirical studies of ontogenetic changes in phenotypic plasticity. To capture a wide variety of empirical paradigms, we vary three dimensions of our original twin study: (1) the degree to which cues sampled by the separated clone differ from those sampled by its identical twin; (2) whether separation is temporary and thus only lasts for a fixed number of time periods, or permanent until the end of ontogeny; and (3) whether twins are compared directly after separation or at the end of ontogeny.

We find that the degree to which cues between separated twins differ does not change qualitative changes in phenotypic plasticity, but merely influences the total magnitude of plasticity across ontogeny (Appendix 1, Figures A1.6-A1.7). Not surprisingly, we observe that larger differences in sampled cues between separated twins result in greater magnitudes of phenotypic plasticity across ontogeny. When the separation of twins is temporary, plasticity measured at the end of ontogeny reflects the long-term effects of this separation. This measure illustrates to what extent the time periods in which twins have reunited buffer against further phenotypic divergence or even initiate phenotypic convergence of twins. When plasticity is measured at the end of ontogeny and cue reliabilities increase, plasticity nonetheless tends to increase towards the end of ontogeny (Appendix 1, Figure A1.8, first row). This indicates that highly reliable cues have a major long-term effect on phenotypic development, even in the later stages of ontogeny. This enduring effect cannot be compensated for by short time windows in which twins have reunited. Plasticity measured directly after temporary separation quantifies the short-term, immediate effects of separation (Appendix 1, Figure A1.9). Immediate effects of separation are largest if separation occurs when twins are uncertain about environmental conditions due to a uniform prior distribution and/or when cues are highly reliable during the window of separation.

Mid-ontogeny sensitive periods might depend on both phenotypes and posteriors

As part of our sensitivity analysis, we explore whether resulting patterns of sensitive periods depend on our assumption of modeling phenotypic states alongside information states. Specifically, we compare patterns of plasticity in a phenotype-and-posterior model (black lines and circles, Figure 3.4), in which developmental decisions are shaped by phenotypic states that are coupled with posterior estimates, and a posterior-only model (gray bars, Figure 3.5), which assumes a one-to-one mapping between phenotypes and posteriors. Unfortunately, we are not able to meaningfully interpret the differences in magnitude of plasticity between both models, because differences in posterior estimates and differences in phenotypes are measured in different units. The former is computed as the difference in posterior probabilities while the latter is computed as the normalized Euclidean distance between phenotypes. Thus, we will only discuss qualitative differences in patterns of plasticity across models.

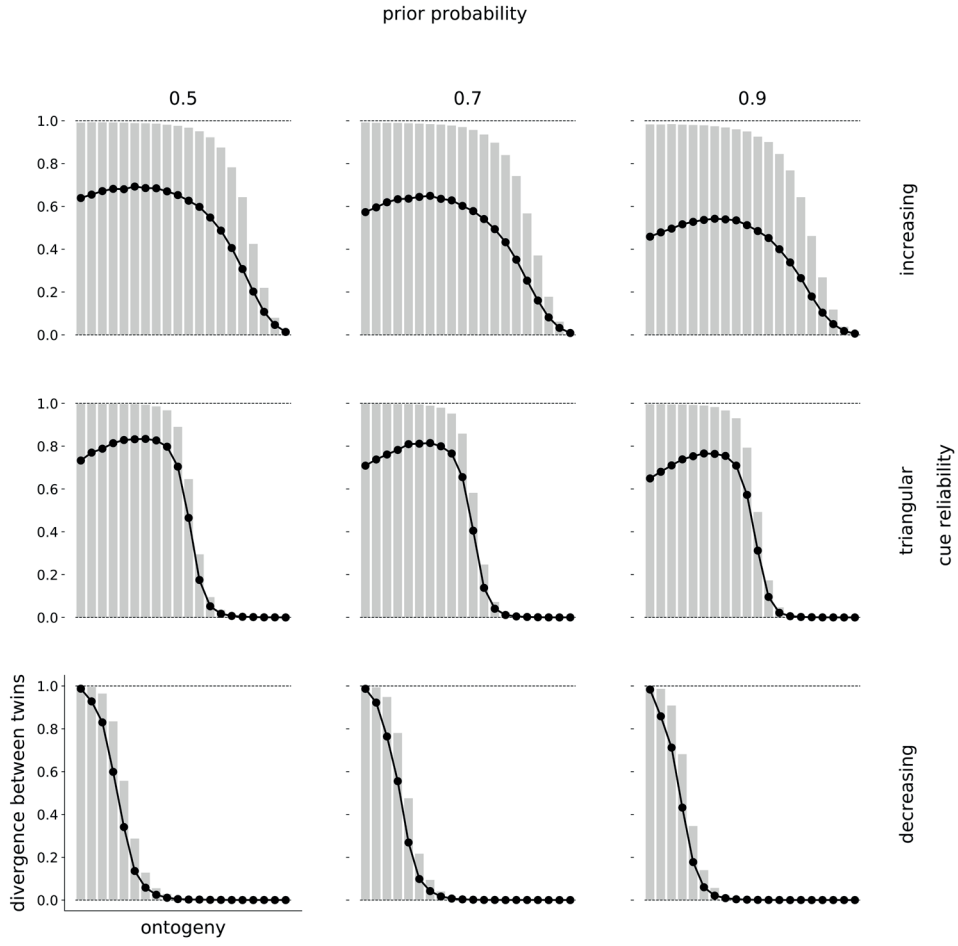


Figure 3.5 Plasticity in phenotype and posterior estimate. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. As in Figure 3.2, each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Black circles correspond to the phenotype-and-posterior and gray bars to the posterior-only model. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts until the end of ontogeny. The vertical axis within each panel depicts the difference between focal individuals and their clones in the phenotype-and-posterior model and the posterior-only model. The horizontal axis depicts the time period in which pairs of twins were separated. Black lines and circles depict the average Euclidean distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range), divided by the maximum possible distance following separation. Gray bars correspond the average absolute distance in posteriors between those same simulated organisms at the end of ontogeny.

Qualitative patterns look largely similar across models. Across all parameter combinations phenotypic plasticity tends to decline with age. However, we also observe differences: phenotypic distances across separated twins might increase when separation occurs later during ontogeny, while differences in posterior estimates between those same twins decrease or remain unchanged. To illustrate this difference, we plot the gradients of both plasticity curves in Appendix 1 (Figure A1.5). This result implies that natural selection favors mid-ontogeny increases in plasticity in the phenotype-and-posterior model but not in the posterior-only model. Both phenotypic state and information available thus act as selection pressures in our model when shaping mid-ontogeny sensitive periods. Thus, only modeling the information state of an organism is not sufficient to explain mid-ontogeny sensitive periods.

3.4 Discussion

We have modeled the evolution and development of sensitive periods when organisms construct their phenotypes incrementally and the reliability of cues varies across ontogeny. We used stochastic dynamic programming to compute optimal developmental policies across a range of evolutionary ecologies, varying the prior distribution of environments, the cue reliability pattern, and the mapping of phenotype onto fitness. From these optimal policies, we derived changes in phenotypic plasticity across ontogeny. We discuss five insights from our model and limitations and future directions.

Mid-ontogeny sensitive periods may evolve when the reliability of cues increases

We find that sensitive periods evolve in mid-ontogeny when the reliability of cues is low at the onset and increases over, at least some portion of, ontogeny. Unlike previous models (Fischer et al., 2014; Stamps & Krishnan, 2017) we find mid-ontogeny sensitive periods when prior and acquired information are consistent with each other, thus identifying increases in cue reliability as the cause of increases in plasticity in our model. Moreover, whereas the previous models find a relatively small plasticity bump at the beginning of ontogeny, our model produces bumps that extend across a substantial portion of ontogeny.

Fuhrman et al. (2015) reviewed evidence for adolescence being a sensitive period of brain development in humans and distinguished three models of plasticity: a discrete and punctuated period of heightened plasticity in adolescence, a continuous and constant sensitive period across childhood and adolescence, or a continuous and gradual decline of plasticity across childhood and adolescence. Our modeling results suggest that natural selection can favor each of these three models, depending on the evolutionary ecology. For example, if cues were, on average, unreliable early and late in ontogeny, with a peak in mid-ontogeny, natural selection might favor a discrete period of heightened plasticity during mid-ontogeny. When cues are, at first, unreliable and gradually increase in reliability across ontogeny, natural selection might favor a continuous sensitive period across childhood. This pattern is especially favored when the distribution of environmental states is uniform and thus making it harder for developing organisms to predict their environment before having sampled any cues. Lastly, if cues are at first highly reliable and decline in reliability

across ontogeny, natural selection might favor an initially high period of sensitivity with a continuous decline in plasticity across childhood.

There are individual differences in the timing of sensitive periods

Using a similar modeling framework to this paper's, Frankenhuis and Panchanathan (2011a, 2011b) and Panchanathan and Frankenhuis (2016) showed that individuals who sample more consistent cue sets might shed plasticity earlier in ontogeny. Here, we add that the opportunity to gather reliable information later in ontogeny might prolong sensitive periods beyond early ontogeny, even resulting in mid-ontogeny sensitive periods. However, organisms of the same population show inter-individual differences in the level of elevation of plasticity during these mid-ontogeny sensitive periods. Because cues are noisy, organisms of the same population will vary in the extent to which they are certain of the state of their environment due to sampling different sequences of cues, with inconsistent sequences of cues resulting in more uncertainty. The more uncertain organisms are when the opportunity to gather reliable information arises, the higher their peaks in mid-ontogeny sensitive periods. This finding suggests that empirical studies are most likely to observe mid-ontogeny sensitive periods if organisms start out uncertain (e.g. have experimentally evolved a uniform prior) and receive highly reliable cues midway ontogeny.

Individual differences tend to stabilize across ontogeny

Kok et al. (2019) suggest that increased exposure to reliable cues across ontogeny might reduce organisms' uncertainty about their environment, leading to fewer adjustments of phenotypic traits, such as exploration behavior. This process, they argue, might cause age-related increases in trait repeatability. When organisms in our model are initially uncertain about their environment due to a uniform prior distribution and the reliability of cues increases early in ontogeny, we observe such a pattern of increased trait repeatability. Trait repeatability develops earlier in ontogeny the earlier organisms have access to reliable cues.

Research on animal personality includes a focus on the repeatability of phenotypic traits across an organism's lifetime (Sih et al., 2004). The typical pattern in this literature is that the repeatability of phenotypic traits increases across ontogeny in a variety of species, including humans (Fraley & Roberts, 2005; Réale & Dingemanse, 2012; Sih et al., 2004). However, not all studies find this pattern. Wuerz and Krüger (2015), for example, observed large variation in repeatability across different traits and life stages in Zebra finches. Some traits showed no repeatability, while others were only repeatable in some portions of ontogeny. In some cases repeatability even decreased across life stages. Although in our model trait repeatability usually increases, we do find, for instance, that with informative prior distributions (0.7 or 0.9) and increasing or triangular (first increasing, then decreasing) cue reliabilities, trait repeatability might decrease in mid-ontogeny rather than monotonically increase across ontogeny. Our model thus suggests hypotheses about the selection pressures that can result in the more common pattern of increasing trait repeatability and in the less common pattern of decreasing trait repeatability.

Results are robust to study paradigm

When applying different paradigms to quantify phenotypic plasticity, we only observed changes in the overall magnitude of plasticity but no qualitative changes in the

patterns of sensitive periods across ontogeny. We found greater magnitudes of plasticity across ontogeny when simulated twins were exposed to drastically different cues during their separation. Although we do not imply that our model findings are readily applicable to empirical studies of phenotypic plasticity, we do think that they raise two empirical questions that are surely worth exploring: first, whether patterns of plasticity observed in empirical studies using different experimental manipulations are comparable; and second, whether plasticity is larger and easier to detect in empirical studies using more extreme manipulations of individuals' experiences. As a first step, studies might compare trajectories in plasticity derived from different study paradigms in the same species and for the same trait of interest.

Sensitive periods depend on information and phenotypic state

Mid-ontogeny sensitive periods only emerge when the state of organisms includes both posteriors as well as phenotypes, not when state only includes posteriors, if plasticity is measured at the end of ontogeny. However, it is unknown to what extent this finding depends on our specific model assumptions. Thus, an open question is whether the inclusion of phenotypic states is generally, across models of sensitive periods, required for the evolution of mid-ontogeny sensitive periods.

Future work could systematically compare outcomes from posterior-only and phenotype-and-posterior models across different study paradigms (e.g. measuring plasticity at different time points) and model assumptions (e.g. fixed or varying cue reliabilities, stable or fluctuating environments) to study whether the inclusion of phenotypic states is necessary for sensitive periods to be favored in later developmental stages. We have made a small step in this direction by comparing both kinds of models, when differences in posteriors and phenotypes between simulated twins are measured right after their separation, rather than at the end of ontogeny (Appendix 1, Figure A1.4). This measure quantifies the immediate phenotypic effects of the experimental manipulation. In that scenario, we find mid-ontogeny sensitive periods for both a posterior-only and a phenotype-and-posterior model.

Limitations and future directions

We first discuss two specific limitations of our model and then two broader limitations of this class of models. We also suggest future directions that can address some of these limitations.

First, in our model the environment remains stable within an individual's lifetime. Whether this assumption is plausible for a given species depends on its generation time relative to the rate of environmental change (Botero et al., 2014). For long-lived organisms, it is less likely that the environment remains stable throughout their lifetime (Nettle et al., 2013). Also, in a seasonally changing environment, natural selection might increase plasticity at those times when learning and development, or changes in behavior, might enhance fitness. For instance, seasonally breeding adult songbirds exhibit seasonal plasticity in song behavior and the associated brain regions (Tramontin & Brenowitz, 2000). Our model assumes a stable environment throughout the organism's lifespan and so is not designed to capture such phenomena.

Second, we constrained ontogeny to a fixed time horizon. In nature, however, different individuals of the same species might mature at different times as a result of phenotypic plasticity or other processes. In our model, the duration of ontogeny is fixed and fitness is accrued at the end of ontogeny. Plasticity might terminate towards the end of ontogeny, because the remaining time is too short to revise estimates and switch developmental trajectories. Future modeling might explore the evolution of sensitive periods when the time horizon is uncertain (e.g. in each time period, there is some fixed or increasing probability of extrinsic mortality; Mangel & Clark, 2019).

Third, models like ours are agnostic about mechanism. They exclusively consider the impact of experience on phenotype (Frankenhuis & Walasek, 2020). As a consequence, such models cannot be used to make predictions about the physiological processes that guide changes in plasticity across ontogeny. Nonetheless, models can help focus research efforts on hypotheses about mechanisms that produce the patterns generated by qualitative models that themselves do not incorporate mechanism.

Fourth, in models like ours the environmental state is typically the only unknown quantity and organisms know how to optimally respond to it. Organisms learn about the state of their environment based on cues and know the optimal developmental decision given their current state. However, real organisms might need to learn about other environmental quantities, such as the reliability of cues, or about the optimal adaptive response to different states. A large body of research on reversal learning shows that organisms are capable to infer the reliability of cues in nature (Izquierdo et al., 2017). Trout, for example, learn to recognize the sight or smell of potential predators (Behrens et al., 2007; Horn et al., 2019). When studying phenotypic plasticity in cases where the optimal response is known to the organism, we study so-called 'switch-like' plasticity (Frankenhuis, Panchanathan, et al., 2019; Snell-Rood, 2012). This captures a variety of traits and species. The development of defensive armor in Daphnia in response to chemical predator cues is a well-known example (Agrawal et al., 1999). However, in other cases organisms learn to respond adaptively via trial-and-error. For example, the circulatory, nervous, and immune systems are able to learn some adaptive responses from feedback (Snell-Rood, 2012). Natural selection has, in these cases, equipped organisms with the ability to learn the adaptive response based on trial-and-error (i.e. developmental selection) (Frankenhuis, Panchanathan, et al., 2019; Snell-Rood, 2012).

Future models of the evolution of sensitive periods might vary the environmental state within the lifetime of an organism, explore the consequences of a probabilistic time horizon on the cessation of ontogeny, or integrate known proximate mechanisms (Frankenhuis, Panchanathan, et al., 2019; Kacelnik, 2012; McNamara & Houston, 2009; Taborsky et al., 2021; Trimmer et al., 2012). Incorporating additional learning mechanisms, such as trial-and-error, represents another interesting future direction that might make models like ours more relevant to empirical studies of phenotypic plasticity. Specifically, organisms might learn about the reliability of cues or the adaptive response to different states. Reinforcement learning models are promising tools to approach both types of learning problems (Frankenhuis, Panchanathan, et al., 2019).

Conclusion

By showing that sensitive periods can be favored by natural selection beyond early life if the reliability of cues increases across ontogeny, our model contributes to a growing set of models exploring the selection pressures shaping the evolution of sensitive periods in development. Together, this family of models has the potential to develop into an integrative theoretical framework of the evolution and development of sensitive periods, which is firmly anchored in a classic and well-developed body of theory exploring the evolution of phenotypic plasticity more generally.

Chapter 4

Sensitive periods, but not critical periods, evolve in a fluctuating environment: a model of incremental development

This chapter is based on:

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2022) Sensitive periods, but not critical periods, evolve in a fluctuating environment: a model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1969), 20212623. https://doi.org/10.1098/rspb.2021.2623

Online supplements can be found alongside the published article (https://doi.org/10.6084/m9.figshare.c.5834303.v1)

4.0 Abstract

Sensitive periods, in which the impact of experience on phenotype is larger than in other periods, exist in all classes of organisms, yet little is known about their evolution. Recent mathematical modeling has explored the conditions in which natural selection favors sensitive periods. These models have assumed that the environment is stable across ontogeny or that organisms can develop phenotypes instantaneously at any age. Neither assumption generally holds. Here, we present a model in which organisms gradually tailor their phenotypes to an environment that fluctuates across ontogeny, while receiving cost-free, imperfect cues to the current environmental state. We vary the rate of environmental change, the reliability of cues, and the duration of adulthood relative to ontogeny. We use stochastic dynamic programming to compute optimal policies. From these policies, we simulate levels of plasticity across ontogeny and obtain mature phenotypes. Our results show that sensitive periods can occur at the onset, midway through, and even towards the end of ontogeny. In contrast to models assuming stable environments, organisms always retain residual plasticity late in ontogeny. We conclude that critical periods, after which plasticity is zero, are unlikely to be favored in environments that fluctuate across ontogeny.

4.1 Introduction

Phenotypic plasticity, the capacity of a single genotype to produce multiple phenotypes depending on environmental and somatic conditions, is widespread in nature (Nettle & Bateson, 2015; Stearns, 1989; West-Eberhard, 2003). There is well-established theory exploring the conditions in which phenotypic plasticity is favored by natural selection over non-plastic development. This work has provided valuable insights (for review, see Snell-Rood & Steck, 2019). For instance, plasticity is likely to be favored when the environment changes between generations at a rate too fast for genetic evolution to track, but slowly enough within generations for organisms to benefit from using early experience to guide phenotypic development (Botero et al., 2015; Snell-Rood & Steck, 2019; Stephens, 1991). However, this work has not focused on the question how natural selection shapes *changes in plasticity* across ontogeny for different species, individuals, and traits.

Modelling sensitive periods

Recently, mathematical models have been used to explore how natural selection shapes sensitive periods, i.e., time periods or life stages during which the impact of experience on phenotypic development is greater than at other times or stages (reviewed in Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017; Frankenhuis & Walasek, 2020). In these models, organisms typically begin ontogeny uncertain about the state of their environment and gradually reduce uncertainty by sampling environmental cues. As a result, plasticity is typically highest at the onset of ontogeny and gradually declines. These models have mainly focused on stable environments across ontogeny (English et al., 2016; Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Stamps & Krishnan, 2014a, 2014b, 2017; for an exception, see Fischer et al., 2014). However, many species and populations experience environmental fluctuations within generations as well (Botero et al., 2015). Little is known about optimal levels of plasticity across ontogeny in such conditions. For instance, when conditions fluctuate across ontogeny, plasticity may be prolonged to enable phenotypic adjustments across all of ontogeny (English et al., 2016; Panchanathan & Frankenhuis, 2016; Pascalis et al., 2020). Such a pattern would differ from that observed in models of stable environments, which often favor critical periods, where plasticity drops to zero (Pascalis et al., 2014, 2020).

Sensitive periods in fluctuating conditions

We know of only one model that has explored the evolution of sensitive periods in an ontogenetically fluctuating environment. Fischer et al. (2014) modelled an environment that fluctuates stochastically between two discrete states. Organisms develop initial phenotypes at the onset of ontogeny based on the long-term distribution of environments. In subsequent time periods, organisms sample imperfect cues to the current environmental state and adjust their phenotypes to maximize survival and reproduction across ontogeny. As with models of stable conditions, in this model plasticity declines with age. However, in contrast to those models, the highest level of plasticity ('peak-plasticity') does not always occur at the onset of ontogeny. In some conditions, organisms delay phenotypic adjustment until uncertainty has been sufficiently reduced, resulting in peak-plasticity shortly after the onset of ontogeny.

The model by Fischer et al. (2014) offered a crucial step forward but also has two limitations. First, it assumes that any phenotype can develop at any age within a single time period. This assumption does not apply when phenotypes are gradually constructed or cannot be reversed. In such cases, the current phenotypic state constrains the range of phenotypes available in the future. Second, the Fischer et al. model measures plasticity as phenotypic change directly following a cue. However, there are other possibilities that afford different insights (Frankenhuis & Walasek, 2020). For example, we can explore the effects of cues on developmental trajectories and mature phenotypes (Mueller, 2018), matching commonly used empirical designs (Haltigan et al., 2013; Humphreys et al., 2019).

Our contribution

Here, we present a model in which traits develop incrementally in an environment that fluctuates across ontogeny, exploring how cues shape plasticity across ontogeny and adult phenotypes. Organisms that gradually tailor phenotypes cannot instantaneously develop any phenotype at any time. Such incremental development is widespread in nature. For instance, plants gradually develop leaf morphology, such as area, thickness, and dissection, in response to light intensity, humidity, and temperature (Callahan et al., 1997; Maugarny-Calès & Laufs, 2018; Schlichting, 1986). Animals develop morphological defenses, such as protective armor, increased body size, or longer tails, as well as changes in coloration, in response to predator cues (Agrawal et al., 1999). In humans, the development of motor skills appears stepwise if measures are taken across weeks or months. However, this pattern reflects smaller incremental changes, which are visible once measures are taken frequently on shorter time scales (Adolph et al., 2008).

In our model, the environment varies stochastically between two discrete states. Organisms incrementally construct phenotypes while sampling cost-free, imperfect cues to the current conditions. Once phenotypic increments have developed, they cannot be undone. We use stochastic dynamic programming to compute optimal policies for a range of environments, varying the rate of environmental fluctuations, the reliability of cues, and how long adulthood lasts relative to ontogeny. These policies specify the optimal decision for each possible state of an organism, depending on its current phenotype and cues sampled. We then simulate populations of organisms following the optimal policy. Finally, we use experimental designs, matching those used in empirical studies, to quantify plasticity across ontogeny and distributions of mature phenotypes.

4.2 Model

Evolutionary ecology

The environment consists of an infinite number of discrete and nonoverlapping patches. Each patch can be in one of two states, E_0 or E_1. From one time period to the next, the state of each patch switches stochastically between E_0 and E_1 with transition probabilities, $P(E_{0,t}|E_{1,\,t\text{-}1})$ and $P(E_{1,t}|E_{0,\,t\text{-}1})$ where t denotes the current time period. For example, a patch might start out rich in one food type and switch to a different food type (e.g., seeds or fruits). We use a Markov process to fully describe the transitions between

states. As the per time period transition probabilities are fixed, we abbreviate $P(E_{0,t}|E_{1,\,t\text{-}1})$ and $P(E_{1,t}|E_{0,\,t\text{-}1})$ with $P(E_0|E_1)$ and $P(E_1|E_0)$.

We explore symmetric, $P(E_0|E_1) = P(E_1|E_0)$ and asymmetric transition probabilities, $P(E_0|E_1) \neq P(E_1|E_0)$. We also vary how likely transitions occur, ranging from 0.1 to 0.45 (positive autocorrelation). At the low end, the environment is relatively 'stable': an environmental switch is unlikely to occur. At the high end, the environment is 'unstable': a switch is almost as likely as no switch. We do not explore transition probabilities equal to or larger than 0.5 (negative autocorrelation).

Phenotypic development

Organisms are born, randomly disperse into a new patch, develop to maturity in the new patch, reproduce, and die. Ontogeny, T_{ont}, is fixed at 10 discrete and non-overlapping time periods. We obtain similar qualitative results for a larger number of time periods (Appendix 2 Figures A2.1–A2.4). We vary the length of adulthood (T_{adult} = 1, 5, and 20 time periods) to explore different ratios of adulthood to ontogeny (see Appendix 2, Figures A2.5–A2.9 for an adult lifespan of 10 time periods). Thus, time runs from $t = 0$ (birth) until the end of the reproductive phase T_{end}, such that $T_{end} = T_{ont} + T_{adult}$.

For each environmental state, there is a corresponding optimal phenotype: P_0 for E_0 (e.g., specialized for foraging seeds) and P_1 for E_1 (e.g., specialized for foraging fruits). These phenotypes represent two different traits, rather than a single trait that increases or decreases. In other words, phenotypes are not arrayed along a single dimension, but along two independent dimensions. Changes in one trait are independent of changes in the other trait. Organisms learn about their environment by receiving cost-free, imperfect cues. After each cue, organisms have three options: specialize one increment towards P_0, specialize one increment towards P_1, or wait and forgo specialization. Once an increment has developed, it cannot be undone, yet organisms may always switch developmental trajectories.

In adulthood, organisms experience the same transition probabilities as during ontogeny, but cannot adjust phenotypes. Instead, they accrue fitness depending on the phenotype-environment fit and reproduce proportional to fitness. In this model, fitness is only a function of fertility. We consider the effects of viability selection in the Discussion section.

Learning about the environment

The organism is adapted to the transition probabilities between states, as well as the long-term probability distribution over states (i.e., the stationary distribution of the Markov process), denoted by: $\pi(E_0) = \frac{P(E_0|E_1)}{P(E_0|E_1)+P(E_1|E_0)}$ and $\pi(E_1) = 1 - \pi(E_0)$. This distribution serves as an organism's evolutionary prior of being in each of the two environmental states at the onset of ontogeny (Stamps & Frankenhuis, 2016). If transitions towards the seed-rich state are more likely than transitions towards the fruit-rich state, i.e., $P(E_1|E_0) > P(E_0|E_1)$, the long-term probability of being in the seed-rich state is higher than that of being in the fruit-rich state, i.e., $\pi(E_1) > \pi(E_0)$. Symmetric transition probabilities produce a uniform stationary distribution, i.e., $\pi(E_0) = \pi(E_1) = 0.5$, while asymmetric transition probabilities produce a non-uniform stationary distribution, where $\pi(E_0) > \pi(E_1)$ if $P(E_0|E_1) > P(E_1|E_0)$.

In each time period, organisms sample a cost-free, imperfect cue to the current state of the environment and update their estimates according to Bayes' theorem (Dall et al., 2015; McNamara et al., 2006; Stamps & Krishnan, 2014a; Trimmer et al., 2011). The cue reliability is defined by the conditional probability of sampling the correct cue in the corresponding state, $P(C_0|E_0) = P(C_1|E_1)$. The probability of sampling an incorrect cue corresponds to $P(C_1|E_0) = 1 - P(C_0|E_0)$ and $P(C_0|E_1) = 1 - P(C_1|E_1)$, respectively. We vary the cue reliability from low (0.55) to high (0.95). The higher the cue reliability, the better organisms can adjust to the current state of the environment and exploit positive autocorrelation to adjust to future states of the environment.

Fitness during adulthood

Organisms who wait and forgo specialization across all of ontogeny attain a baseline fitness. Any developed specializations lead to increases or decreases from this baseline. During each time period in adulthood, fitness depends on the current phenotype-environment match. Total fitness corresponds to the sum of the fitness scores across adulthood.

We consider phenotypic specializations matching the environmental state as 'correct' and specializations towards the other state as 'incorrect'. We assume that correct phenotypic specializations increase fitness and incorrect ones decrease it relative to baseline fitness (Innes-Gold et al., 2019). The fitness in each period of adulthood is calculated by summing the marginal rewards for correct specializations, marginal penalties for incorrect ones, and baseline fitness. We explore three mappings between phenotypes and marginal fitness rewards and penalties (linear, increasing, and diminishing) and three penalty weights (0.5, 1, and 2) (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Walasek et al., 2021). The specific combination of mappings and penalty weight determines how organisms accrue fitness. Returning to our example of seed and fruit specialization, imagine the following two organisms: organism A has developed equal numbers of specializations for both states, while organism B has waited throughout ontogeny, developing zero specializations for either state. If we assume linear reward and penalty functions and a penalty weight of 1, then both organisms accrue zero fitness. If, instead, we assume a higher penalty weight or a diminishing penalty function, then, all else equal, A would attain lower fitness than B. With a lower penalty weight or an increasing penalty function, B does better than A. In the paper we set the penalty weight to 1 and the reward and penalty mappings to linear. We present the other combinations in Appendix 2 (Figures A2.16-A2.21) and address them in the Discussion section.

We describe fitness functions and formulas of all mappings in Appendix 2, 'Dynamic programming'.

Optimal developmental policies

To obtain optimal policies, we use the posterior estimates across ontogeny to compute expected fitness across adulthood. We treat the states of the environment during ontogeny as 'hidden', unobserved states and sampled cues as 'observed' states of a Hidden Markov Model. We then apply the forward algorithm to compute the posterior probabilities, $P(E_0|D_t)$ and $P(E_1|D_t)$ for all possible orderings of sampled cues D_t (Rabiner, 1989). $D_t = \{x_{1,} x_{2,} \ldots x_t\}$

denotes the sequence of cues until time period t, where x_1, x_2, and so forth until x_t denote the cue (C_0 or C_1) sampled in each time period. We provide the formulas of the forward algorithm in Appendix 2, 'Dynamic programming'.

In contrast to ontogeny, we model adulthood as a Markov Model with environmental states as observed states and no hidden states. The probabilities of starting adulthood in E_0 or E_1 equal the posteriors in the final time period of ontogeny. We compute the probabilities of being in each of the two states across adulthood, $P(E_0)$ and $P(E_1)$, based on these posteriors and the transition probabilities. Then we use $P(E_0)$ and $P(E_1)$ to compute expected fitness across adulthood (Appendix 2, 'Dynamic programming').

Finally, we compute optimal developmental policies using stochastic dynamic programming via backwards induction (Appendix 2, 'Dynamic programming'). The algorithm uses the posterior probabilities at the end of ontogeny and the expected fitness across adulthood as a starting point to determine the optimal decision in the final time period and then works its way backward in time. All code is written in Python 2.7. and available on GitHub (https://github.com/Nicole-Walasek/SensitivePeriodsInFluctuatingEnvironments).

4.3 Analysis

From transition probabilities to autocorrelation

Empirical studies often use temporal autocorrelation to measure environmental change. To facilitate comparisons between our model and such studies, we compute autocorrelation values from transition probabilities (Appendix 2, 'From transition probabilities to autocorrelations'). Higher transition probabilities produce lower autocorrelations.

The magnitude of the difference between $P(E_0|E_1)$ and $P(E_1|E_0)$, the 'asymmetry', also affects the autocorrelation. Suppose transitions to one state are more likely than to the other, for example, $P(E_0|E_1)$ = 0.1 and $P(E_1|E_0)$ = 0.2. In this case, the asymmetry is 0.1. If the patch starts in E_1, transitions are initially quite unlikely. However, once the state switches, the probability of another switch is higher. Overall, there would be more switches and lower autocorrelation compared to a scenario in which $P(E_0|E_1) = P(E_1|E_0)$ = 0.1. Higher asymmetries thus imply a smaller range of autocorrelations. Different sets of transition probabilities and asymmetries can approximate the same autocorrelation (see Appendix 2 Figure A2.10).

We have explored different asymmetries (i.e., 0.02, 0.05, 0.1, and 0.2, see Appendix 2, Figures A2.11 – A2.12). In the main text, we depict only autocorrelations characterized by an asymmetry of 0.1. This value can reveal the qualitative differences between symmetric and asymmetric transition probabilities, while still covering a large range of autocorrelations. Specifically, we set $P(E_0|E_1)$ - $P(E_1|E_0)$ = 0.1, so E_0 is the more likely environmental state. For both symmetric and asymmetric cases, we present results for values of $P(E_0|E_1)$ and $P(E_1|E_0)$ that approximate autocorrelations of 0.2, 0.5 and 0.8.

Quantifying plasticity

We simulate experimental designs resembling empirical adoption studies. These studies compare mature organisms, often twins or siblings, separated at a particular point during ontogeny for a specific duration. Researchers investigate how the age at which organisms are separated (and possibly later reunited), and the conditions during separation, determine variation in mature phenotypes. We have previously shown that different manipulations of experiences during separation – for instance, receiving reciprocal opposite cues or cues from a different patch, and temporary or permanent separations – yield similar qualitative patterns (Walasek et al., 2021). These patterns are most pronounced for reciprocal opposite cues and permanent separation, as experience is maximally divergent for a longer time. Therefore, we analyze only this manipulation here.

We use the optimal policy to simulate developmental trajectories. The level of plasticity corresponds to the extent to which phenotypic development depends on cues during ontogeny. We compute plasticity for each $t \in \{1, T_{ont}\}$. We start by simulating pairs of clones, following the optimal policy. Organisms start in either environment, E_0 or E_1. We simulate all possible sequences of cues, resulting in one pair of clones per sequence. Each pair of clones receives a weight according to the likelihood of its particular cue sequence.

Clones develop together until time period t, experiencing the same sequence of cues and making the same phenotypic decisions, resulting in identical phenotypes. At this point, the clones are separated, with one (the focal) remaining in the original patch and the other (the copy) developing in a mirror patch. The sequence of environmental states in the mirror patch is the same as in the original patch. However, the cues in the mirror patch are opposite those in the original patch. Whenever the focal individual samples C_0, the copy samples C_1, and vice versa. Focal-and-copy pairs continue development until maturation. Mature phenotypes are described by the number of time periods specialized towards P_0 and P_1. Together with the number of time periods waited, these numbers sum to T_{ont} = 10.

At the end of ontogeny, we compute the weighted average Euclidean distance between the two-dimensional phenotype vectors across all simulated pairs of clones. To control for the number of time periods the focal and copy have developed together, we normalize this measure by dividing the weighted average by the maximally attainable Euclidean distance, resulting in a range from 0 to 1. We show a schematic overview of our adoption study paradigm in Appendix 2 (Figure E2.13).

4.4 Results

First, we present the optimal phenotypic decisions across ontogeny. Next, we present the levels of plasticity resulting from these policies. We present the linear reward and linear penalty combinations (penalty weight of 1) below and all other combinations in SM 7 of the online supplements. Additionally, we show distributions of mature phenotypes and compare the terminal fitness of the optimal policies against two non-plastic strategies in SM 7 of the online supplements.

Optimal decisions across ontogeny

All organisms start out with the same estimate of the environmental state. These estimates diverge across ontogeny based on individual variation in sampled cues and then converge in adulthood towards the stationary distribution after learning stops (Appendix 2, Figure A2.13). While adult organisms no longer sample cues, their estimates of the environmental state continue to change, converging towards the stationary distribution. If adulthood is long enough, the estimates across individuals fully converge (Appendix 2, Figure A2.13).

With symmetric transition probabilities, the optimal developmental decision is to specialize toward the environment with the higher posterior in every time period (Figure 4.1 and Appendix 2, A2.14). This result follows from two facts. First, the stationary distribution implies that, on average, organisms will encounter each environmental state equally often. Second, organisms never change their estimates about which state is more likely during adulthood (Appendix 2, Figure A2.13). This means that, if an organism estimates that E_0 is the more likely state at the onset of adulthood, it will continue to do so regardless of the duration of adulthood. Taken together, organisms should specialize according to their posteriors regardless of the adult lifespan.

With asymmetric transition probabilities, the optimal decision depends on the relative length of adulthood, the cue reliability, and the autocorrelation between environmental states (Figure 4.1 and Appendix 2, A2.15). When adulthood is long relative to ontogeny (20 time periods) or the cue reliability is low (0.55), organisms always specialize towards E_0, the more likely state in the stationary distribution. With a long adult lifespan, organisms will more often encounter E_0 during adulthood and specialize accordingly. When cue reliability is poor, organisms remain uncertain about the environmental state when entering adulthood and so choose the more likely state (Figure 4.1). When adulthood is short relative to ontogeny (1 or 5 time periods), there is a high probability that the adult environment differs from the most likely state in the stationary distribution (Appendix 2, Figure A2.13). That is, a substantial proportion of organisms – though never, the majority – spends more time in E_1. The autocorrelation and cue reliability determine when during ontogeny, and with which posteriors, organisms start specializing toward the less likely state. The higher the autocorrelation, the sooner organisms start specializing towards E_1 because they can better anticipate the adult environment. With reliable cues (0.75 and 0.95) organisms achieve more extreme posterior estimates and are more likely to specialize based on their posteriors at the end of ontogeny (Figure 4.1 and Appendix 2, A2.15).

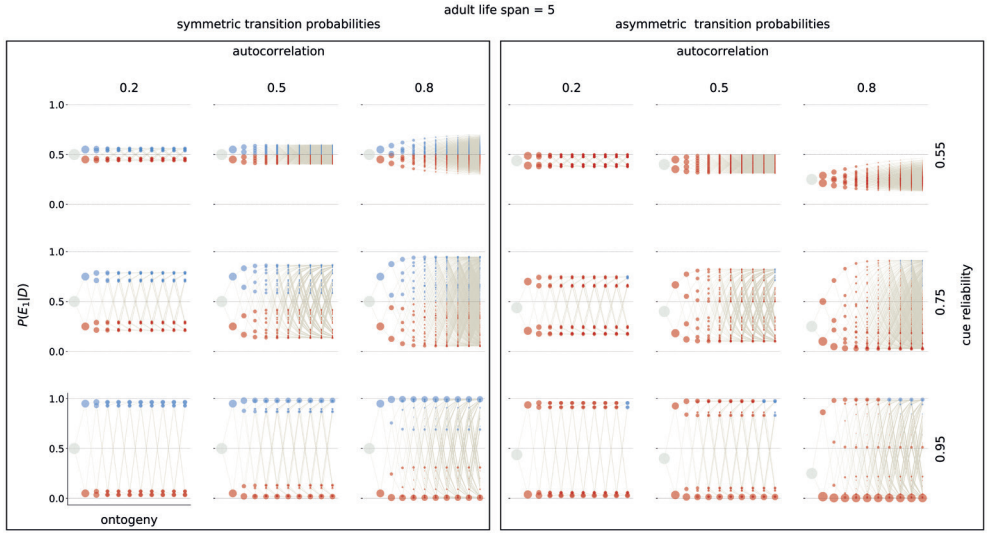


Figure 4.1 Optimal policies. Optimal policies are shown for linear rewards and linear penalties (penalty weight of 1), $T_{adult} = 5$, and symmetric (left panel) and asymmetric (right panel) transition probabilities. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior estimate of being in E_1 according to the stationary distribution (large grey circles). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the colored circles. Colors indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same posterior estimates (but different phenotypic states) make different phenotypic decisions. Black corresponds to waiting (not visible here because organisms never choose to wait), blue to specializing towards P_1, red to specializing towards P_0. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

Optimal levels of plasticity across ontogeny

Fixed, non-plastic policies are favored only under a narrow range of conditions. When plasticity is favored, it is retained until the end of ontogeny, though the timing of peak-plasticity varies. With low autocorrelation, plasticity peaks towards the end of ontogeny. With high autocorrelation, the timing of sensitive periods depends upon the cue reliability, with plasticity peaking at the onset, halfway through, or towards the end of ontogeny. We elaborate below.

Plasticity is not favored when one environment is more likely and adulthood is long or cues are unreliable.

All else equal, asymmetric transition probabilities reduce the scope for plasticity. After all, with one state more likely than the other, the organism faces less uncertainty and will rely less on plasticity and more on its prior. Asymmetric transition probabilities coupled together with long relative adulthoods result in low plasticity, or even no plasticity, across ontogeny (Figure 4.2). Longer adult lifespans allow organisms to rely on the stationary

distribution to adjust to their adult environment, reducing the need for plasticity. The stationary distribution implies that, on average, organisms will encounter the more likely environmental state more often than the less likely state (Appendix 2, Figure A2.13). Asymmetric transition probabilities coupled together with unreliable cues (0.55) favor zero plasticity across ontogeny (Figure 4.2). To avoid phenotype-environment mismatches due to unreliable cues, organisms use their priors at the onset of ontogeny to specialize towards the more likely environmental state. With symmetric transition probabilities, shorter adult lifespans, or reliable cues, plasticity is favored across ontogeny even when the environment fluctuates frequently (i.e., autocorrelation equals 0.2).

Environmental fluctuations favor sensitive but not critical periods.

Unlike previous models that assume stable environments (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Walasek et al., 2021), we find that 'critical periods' in which plasticity drops to zero are never favored. When the environment fluctuates, organisms always benefit from adjusting their phenotypes—even late into ontogeny. The exact level of plasticity at the end of ontogeny depends on the adult lifespan and the autocorrelation.

Short adult lifespans (1 time period) favor higher levels of plasticity at the end of ontogeny compared to longer adult lifespans (5 or 20 time periods) (Figure 4.2). When adulthood is short organisms rely on the most recent cues prior to the onset of adulthood. When adulthood is moderately long (5 time periods), organisms rely on a combination of recent cues and the prior (Appendix 2, Figures A2.14-A2.15). Only those organisms with highly certain posterior estimates specialize towards the less likely state. Those with less certainty specialize towards the more likely state in the stationary distribution. When the adult lifespan is long, natural selection favors non-plastic strategies.

Lower autocorrelations typically result in higher levels of plasticity at the end of ontogeny (Figure 4.2). The more frequent environmental fluctuations are, the more cues can shift posterior estimates throughout all of ontogeny, increasing the scope for plasticity (Figure 4.1). When the autocorrelation is high (0.8), organisms are less likely to attend to cues towards the end of ontogeny, resulting in lower levels of end-of-ontogeny plasticity. A relatively stable environment allows them to reduce uncertainty about their adult environment earlier during ontogeny. However, when cues are highly reliable (0.95) and the adult lifespan is short (1 time period), the chance of sampling incorrect cues is so low that the expected benefits from additional information about the environment outweigh potential mismatch costs. Under these conditions, end-of-ontogeny levels of plasticity can match those of ecologies with lower autocorrelations.

The timing of sensitive periods across ontogeny

Sensitive periods can evolve at the onset of ontogeny when environmental fluctuations are rare (autocorrelation 0.8). With symmetric transition probabilities and moderately reliable cues (0.75), organisms initially use cues to reduce uncertainty about their environment, resulting in a constant level of plasticity over large portions of ontogeny (Figure 4.2). However, plasticity declines towards the end as some organisms achieve more extreme posterior estimates, and consistently specialize towards one phenotypic target

(Figure 4.1). When cue reliability is low (0.55), natural selection favors constant, non-zero levels of plasticity across ontogeny. Neither the stationary distribution nor the sampled cues provide sufficient information to reduce uncertainty about the state of the environment. Organisms remain fairly uncertain, and thus attend to noisy cues across all of ontogeny.

Sensitive periods may evolve halfway through ontogeny when environmental fluctuations are rare (autocorrelation 0.8). With asymmetric transition probabilities and reliable cues (0.75 or 0.95), organisms specialize early in ontogeny according to the stationary distribution, ignoring sampled cues. As ontogeny proceeds, plasticity increases because organisms become more uncertain when they sample cues that contradict their posteriors. Plasticity peaks when organisms reach states after which they are likely to consistently specialize towards one phenotypic target, reducing the scope for plasticity in subsequent time periods. When cues are moderately reliable (0.75) and the adult lifespan is short (1 time period), organisms may reach such states halfway through ontogeny. Highly reliable cues (0.95) increase the probability that organisms start to specialize towards the less likely state already halfway through ontogeny (Figure 4.2), for both short and moderate adult lifespans (1 or 5 time periods).

Sensitive periods often evolve towards the end of ontogeny. Frequent environmental fluctuations favor sensitive periods towards the end of ontogeny. In such conditions (autocorrelations of 0.2 and 0.5), organisms specialize according to the most recent cues prior to the onset of adulthood (Figures 4.1-4.2). When environmental fluctuations are rare (autocorrelation of 0.8), plasticity sometimes peaks towards the end of ontogeny. When the adult lifespan is moderate (5 time periods) and cues are moderately reliable (0.75), a small proportion of the population specializes towards the less likely state in later time periods, resulting in sensitive periods towards the end of ontogeny. Plasticity may also peak at the end of ontogeny when the adult lifespan is short (1 time period) and cues are highly reliable (0.95), because organisms always choose to specialize according to cues in the final time period (Figures 4.1-4.2). These are also the only conditions in our model that favor two peaks in plasticity: one smaller peak halfway through ontogeny and one larger peak in the final time period. In the second half of ontogeny, plasticity decreases because many organisms are locked into developmental trajectories on which they consistently specialize towards the same state. However, to reduce mismatch penalties during a short adulthood, organisms always specialize according to the final cue as a form of insurance.

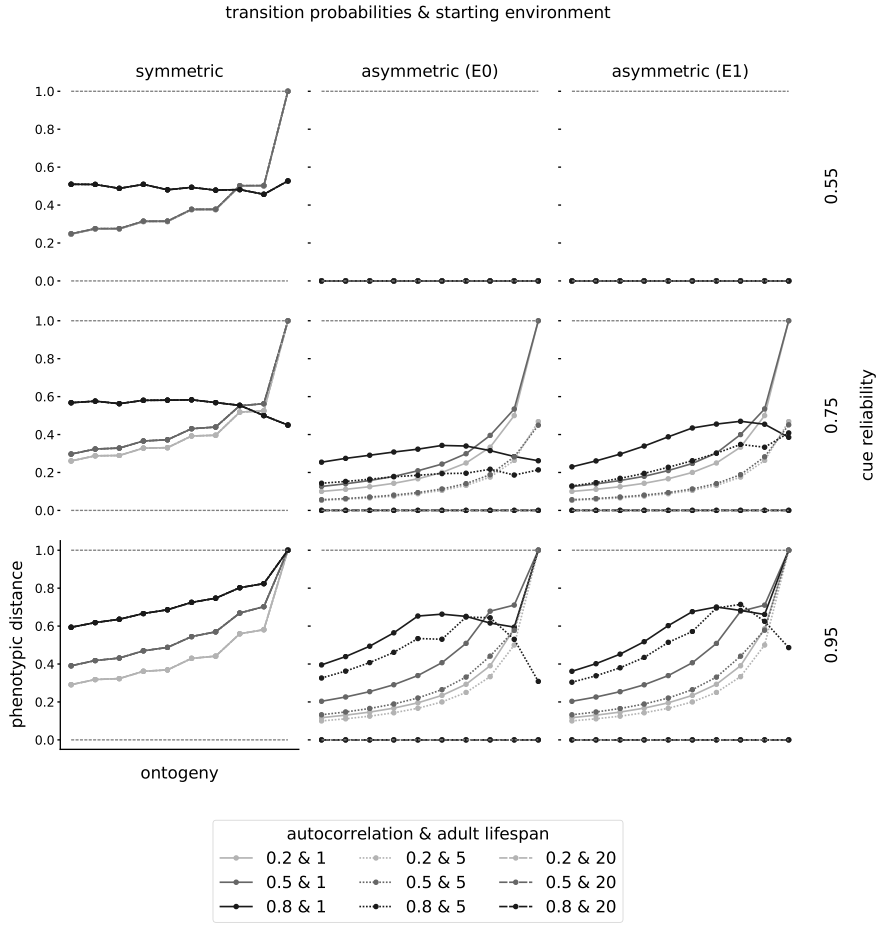


Figure 4.2 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties (penalty weight of 1). Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. Rows indicate different cue reliabilities. Within each panel, we show separate lines for different levels of autocorrelation (indicated by the greyscale) and different adult lifespans (indicated by the line type). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct $T_{ont} = 10$ experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time period t during ontogeny (horizontal axis). We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of separation. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

4.5 Discussion

Sensitive periods are more likely to evolve than critical periods

Unlike in models assuming stable environments, we find that critical periods, in which plasticity drops to zero, are never favored. In a fluctuating environment, organisms always use cues at the end of ontogeny to reduce uncertainty about their adult environment. Combining insights across models, we may expect empirical researchers to observe critical periods for traits that have evolved in stable ontogenetic environments and sensitive periods for traits that have evolved in fluctuating ontogenetic environments.

Our finding that plasticity always persists at the end of ontogeny is striking for two reasons. First, previous work shows that the fitness costs of plasticity may outweigh the fitness benefits when organisms need to continuously readjust to fluctuating environmental conditions and pay the associated costs (Leung et al., 2020; Pfab et al., 2016; Snell-Rood & Steck, 2019). Our model assumes no costs to building, maintaining, and running the physiological machinery for plasticity; to sampling cues; and to making phenotypic adjustments. Our model does, however, assume that plasticity is incremental and irreversible, and that there is a cost to phenotype-environment mismatch in adulthood. With these assumptions, the level of plasticity at the end of ontogeny is highest when adulthood is short and the rate of environmental fluctuations is high. Second, previous models exploring the evolution of plasticity in fluctuating environments have assumed that developed phenotypes can be undone, allowing organisms to continuously readjust their phenotypes to changing conditions (Pfab et al., 2016; Utz et al., 2014). The ability to reverse development may reduce phenotype-environment mismatch and thus make plasticity across all of ontogeny more viable. In our model, organisms cannot reverse phenotypic increments. Developmental trajectories, however, are reversible, such that organisms may specialize towards the opposite phenotypic target at any point during ontogeny. This allows phenotypic plasticity to be favored and even persist until the end of ontogeny when the environment fluctuates frequently.

A study by Relyea (2003) has shown that tadpoles (*Hyla versicolor*) – which cannot undo developed phenotypes but are able to switch developmental trajectories – retain plasticity towards the end of ontogeny in a fluctuating environment. Tadpoles were exposed to variation in predation risk across ontogeny and showed the induction of morphological defenses, such as greater mass, deeper tails, or shorter bodies, throughout all of ontogeny, albeit to a lesser extent at later stages. Relyea did not operationalize reversibility as the deconstruction of phenotypic adjustments, but as the relative readjustment of different morphological features. For example, if a tadpole developed a deeper tail relative to its body size in response to predators and increased body size after predators were later removed, this counted as a reversal. Such reversals are similar to the reversibility of specialization trajectories in our model. The study showed that reversibility of phenotypic inductions was high early in ontogeny and lower later during ontogeny. Our model predicts a decline in organisms' ability to switch between trajectories as ontogeny progresses, when environmental fluctuations are rare and cues are moderately reliable. In such conditions, the remaining time is too short for organisms to revise estimates. Relyea allowed for a switch in predation risk only once during ontogeny. These conditions resemble those of high

autocorrelation values in our model, more so than those of low autocorrelation. However, we do not know the reliability of predation cues used in the study. Future research could replicate the experiment while manipulating the reliability of cues. Our model predicts a steep increase in plasticity at the end of ontogeny when cues are highly reliable.

The timing of sensitive periods

We find that sensitive periods typically occur at the end of ontogeny. This finding contrasts with results from models of stable environments (English et al., 2016; Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016; Stamps & Krishnan, 2014a, 2014b, 2017), as well as the results of one model exploring fluctuating environments (Fischer et al., 2014). Our finding indicates that natural selection may heighten sensitivity to cues towards the end of ontogeny when the environment changes rapidly and phenotypes develop incrementally. Developing organisms then use experiences towards the end of ontogeny to adjust phenotypes right before maturity. This makes sense: when the environment fluctuates frequently, cues towards the end of ontogeny tend to better predict conditions in adulthood than earlier cues. Responses to cues can be behavioral or morphological. Examples of greater reliance on cues later during development exist for both. For example, migratory bird, bat, and fish species use cues throughout their journey to predict remote conditions and adjust their arrival time and destination (Winkler et al., 2014). Often, these animals rely the most on cues towards the end of their journey to make such predictions. In bulb mites (*Rhizoglyphus robini*) nutritional conditions during the final ontogenetic stages determine whether males mature as 'fighters' or 'scramblers' (Smallegange, 2011). The extent to which these patterns depend on the rate of environmental change, the reliability of cues, or the duration of adulthood relative to ontogeny remains to be explored. Experimental evolution studies of bulb mites and other insect species can be used to explore different parameter combinations and test predictions from models like ours (English & Barreaux, 2020).

While sensitive periods often emerge at the end of ontogeny in a fluctuating environment, they may occur midway through ontogeny when autocorrelation is high. Models of stable environments have obtained this same result (Stamps & Krishnan, 2017; Walasek et al., 2021), and so did Fischer et al.'s model of fluctuating environments (Fischer et al., 2014). These sets of models produce this pattern, at least in part, for the same reason: the initial discrepancy between posteriors and cues increases uncertainty, which increases plasticity early in ontogeny, before plasticity declines when later cues reduce uncertainty. Our model also produces sensitive periods midway through ontogeny, but for different reasons than those models. Fischer et al. (2014) and Stamps & Krishnan (2017) assume that organisms start development with specialized phenotypes and that specializations are fully reversible. Under these conditions, sensitive periods may arise midway through ontogeny because organisms first sample multiple cues to reduce uncertainty, before changing their specialized phenotypes. This effect may be strongest when phenotypic adjustments are costly, as in Fischer et al.'s (2014) model. Complete reversibility further allows organisms to delay developing specializations because the scope for phenotypic adjustment is not constrained by the duration of ontogeny. In the current model, and in a previous model of stable environments with variation in the cue reliability across ontogeny (Walasek et al., 2021), sensitive periods midway through ontogeny are favored even if organisms do not start out with specialized phenotypes that are costly to switch away from and adjustments

are irreversible. Thus, sensitive periods midway through ontogeny may evolve across a range of environments and life histories.

Long adult lifespans disfavor plasticity

When adulthood is short relative to ontogeny, high levels of plasticity are favored across ontogeny. By contrast, when adulthood is long, organisms rely less on (or ignore) their experiences and specialize towards the more likely state in the stationary distribution. These findings may appear at odds with those of a co-evolutionary model by Ratikainen & Kokko (Ratikainen & Kokko, 2019) showing that longevity favors plasticity and vice versa. However, this difference can be understood in light of assumptions about plasticity in adulthood. Our model assumes that phenotypic development is limited to ontogeny. Their model allows adults to continue tailoring their phenotypes. Thus, in their model (but not in ours), long-lived organisms can do better than adapting to the stationary distribution. Combining both models, we may predict that longevity is associated with higher levels of plasticity when adult phenotypes are malleable, and with lower levels of plasticity when adult phenotypes are fixed. Cross-species comparisons have shown that higher levels of plasticity are associated with longer lifespans in some groups of animals (Gopnik, 2020; Sol et al., 2016), but with shorter lifespans in other groups of animals (Sowersby et al., 2021). Future work may test whether the malleability of adult phenotypes moderates these opposite patterns of association.

Limitations and future directions

Our model assumes only two environmental states. Another possibility would be to assume a larger number of discrete states or a continuum of states. This would make it possible to independently manipulate the means and variances of both the prior distribution and the reliability of cues. Doing so might influence the findings from our model. However, previous models of stable environments that incorporate a continuum of environmental states (Stamps & Krishnan, 2014b, 2017) have found similar qualitative patterns as those assuming two discrete states (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016). Future modelling could explore whether our results replicate when increasing the number environmental states.

In our model, fitness is only a function of fertility. Other state-dependent models assume that fitness depends on fertility and mortality (Fischer et al., 2014; Houston et al., 1988). Our model could be extended to include mortality. Mortality would be a function of phenotype-environment match during ontogeny, adulthood, or both, depending on how the trait influences mortality across these stages.

In our model, fitness is proportional to the difference between correct and incorrect specializations. We instantiate this through specific reward-penalty mappings and penalty weights. In the main text, we set the penalty weight to 1, implying that rewards and penalties contribute equally to fitness. In Appendix 2, we show that penalty weights of 0.5 and 2 yield the same qualitative patterns, but there is one difference: when the penalty weight is 2, organisms sometimes wait to reduce phenotype-environment mismatch. Though we have explored a wide parameter range, future work could investigate a more general model where fitness is an arbitrary function of phenotype.

Our model assumes that organisms 'know' (i.e., have evolutionarily adapted to) the cue reliability, autocorrelation, and the durations of ontogeny and adulthood, because these parameters were fixed across generations. However, if these parameters were variable, organisms may estimate them based on experience. Future modelling could explore a scenario in which the cue reliability, autocorrelation, and the duration of life stages vary between generations, but are stable within generations. For instance, an organism might be born into one of several patches, each of which has its own cue reliability, autocorrelation, or duration of life stages. Future modelling could also explore a scenario in which these parameters vary within generations as well. For instance, the weather might change at different rates in different seasons. Under these conditions, organisms may need to learn the pattern of change of environmental parameters across their lifespan (Frankenhuis, Panchanathan, et al., 2019). In experimental studies, humans, non-human primates, and rodents are able to learn the cue reliability and adaptively adjust their behavior (Behrens et al., 2007; Izquierdo et al., 2017). It would be interesting to see whether organisms that are uncertain about multiple parameters retain higher end-of-ontogeny levels of plasticity, as we see in the current model. Organisms may develop sensitive periods late in ontogeny, if conditions favor attaining confident estimates of environmental parameters prior to committing to phenotypic specialization.

As noted, animals and plants may experience fluctuations in different environmental statistics during their lifetimes (Grosbois et al., 2008). For example, the reliability of cues varies across ontogeny for a variety of aquatic species, such as larval mosquitos (*Culex restuans*), common roaches (*Rutilus rutilus*), fathead minnows (*Pimephales promelas*), and goldfish (*Carassius auratus*) (Ferrari et al., 2010). However, for many species and traits, there is little information about the values of environmental statistics across ontogeny (Frankenhuis, Nettle, et al., 2019). As a future direction, we envision a repository of environmental statistics across ontogeny (e.g., autocorrelation, cue reliability) for a range of species and populations. Such a repository can benefit empirical researchers who study how environmental conditions shape development, as well as theoreticians modelling the evolution of developmental phenomena, such as sensitive and critical periods. For instance, it would allow modelers to make informed decisions about which parameters to fix or vary across ontogeny, depending on their research questions about groups of organisms (e.g., taxa, clades) or particular species or populations. Modelers and empirical researchers may use the repository to focus on those rates of variation that are most relevant for a given taxonomic group or species when developing theory and experiments. In this way, a repository of environmental statistics has the potential to strengthen connections and create synergies between empirical and theoretical work, thus accelerating progress in our understanding of the evolution and development of sensitive periods.

Chapter 5

A framework for studying environmental statistics in developmental science

This chapter is based on:

Walasek, N., Young, E. S., & Frankenhuis, W. E. (2022) A framework for studying environmental statistics in developmental science. *Manuscript in preparation*

5.0 Abstract

People are shaped through interactions with environments. Therefore, studying environmental stability and change are key to developmental science. The current norm is to describe notions of environmental stability and change (e.g., variability, unpredictability, instability) using natural language, which tends to be ambiguous. This ambiguity weakens the match between theory and methods (e.g., between constructs and measures, research questions and analyses) within studies, and leads to inconsistencies across studies. This puts developmental studies in disarray: when researchers speak different languages, they learn less from each other, which in turn impedes cumulative science and interdisciplinary integration. The field needs a shared framework that organizes notions of environmental stability and change in unambiguous, formal terms. Here, we present such a framework. Although our framework is novel, it draws on statistical definitions of environmental stability and change that are already widely used in the other disciplines, such as biology, ecology, and economics. To demonstrate feasibility, we apply our framework to a dataset of crime rates in New York City across 15 years, focusing on 'unpredictability' as a case study. We explore different ways to statistically quantify unpredictability, for example, by using the autocorrelation, entropy, or the number of changepoints. We find that some results generalize across statistical definitions, and others depend on which statistical definitions are used. For instance, regions rank similarly in terms of unpredictability based on different statistical definitions, and the associations between unpredictability and poverty across regions are also similar. However, if we simulate residential changes, individuals rank differently in terms of their exposures to unpredictability. This matters for psychological research focusing on individual outcomes: it means that different notions lead to different conclusions about the impact of unpredictability on key life outcomes such as health, wellbeing, and psychopathology. This case study illustrates the merit of having a shared framework that creates a structured space for a cumulative science of environmental stability and change. To facilitate its use, we include accessible, step-by-step guidelines for applying our framework to developmental datasets.

5.1 Introduction

Environmental stability and change are both central to developmental science. Most developmental research assumes, claims, or examines some notions of environmental stability and change. For example, researchers study the effect of *stability* in parental warmth on children's socio-emotional adjustment, or how unpredictable environmental *changes* shape health and cognition. Stable and high parental warmth are associated with better adjustment; high unpredictability with worse health. However, limited attention is given to defining stability and change in constructs, such as parental warmth or unpredictability, in unambiguous ways. As a result, constructs are often loosely connected to measures and data analyses approaches are inconsistent across studies. Here, we argue, that to achieve better integration of research, we need to express notions of environmental stability and change in unambiguous, formal terms; that is, as statistical definitions.

To illustrate, suppose a researcher is interested in the effects of parental warmth on emotional adjustment. Parental warmth is assessed by measuring parents' responsiveness and supportive behaviors during a stressful task (Luby et al., 2012, 2016). A researcher who is interested in stability might compute the average across measurements. However, a simple average might suffice in some cases but not others. If the researcher cares about the overall level of warmth across measures, the average may suffice. If the researcher wants to capture not only the overall level of warmth scores but also their consistency, the average alone is not enough. In this case, the researcher should also quantify how variable warmth scores are across time. Low variability implies similar warmth scores over time and, therefore, consistency. Thus, to answer the research question it is necessary to consider the distribution of parental warmth and not just the average.

The need for statistical definitions comes into sharp focus when talking about change. Much developmental research is aimed at understanding the challenges that environmental change poses to individuals. To identify these challenges researchers study how environmental change (e.g., unpredictability) shapes developmental outcomes (e.g., health and cognition). Different types of change may result in different developmental outcomes. This is important both for understanding functional responses and maladaptive responses. Suppose a researcher is interested in whether impulsivity is adaptive in unpredictable environments. Theoretical modeling has found that this depends on the definition of unpredictability. Impulsivity may be adaptive when the collection of resources is often interrupted ('collection risk') but not when resource quality is highly variable (Fenneman & Frankenhuis, 2020). To test empirical predictions derived from this model, it is necessary to explore these different notions of unpredictability and to develop appropriate statistical definitions. Similarly, another researcher may be interested in the effects of environmental variability on health. She expects that unpredictable changes in environmental conditions, but not predictable ones, deteriorate health. A widely accepted definition of unpredictability is random variation in harshness across time (Ellis et al., 2009). However, such a definition is consistent with multiple statistical formalizations of change over time. For example, one researcher might compute variance in harshness across the measurement period (e.g., Li et al., 2018). Another might track abrupt shifts in the mean or variance in harshness (changepoints; described in Young et al., 2020), and a third one might

measure how well current levels of harshness predict future levels (i.e., autocorrelation; see, for example, Burgess & Marshall, 2014; Marshall & Burgess, 2015). But which statistical definition best captures 'random variation in harshness'? Do all statistical definitions of unpredictability predict the same developmental outcomes?

The need for a framework

Clear definitions of stability and change create a structured space for science to progress. Statistical definitions offer an unambiguous way to define these concepts. That is, these definitions serve as the building blocks for exploring and testing questions related to stability and change. In some cases, our hypotheses and constructs are explicit enough to narrow the range of possible building blocks. In other cases, researchers may use different building blocks to study the same construct or hypothesis. However, as is often the case, they do so because the ambiguity of natural language allows different interpretations (Frankenhuis & Walasek, 2020). It is, of course, fine if researchers employ different statistical definitions of the same construct in the literature, as long as their definitions are explicit. But there are costs. Researchers may not be aware that they are adopting one view over the other or realize they unintentionally study a different construct or question than their colleagues. Similarly, readers of the resulting literature likely focus on the studied constructs and may not notice inconsistencies in statistical definitions across studies.

The lack of formal statistical definitions of stability and change can also invite flexibility in measurement. A recent review identified 15 different measures of unpredictability in 21 empirical studies (Young et al., 2020). Unpredictability is measured as the number of household moves, family disruptions, or changes in parental financial status. These measures are used as proxies of 'stochastic variation in harshness', even though it is not obvious how they relate to possible statistical definitions. However, there exist studies which use measures and statistical definitions that come closer to this conceptual definition of unpredictability. For example, Li et al. (2018) measure harshness through socioeconomic status and estimate linear slopes in harshness for each individual. They quantify unpredictability as the residual variance around these slopes. What this statistical notion of unpredictability does not capture is whether variation in harshness is predictable or unpredictable (Young et al., 2020). For instance, the residual variance may be correlated across time or not. If it is, this increases predictability (see Table 5.1).

Clear and unambiguous statistical definitions narrow which measures are valid and limit the use of questionable proxies. Inconsistent and vague statistical definitions and measures weaken the match between theory and methods (e.g., between constructs and measures, research questions and analyses) within studies, and lead to inconsistencies across studies. This puts developmental studies in disarray. But, before we can compare and evaluate different statistical definitions and use them to inform our measures, we need to organize existing definitions. To this end, we present a framework that organizes statistical definitions of environmental stability and change (Table 5.1).

Our framework

Building on existing work in the social and biological sciences, we integrate familiar approaches with elements borrowed from other fields. Our framework provides clear

definitions of 'environmental statistics' for repeated measures of an environmental variable, such as harshness or parental warmth. For each individual, we extract various measures of stability and change with different statistical definitions, which can be included as predictors in any analysis.

Our framework focuses on stability and change. *Stability statistics* describe the overall level and distribution of an individual's data, such as its mean and spread. *Change statistics* describe patterns in data across time. There are two broad classes of change statistics: *predictable changes* and *unpredictable changes* (Jebb et al., 2015; Ram & Gerstorf, 2009). For example, a linear trend describes predictable changes in the data (constant increase or decrease), whereas changepoints describe abrupt shifts in the mean, variance, or both. Often, multiple statistical definitions are plausible for different constructs invoking stability or change. The appropriate statistic depends on the specific research question, hypothesis, or theory. For example, a researcher might be interested in how exposure to crime in her everyday environment shapes health and cognition. If she is interested in the overall level of crime exposure, she might choose the mean. If, instead, she is interested in how changes in crime shape health, she might estimate the linear slope in the data. Hypotheses may also relate to variability in exposure to crimes. Sudden changes in variability in crime rates can indicate a highly unpredictable environment. One way to quantify such unpredictable changes is to estimate the number of changepoints in the mean or variance. We describe a range of stability and change statistics that are part of our framework in Table 5.1.

Within individuals, we can quantify the statistical properties of an individual's environment, experiences, or exposures. Across individuals, we can explore how various indices of within-person stability and change are distributed in an entire sample. This allows us to assess how different environmental statistics relate to each other in a specific sample. For example, we may find that environmental statistics indicating high unpredictability in samples of individuals from large cities look different in samples from rural areas. Our framework is designed to include a selection of environmental statistics particularly relevant to development science; it is no exhaustive collection of all possible statistics.

The benefits of a framework

Our framework has four major benefits. First, the framework increases conceptual clarity by highlighting the different ways in which patterns of environmental stability and change, such as unpredictability, may be defined and computed. Second, the framework provides guidance by offering tools to explore and compute these statistical definitions. Third, the frequent application of the same statistics can foster integration and comparability of findings across different studies. Fourth, the framework may also help to integrate different types of data.

Developmental research often uses individual-level data, parsing the environment through the individual's lens. Such data may include subjective measures through self-reports, questionnaires, or interviews. Or they may include objective measures through scores on a standardized task, the number of specific life events (e.g., household moves or bereavement experiences), socioeconomic data (e.g., income or educational attainment), or observations in the laboratory. Subjective measures of individuals' experiences are central in

clinical research using complex systems theory to study well-being (e.g., Olthof et al., 2020). Our framework complements such work by focusing on stability and change of the external (rather than internal) environment. Publicly available datasets can provide insights into an individual's surrounding ecology, such as their home, work and school environment (Kievit et al., 2021). Such environment-level data typically offer many more repeated measures than is typical in developmental studies using individual-level data. Datasets with information about crime, violence, prevalence of disease, or access to educational and healthcare facilities are readily available and can be feasibly linked to individual data. Although, such data are still underrepresented in developmental studies, few examples exist (Hatzenbuehler et al., 2021; Miller et al., 2018; Snyder et al., 2011). For example, Hatzenbuehler et al. (2021) incorporated environment-level measures of stigma (e.g., institutional policies) alongside individual-level self-reports (e.g., experienced discrimination) in a recent study exploring the association between stigma and brain volume in Black and Latinx youth. Our framework can facilitate the computation of stability and change statistics from such environment-level data alongside individual-level data. Studies that bridge both types of data would allow us to assess their individual contributions to development.

Environmental unpredictability: a case study

We use the case of environmental unpredictability to illustrate the four benefits of our framework. A framework of environmental statistics provides tools to compute a range of unpredictability statistics. It can help distinguish between competing statistical definitions of unpredictability and increase comparability across studies that use the same statistics. The framework can also enrich existing measures by facilitating the linking between individual-level and environment-level data. For example, in addition to the number of household moves as a measure of unpredictability, we could integrate environmental data for each location an individual has lived in. From such data we can extract unpredictability statistics, resulting in a higher resolution of the experienced environment.

Statistic	Quantifies	Definition	Interpretation
Mean	Stability	Arithmetic mean	Average harshness levels across measurement period
Standard deviation	Stability	Standard deviation	Average fluctuations around mean level of harshness
Minimum	Stability	Lowest value of environmental variable	Lowest harshness level during measurement period
Maximum	Stability	Highest value of environmental variable	Highest harshness level during measurement period
Interquartile range (IQR)	Stability	Range of the middle 50% of the environmental variable	Range of the most common harshness levels; a large IQR indicates a large range of harshness levels.
Slope (linear model)	Predictable change	Linear association between time and environmental variable	Indicates the trend of harshness across time; a positive slope indicates an increase in harshness across time and a negative slope a decrease.
Period	Predictable change	Length of a cycle in the environmental variable, if any is present	Presence of a cycle indicates that similar harshness values occur every cycle.

Statistic	Quantifies	Definition	Interpretation
Autocorrelation	Unpredictable change	The extent to which current values correlate with future values of the environmental variable	High absolute autocorrelation indicates that current harshness levels are predictive of future harshness levels.
Partial autocorrelation	Unpredictable change	The extent to which current values correlate with future values of the environmental variable correcting for the correlation with intermediate values	Absolute partial autocorrelation indicates the extent to which current harshness levels are predictive of future harshness levels without considering intermediate harshness values.
Entropy	Unpredictable change	Approximate entropy for time series; the extent to which environmental fluctuations are irregular	Low entropy indicates that harshness levels across the measurement period fluctuate regularly and predictably.
Spectral coefficient	Unpredictable change	Describes patterns of noise in the environmental variable and is sometimes called color of noise	Indicates the extent to which noise in harshness is predictable across time. Noise can change randomly across time (white noise, spectral coefficient around 0). It can change slowly, resulting in long runs of above or below average conditions (red and brown noise, coefficient between 1 and 2). Or, it can change rapidly but predictably (blue noise, coefficient below 0).
Number of changepoints in mean	Unpredictable change	Number of times the mean of the environmental variable changes	Indicates how often the average harshness level changes across the measurement period. We can look separately at positive or negative changes in mean.
Number of changepoints in variance	Unpredictable change	Number of times the variance of the environmental variable changes	Indicates how often variance in harshness levels changes across the measurement period. We can look separately at positive or negative changes in variance.
Average time between changepoints	Unpredictable change	Average time between any two changepoints	A short average time between changepoints indicates that changes in harshness (mean and/or variance) occur at a high frequency.
Standard deviation of time between changepoints	Unpredictable change	Standard deviation of the time between changepoints	Indicates variability in the time between changepoints. High variability indicates that changepoints can occur abruptly.
Longest period without changepoints	Unpredictable change	Longest period without changepoints	Indicates the duration of the longest period with stable mean and/or variance in harshness.

Table 5.1 Glossary of environmental statistics. Environmental statistics are used to quantify stability and change. The latter is further divided into predictable and unpredictable change. In the fourth column, we provide an interpretation of each statistic using harshness as an example.

In this case study, we conceptualize unpredictability as random variation in harshness across space, time, or both (Ellis et al., 2009). Harsh environments are characterized by greater risks of disability and death (Brumbach et al., 2009; Ellis et al., 2009). In non-human

animals, resource scarcity and predator density are indicators of harshness. In humans, poverty and crime rates are often used to measure harshness (Brumbach et al., 2009; Young et al., 2020). We illustrate our framework using existing, publicly available crime records in New York City (USA) as indices of harsh environmental conditions. The data span 15 years from January 2006 until December 2020. For different regions in NYC, we present a range of statistical definitions of unpredictability.

We use six different statistical definitions to quantify unpredictability as 'random variation in harshness': the standard deviation, changepoints in mean, changepoints in variance, autocorrelation, entropy, and color of noise. The standard deviation describes the average deviation from the mean in harshness. In Table 5.1 we have referred to the standard deviation as a stability statistic because it summarizes the distribution of the data and not how they change across time. To illustrate, suppose that we compute the standard deviation of an individual's time series. The resulting value tells us something about how the data fluctuate around the mean. However, if we would randomly shuffle the data points, the resulting standard deviation would not change. The standard deviation does not consider the order of individual data points across time. Our other unpredictability statistics do take the order into account. Nonetheless, we included the standard deviation as an unpredictability statistic because previous work has used it to quantify unpredictability (e.g., Li et al., 2018).

Changepoints describe abrupt shifts in the mean or variance in harshness (Haynes et al., 2016; Killick & Eckley, 2014; Young et al., 2020). The autocorrelation indicates how much current harshness values predict future values. Suppose we have collected monthly measures of harshness across one year. The autocorrelation then corresponds to the correlation between these data and the same data shifted by one month (i.e., by one time unit) (Burgess & Marshall, 2014; Marshall & Burgess, 2015). This is called a lag-1 autocorrelation. Lag-2 autocorrelation would shift the data by 2 months. Entropy of a time series quantifies the extent to which harshness values change regularly or irregularly (Pincus, 1991; Richman & Moorman, 2000). For example, a time series that consists of only two alternating harshness values is perfectly regular, resulting in low entropy. If changes between those values occur randomly, entropy would be high, indicating high unpredictability. Color of noise indicates the extent to which noise in harshness is predictable across time (Burgess & Marshall, 2014; Marshall & Burgess, 2015; Ruokolainen et al., 2009; Vasseur & Yodzis, 2004). Noise is what is left of the data after subtracting systematic patterns, such as trend and season. Noise can change randomly across time (white noise, color of noise around 0). It can change slowly, resulting in long runs of above or below average conditions (red and brown noise, color of noise between 1 and 2). Or, it can change rapidly but predictably (blue noise, negative color of noise). Autocorrelation, entropy, and color of noise tackle quantifying unpredictability in slightly different ways. We refer the reader to the respective references for technical details on each of the approaches.

Using simulations, we also illustrate the potential for linking environment-level data to individual-level data. Based on the crime data, we simulate a sample of individuals who moved across different regions in NYC during the measurement period. For these simulated data, we present a range of stability and change statistics describing per-individual exposure

to crime rates across time. All our code, including the simulated data, is available at https://github.com/Nicole-Walasek/environmental_statistics/.

5.2 Methods

For our case study of unpredictability, we first apply our framework to different regions in NYC. This allows us to compute and compare a range of unpredictability statistics within and across regions. Second, we apply our framework to a simulated dataset of individuals.

Public datasets can span years or decades and often provide daily measurements. They typically record crimes for specific regions within a city or country. However, developmental researchers are interested in environmental stability and change across an individuals' lifespan. Therefore, we use the crime data to simulate individuals moving between different regions within NYC. In principle, such data could be obtained by linking people's residential history to public crime records of each residential location (see Hatzenbuehler et al., 2021; Miller et al., 2018; Snyder et al., 2011 for examples).

From here on, we first describe the dataset (section 'New York crime data'), then explain how we identified regions of interest (section 'Regions of interest'), and finally use those data to simulate a sample of individuals (section 'Simulating individuals').

New York crime data

The dataset is part of the *NYC Open Data project*[1]. This initiative provides open access to the information that is available to the New York City government. The database contains data related to business, governance, education, environment, safety, and health. We will focus on the *NYPD Arrests Data (Historic)*[2]. This dataset records every arrest by the NYPD dating back to 2006 and is still being updated every quarter. Each entry holds information about the type of crime, the arrest location, and the time of arrest. We only use information about crimes related to assault resulting in 644,684 entries across 15 years, from January 2006 until December 2020. Table A3.1 in Appendix 3 lists all offenses that we included under 'assault' and their description provided by the police department.

Regions of interest

Imagine a person living in NYC between 2006 and 2021. We make two assumptions for our simulation. First, she likely spends most of her time (on average) within a 5 km distance from her home. Second, over those 15 years, she may change residency between different regions within the city, where she might experience varying levels of violent crimes. These crimes occur more frequently in some regions than others.

We simulated a sample of individuals who may move to different regions of NYC between 2006 and 2020 and measured their exposure to crime across time and space (i.e.,

1 https://opendata.cityofnewyork.us/

2 https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

different regions within NYC). To do so, we first identified regions of interest where our simulated individuals might live. We picked regions in NYC that vary in crime rates. In addition, we used reports from the *New York City Institute for Children, Poverty, and Homelessness*[3] to identify regions that vary in wealth, because crime rates tend to be positively associated with poverty (Barone & Mocetti, 2016; De Courson & Nettle, 2021; Rufrancos & Power, 2013). Based on the report, we selected five regions of interest: Morrisania, Brownsville, Ozone Park, Upper East Side, and Tottenville (from poorest to wealthiest). We used Google Maps to pick central locations within these regions and extracted their geographic coordinates (longitude and latitude). Next, we defined a 5 km radius around each region's central point and identified crimes occurring within these regions. Some assaults may belong to multiple regions due to small overlap between regions (Figure 5.1).

Figure 5.1, panel A shows the five regions of interest and a 5 km radius (yellow circles) delineating the crimes relevant to each region in NYC. We also plot the locations of all assaults occurring in NYC throughout 2013 (arbitrarily chosen to illustrate our approach) to visualize the spatial variance across regions. Each green dot corresponds to one reported assault. We show the temporal variation by plotting the monthly number of assaults recorded across the measurement period for each region in Figure 5.1, panel B. The vertical axis displays the number of assaults per 100,000 inhabitants to correct for population density. The grey shaded area highlights the year 2013. In addition, we show different temporal resolutions (i.e., daily, weekly, biannually, and yearly) of assault rates in Appendix 3, Figure A3.1. Although our framework can be applied to any resolution of the data, we chose a monthly resolution for all analyses. The biannual resolution hides a large portion of the variation in crime rates visible at higher resolutions. The weekly and monthly resolution show qualitatively similar patterns of variation. We chose the monthly resolution because it reduces the number of data points per region by a factor of 52, from 9,360 to 180, making it computationally less expensive to extract environmental statistics. Of course, researchers should choose resolutions that match their research questions and theories (Hopwood et al., 2021). For example, if a theory states that monthly, weekly, or yearly fluctuations in crime rates have a different impact on development, those resolutions are an appropriate focus. However, we acknowledge that sampling frequency in developmental research is often dictated by practical limitations, such as funding and the availability of participants.

Simulating individuals

We used the region-wise data as a basis for our simulation. Suppose an individual lives in Brownsville in January 2006. At each time period (month), there is a probability P_{stay} that she will keep living in Brownsville and a probability $P_{move} = 1 - P_{stay}$, that she will move to a different region. We then specify whether moves from Brownsville to the other five regions are equally likely or whether some moves are more likely than others. We chose $P_{stay} = 0.98$ for all regions of interest and specify an equal probability of moving to any of the other five regions, $\frac{P_{move}}{4} = \frac{0.02}{4} = 0.005$. With this setup, we can simulate a sequence of regions for a person who lives in Brownsville at the onset of 2016. Rather than deciding on a fixed starting region for the entire sample, we randomly sampled starting regions, one for each simulated individual. Each region was equally likely to be sampled. We simulated 500

3 https://www.icphusa.org/reports/on-the-map-the-dynamics-of-family-homelessness-in-new-york-city-2/

individuals. Figures 5.2-5.3 and 5.5-5.6 show examples of time series for a subset of our simulated individuals.

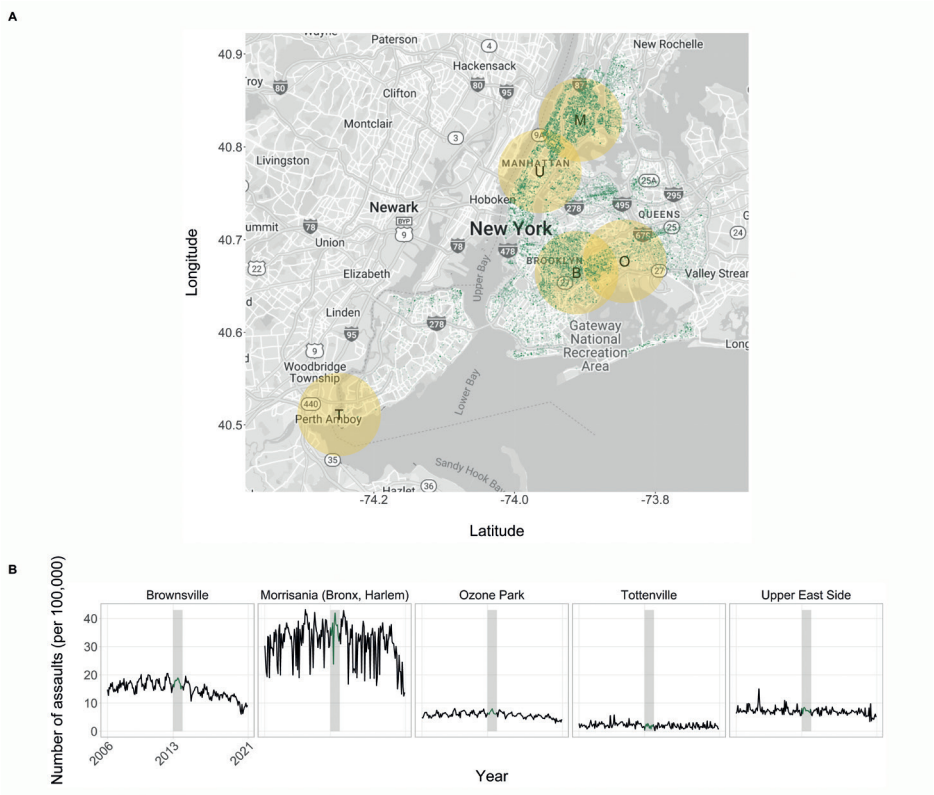


Figure 5.1 Regions of interest in the New York Crime data. Panel A shows the five regions of interest Morrisania, Brownsville, Ozone Park, Upper East Side, and Tottenville (from poorest to wealthiest) and a 5 km radius around them in yellow. The horizontal axis shows the longitude and the vertical axis the latitude. Green dots indicate assaults occurring throughout 2013. Each dot corresponds to one reported assault. Panel B shows monthly assault rates across the measurement period for each region. The vertical axis displays the number of assaults per 100,000 inhabitants and the horizontal axis denotes time in years. The grey column highlights the year 2013.

Our method is flexible. We can compare individuals who have moved often to those who moved less frequently by varying the overall probability of moving. Similarly, we can compare individuals that have experienced high levels of harshness to those who have experienced lower levels. We can do this by making moves within harshness levels more likely and moves across harshness levels less likely. Based on this assumption, a person currently living in Brownsville would be more likely to move to Morrisiania than to Tottenville (Figure 5.1). Such flexibility allows us to explore the effects of different assumptions. For example, we could explore how a sample with low upward mobility (i.e., difficulty escaping poor economic circumstances) compares to a sample with higher mobility. Alternatively, we may use actual demographic data – indicating where people live and how often they move – to inform our initial distribution of starting regions and move probabilities.

5.3 Results

We divide the results into two parts. In the first part, we present results of applying our framework to the New York Crime data. We show unpredictability statistics for each region of interest (Table 5.2, Figure 5.2) and simulated individuals who move between regions (Figures 5.3-5.4). We illustrate to what extent different definitions of unpredictability result in different rank-orderings of regions and individuals. In addition, we present examples of stability and change statistics for the simulated individuals (Figures 5.5-5.6). In the second part, we provide an accessible, step-by-step guide on how to apply our framework.

Environmental statistics – New York Crime data

Statistical definitions of unpredictability

We show six different statistical definitions to quantify unpredictability as 'random variation in harshness': the standard deviation, entropy, color of noise, autocorrelation (at a lag of one month), changepoints in mean, and changepoints in variance (see Tables 1-2). Higher standard deviation, entropy, and number of changepoints in mean and variance indicate higher levels of unpredictability. The same is true of lower absolute (i.e., the magnitude ignoring the sign) autocorrelation and color of noise.

We consider the NYC regions of interest. Figure 5.2 shows changepoints in mean (green line) and variance (grey rectangles) in monthly assault rates for each region. On average, a higher green line implies more assaults and a longer rectangle more variance. Individual horizontal, green lines and individual grey rectangles mark segments of stable mean and variance in the data. A shift in the green line or a new grey rectangle indicate when the mean or variance in the data have changed. The total number of such changes across the measurement period indicates the number of changepoints in mean or variance. We observe differences in the number of changepoints in mean across the different regions with Morrisania having the most and Tottenville the fewest. The differences in the number of changepoints in variance are smaller (Table 5.2). Almost all regions show an increase in mean or variance of assault rates between 2007 and 2010 and a decrease between 2019 and 2021. The increase might reflect the global financial crisis in 2007 and 2008. The decrease may be due to lockdowns and other measures against the Corona virus in 2020 and 2021.

Table 5.2 provides statistics of unpredictability for all regions of interest. Per column, grey bars indicate how different regions rank in unpredictability for a given statistical definition. The longer the bar, the more unpredictable that region is relative to the others. Table 5.2 shows that Tottenville ranks lowest in unpredictability according to the majority of statistical definitions, while Morrisania ranks highest. The high entropy value and high number of changepoints in variance of the Upper East Side may be due to its spatial overlap with Morrisania (Figure 5.1). Across our regions of interest, poverty appears to be associated with higher levels of unpredictability.

Region	SD	Color of noise	Entropy	AC (lag 1)	CP (mean)	CP (var.)
Brownsville	2.91	-0.14	0.89	0.83	18	8
Morrisania (Bronx, Harlem)	6.96	-0.78	1.00	0.33	23	2
Ozone Park	0.99	-0.70	0.83	0.73	4	4
Tottenville	0.86	-0.97	0.51	0.09	2	3
Upper East Side	1.31	-0.84	0.96	0.22	8	12

Table 5.2 Unpredictability statistics in regions of interest in New York City between January 2006 and December 2020. Columns indicate statistical definitions of unpredictability and rows regions in NYC. Within each column, the length of grey bars ranks all regions according to their degree of unpredictability. The longer the bar, the more unpredictable that region is relative to the other regions. 'SD' refers to the standard deviation, 'AC' to the autocorrelation, and 'CP' to changepoints.

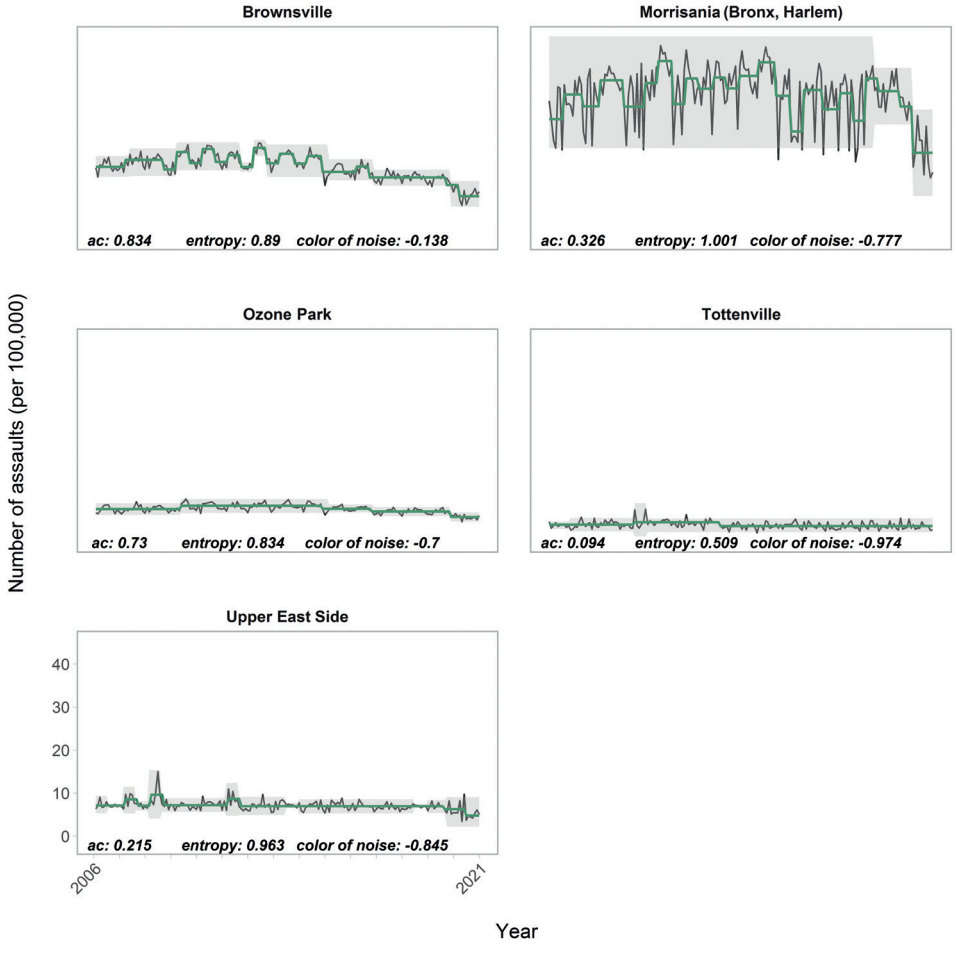


Figure 5.2 Changepoints in mean and variance in NYC regions of interest. Each panel depicts one region of interest. Within each panel, the horizontal axis shows time in years and the vertical axis number of assaults per 100,000 inhabitants. The green line tracks changes in mean and grey rectangles track changes in variance in monthly assault rates. A higher green line implies more assaults; a longer rectangle more variance. We show the values of other unpredictability statistics at the bottom of each subplot. The abbreviation 'ac' refers to the autocorrelation.

Simulated individuals

Next, we examine statistics of unpredictability for simulated individuals who moved across our regions of interest. Figure 5.3 shows statistics of unpredictability for individuals with the same standard deviation. Although matched in their standard deviation, their time series look quite different. Individual 382's data show a high number of changepoints in variance and moderately high entropy, all suggesting exposure to high levels of unpredictability. Individual 80's data show fewer changepoints in variance, and higher autocorrelation, indicating relatively low exposure to unpredictability. However, the number of changepoints in mean is higher for individual 80 and the color of noise values are similarly low, suggesting mixed patterns of unpredictability. Had we just looked at the standard deviation, we would have not been able to distinguish the time series. Even if we compare individuals who have the same standard deviation and number of changepoints in mean (Figure 5.4), their time series look rather different. However, statistics across all three individuals are more similar than they were when only matching the standard deviation. Different unpredictability statistics contribute to the overall picture, creating a higher resolution description of the environment experienced by an individual. This helps us to detect nuances in levels of unpredictability among individuals.

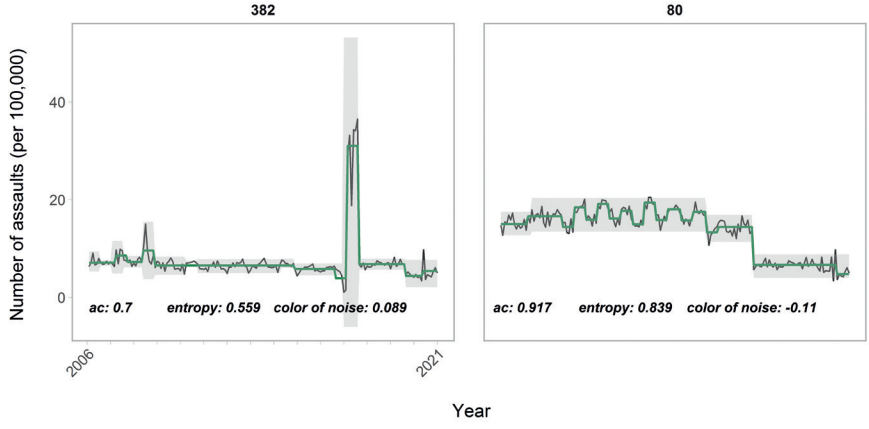


Figure 5.3 Unpredictability statistics for two arbitrarily chosen individuals with the same standard deviation (SD_{382} = 4.75, SD_{80} = 4.76). The panel headers identify two simulated individuals (382 and 80). Within each panel, the horizontal axis shows time in years and the vertical axis number of assaults per 100,000 inhabitants. The green line tracks changes in the mean and grey rectangles track changes in variance in monthly assault rates. A higher green line implies more assaults; a longer rectangle more variance. We show the values of other unpredictability statistics at the bottom of each subplot. The abbreviation 'ac' refers to the autocorrelation.

Comparing regions and simulated individuals

In most regions of interest, unpredictability statistics paint a similar picture. Regions that score low on unpredictability on one of the statistics also tend to score low on others. The autocorrelation breaks this pattern. A reason for this is that the autocorrelation picks up trends in the data. Trends in mean often cause high autocorrelation values of an entire time series, even if subsets of the time series only show relatively low autocorrelation (see Brownsville Figure 5.2, Table 5.2). A positive (negative) trend implies that current values

in the time series are associated with higher (lower) values later in time, causing high autocorrelation. To account for this, we may choose to first remove trends in the time series prior to computing the autocorrelation (see *Steps 2-4* in 'A guide to using the framework').

Unpredictability statistics of simulated individuals who moved across regions are less consistent. Even when we match individuals in standard deviation and number of changepoints, the other statistics do not always point in the same direction. That is, individuals rank differently in terms of their exposures to unpredictability. This may indicate that quantifying unpredictability is more difficult in time series of individuals than in time series of regions. A reason for this may be that assault rates within regions only vary across time, whereas assault rates within simulated individuals vary across both time and space. To represent data that vary temporarily and spatially, we may need to use a higher number of environmental statistics or statistics that are specialized for both types of variation (Burgess & Marshall, 2014; Guélat & Kéry, 2018; Marshall & Burgess, 2015; Stimson, 1985).

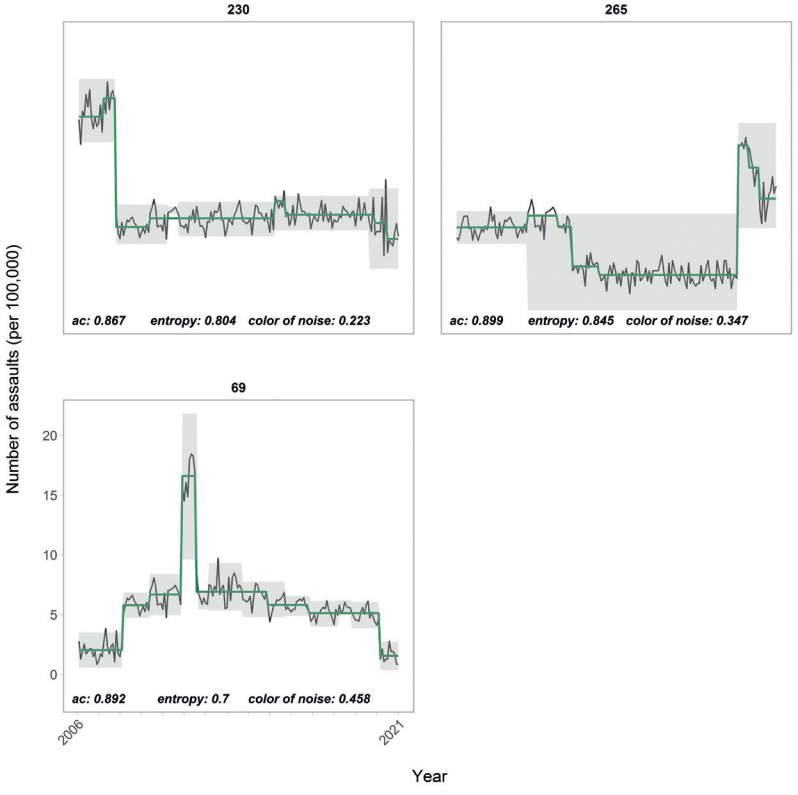


Figure 5.4 Unpredictability statistics for arbitrarily chosen individuals with the same standard deviation (SD_{230} = 3.09, SD_{265} = 3.05, SD_{69} = 3.05) and the same number of changepoints in mean (n = 7). The panel headers identify the simulated individuals (230, 265 and 69). Within each panel, the horizontal axis shows time in years and the vertical axis number of assaults per 100,000 inhabitants. The green line tracks changes in mean and grey rectangles track changes in variance in monthly assault rates. A higher green line implies more assaults; a longer rectangle more variance. Additionally, we show the values of other unpredictability statistics at the bottom of each subplot. The abbreviation 'ac' refers to the autocorrelation.

Quantifying stability and change

Figure 5.5 visualizes one stability, one predictable, and one unpredictable change statistic for two simulated individuals. The yellow, dashed line indicates the mean of assault rates across the measurement period (stability statistic). The green line indicates the linear slope in the data (predictable change). Grey rectangles track changes in variance in monthly assault rates (unpredictable change). A longer rectangle implies more variance within each segment.

The different statistics tell different stories. Individual 55 (left panel) shows a slightly lower average exposure to assault rates than individual 105 (right panel). However, the mean hides the fact that individual 55 has also experienced a steep increase in assault rates and individual 105 experienced a decrease. We also observe that individual 55 experienced fewer changepoints in variance.

To better understand the variability in the data, we can compute stability and change statistics for the variance. This can help us answer different questions about the data. For example, what is the mean level of variability in crime rates an individual has experienced? Has an individual experienced long-term trends in variability in crime rates? To this end, Figure 5.6 shows mean, slope, and changepoints in variance of the squared deviations from the mean. Variability in the data may also be conceptualized in different ways. For example, similar to Li et al. (2018), we could use deviations from the slope instead of deviations from the mean. Our framework allows for both possibilities. Figure 5.6 shows statistics for the same individuals as in Figure 5.5. These statistics reveal that both individuals experienced very similar average levels of variability in assault rates. Individual 55 experienced an increase in variability, whereas variability in assault rate exposure slightly decreased across time for individual 105. However, for both individuals, non-linear models (rather than a linear slope) may be better suited to capture how variability behaves across time. Our framework provides the possibility to fit non-linear models.

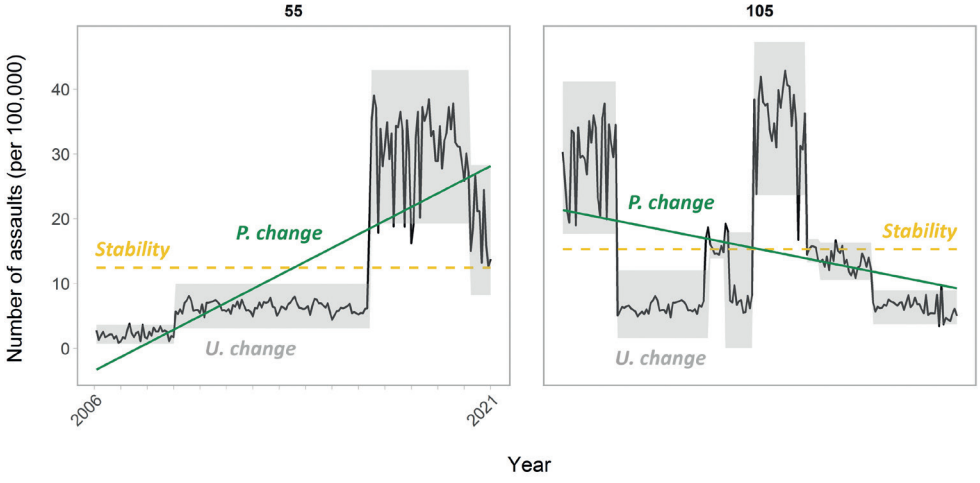


Figure 5.5 Statistic of stability and change for arbitrarily chosen individuals. We show three types of statistics: one stability statistic, one predictable change statistic, and one unpredictable change statistic, for two simulated individuals. The panel headers identify the simulated individuals, i.e., 55 and 105. Within each panel, the horizontal axis shows time in years and the vertical axis number of assaults per 100,000 inhabitants. The yellow, dashed line indicates the mean (stability statistic) of assault rates across the measurement period, the green line indicates the linear slope in the data (predictable change), and the grey rectangles track changes in variance (unpredictable change). The height of a rectangle is proportional to its variance.

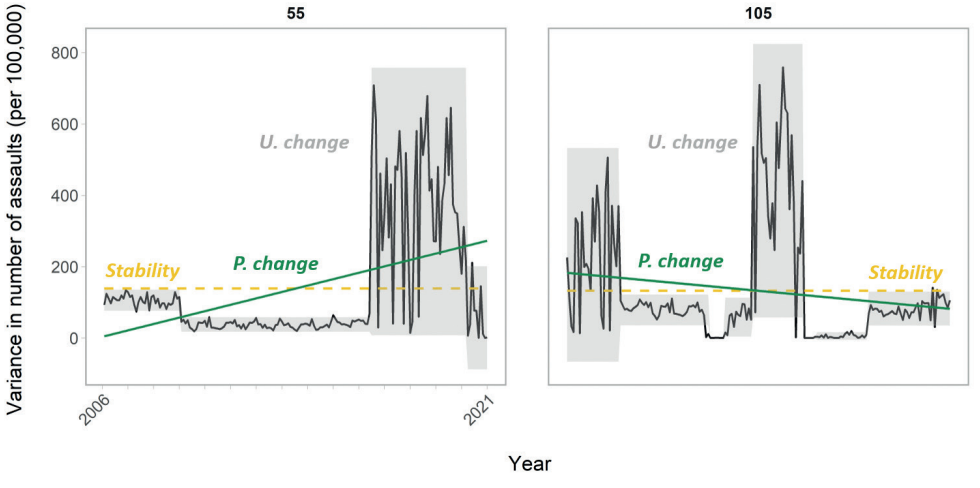


Figure 5.6 Statistic of stability and change applied to the squared deviations from the mean. We show three types of statistics: one stability statistic, one predictable change statistic, and one unpredictable change statistic, for two simulated individuals. The panel headers identify the simulated individuals, i.e., 55 and 105. Within each panel, the horizontal axis shows time in years and the vertical axis variance in the number of assaults per 100,000 inhabitants. The yellow, dashed line indicates the mean (stability statistic) of assault rates across the measurement period, the green line indicates the linear slope in the data (predictable change), and the grey rectangles track changes in variance (unpredictable change). The height of a rectangle is proportional to its variance.

A guide to using the framework

We provide an accessible step-by-step guide for applying our framework (Figure 5.7). The framework can be understood as a tool for data compression: it reduces the complexity of a raw time series to a few statistics, while trying to maintain information about how the data behave across time. Here, we describe some of the specific properties of time series data that are relevant to our methodology. We provide more detailed instructions about how to use the framework in Appendix 3.

Step 1

First, we check whether the data meet the criteria for time series analyses. Our framework takes a dataset with repeated measures of an environmental variable as input. Ideally, the dataset has at least 20 repeated measures per individual. The more repeated measures, the better: some time series modeling techniques require at least 50 observations (Haslbeck & Ryan, 2021; Jebb et al., 2015). As is the case with all cut-offs, they should not be understood as strict rules but as guidelines. A dataset with 15 or 19 repeated measures may also be suitable for our framework (Hyndman & Athanasopoulos, 2018). However, the lower the number of repeated measures, the more likely statistics may be tracking noise in the data, increasing uncertainty in estimates. In these cases, we should be cautious when interpreting the values of environmental statistics. Also, smaller samples are more problematic for some statistics than others. For instance, the mean over 15 data points might be a good representation of the data if the variance is low. However, computing the autocorrelation or changepoints for those same data might result in worse estimates. All else being equal, shorter time series imply noisier estimates.

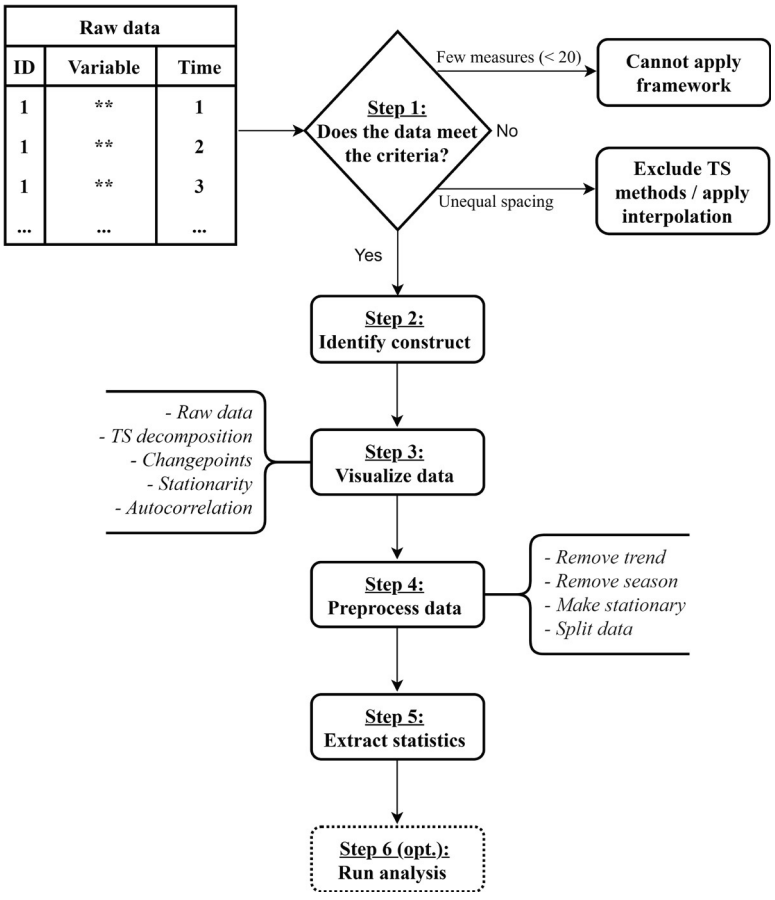


Figure 5.7 A step-by-step guide to using our framework. Our framework takes longitudinal data as input and returns a set of environmental statistics. First, we check whether the data meet the criteria (Step 1) and identify our construct of interest (Step 2). Then we can visualize and preprocess your data (Steps 3-4), before extracting environmental statistics of stability and change (Step 5). Lastly, we can choose to use environmental statistics relevant to our construct of interest as predictors in subsequent analyses (Step 6).

Our framework is more easily applied when measures are equally spaced in time (e.g., once a month or year). Equal spacing is especially important for computing autocorrelation and decomposing the time series (see Steps 2-4). These computations are common in time series analyses (de Haan-Rietdijk et al., 2017; Jebb et al., 2015; Jones, 1984). Equal spacing matters because time series methods typically make assumptions about the sampling frequency of the data. For example, autocorrelation can be computed at different lags in the data. If data is measured monthly, autocorrelation at lag 1 corresponds to the correlation of the time series with itself shifted by one month. Lag 2 autocorrelation uses a shift by two months. To meaningfully estimate the autocorrelation we require equal, monthly spacing.

There are different ways of dealing with unequal spacing. If the irregularity in spacing is small, one option is to ignore it while being explicit about this as a (minor) limitation. For example, if measures are taken every first of the month but some are taken on the second or third, the spacing can still be considered equal. However, if the degree of irregularity is more serious one solution is to exclude statistics and preprocessing steps that assume equal spacing (i.e., autocorrelation and time series decomposition). Thus, in those cases, we advise to only compute statistics that do not make assumptions about regularity. Or, the data can be transformed to become equally spaced. This can be achieved by interpolation (Pavía-Miralles, 2010). Interpolation uses available data points in a time series to fill the empty slots that would not be there if the series had been perfectly regular. Interpolation estimates the most likely values for those slots using the available data. However, interpolation can also bias the data and distort their true dynamics, so before using this method, we advise exploring additional literature (Dezhbakhsh & Levy, 1994; Erdogan et al., 2005).

We do not offer recommendations on sample size because our framework does not fit models to the entire sample. Fitting models to an entire sample is common when exploring research questions about a population. Our framework only fits models (e.g., estimating a linear slope) to individual time series. Thus, it can be applied to small samples if the goal is to only explore these individual time series. However, if the extracted statistics will be used as predictors in an analysis to explore questions about a population, the appropriate steps should be taken to ensure an adequate sample size (see Steps 5-6).

Steps 2-4

Second, we need to decide on the construct that we want to compute environmental statistics for. To show how our framework can be applied, we explored unpredictability as a case study. Generally, it holds that more precise construct definitions make it easier to narrow the range of possible statistics.

Third, we suggest visually exploring the data. Our framework offers various options. We can plot the raw time series, the autocorrelation, and changepoints of randomly selected individuals (Appendix 3, Figure A3.2). Additionally, we can decompose the time series into its individual components: trend, season, and random component. The *trend* describes how the level of a time series changes with time. *Season* refers to the presence of regularly occurring patterns within a particular time period. The *random component* is the residual data after subtracting the trend and seasonal patterns. Visual exploration is necessary to get familiar with the data and to decide on whether and how we should preprocess the data (see Step 4).

By visually exploring the data, we can assess whether they are stationary or not. If the mean, variance, or autocorrelation of a time series are constant, it is stationary. However, if the mean, variance, or autocorrelation change, the series is non-stationary (Jebb et al., 2015; Young et al., 2020). Time series with a trend or increases or decreases in variance are non-stationary. Some time series models assume stationarity, and non-stationary series need to be transformed to meet this assumption (Jebb et al., 2015). Removing changes in mean and variance is usually sufficient to achieve stationarity. Our framework does not require stationary data. However, as noted earlier trends in the mean or variance might distort

some of the statistics in Table 5.1. To estimate how much current values are correlated with later values irrespective of the trend, we can compute the autocorrelation of the stationary series. Our framework, therefore, extracts some statistics for both the raw data and the stationary data.

Fourth, we can preprocess the data before extracting statistics. For example, if we want to explore statistics over stationary series, our framework provides tools to remove the trend in mean or variance, and seasonal components from the data. It also offers the option to split the data at different time periods, resulting in several smaller datasets. This may be useful if we are interested in separately exploring environmental statistics before and after a specific age or time period.

Steps 5-6

Fifth, we extract environmental statistics of stability and change. The framework takes our input data (raw or preprocessed) and returns a dataset that contains a range of environmental statistics for each individual. To estimate a slope, you can specify a model that is fitted to each individual's time series. In the simplest case, the framework will fit a linear model with time as the predictor and the environmental variable as the outcome. We could also apply more complicated models, such as a polynomial. We can then select and plot resulting environmental statistics that fit the conceptual definition of our construct. This step will result in environmental statistics and plots as seen in Table 5.2 and Figures 5.2-5.6.

Sixth, we may use the resulting statistics (e.g., unpredictability scores for each individual) as variables (e.g., predictors) in subsequent analysis. For instance, a researcher may explore the impact of unpredictability on key life outcomes such as health, wellbeing, and psychopathology. As noted earlier, it is important to ensure a sufficient sample size for such an analysis.

5.4 Discussion

Developmental science struggles with a flurry of conceptualizations and operationalizations of constructs central to the field, such as unpredictability. We have offered a step towards coherence by proposing a framework that provides statistical definitions of environmental stability and change. These definitions can serve as building blocks that clarify connections between empirical studies, enabling cumulative construction of knowledge. Researchers using our framework will also be able to better integrate their theory and findings with those in the other biological and social sciences (e.g., biology, ecology, anthropology), which have been using similar frameworks for decades (Bernardi & Hutter, 2007; Burgess & Marshall, 2014; Hammel, 2005; Marshall & Burgess, 2015; Vasseur & Yodzis, 2004; Warlaumont et al., 2021).

To showcase our framework, we have presented a case study of unpredictability, using publicly available crime data from New York City. In this section, we discuss how our framework can help advance research using constructs of environmental stability and

change, such as unpredictability. First, our framework affords comparing and correlating different statistical definitions of the same construct. Second, it can facilitate linking individual- and environment-level data. Third, it can inform conceptual definitions and measures of constructs. Fourth, it can aid the development of a database of environmental statistics. Fifth, it can bridge theoretically- and empirically-driven approaches to studying environmental exposures. The first three points highlight benefits for individual studies, whereas the latter two discuss benefits that emerge across studies. We close by discussing limitations.

Comparing and correlating environmental statistics

Our framework can be used to examine the correlations between different statistical definitions of the same construct. For instance, our analysis showed that different statistical definitions produced similar rank orderings of regions along the dimension of unpredictability. Across regions, we also observed a positive association between different unpredictability statistics and poverty. Unpredictability statistics of simulated individuals who moved across these regions resulted in less consistent rankings of individuals. Suppose that we would have also computed unpredictability statistics for regions within other cities in the US, or more rural areas. We could explore whether statistical definitions of unpredictability behave similarly within different cities and rural areas. With a larger sample of regions, we could compute correlations between different statistical definitions of unpredictability. We may also test the associations between different definitions and poverty. Such work would reveal the extent to which different types of unpredictability covary with each other, and covary with poverty. It may also help us understand the factors that make some environments more unpredictable than others (e.g., temporal fluctuations in crimes). Similarly, unpredictability statistics may also be computed at other spatial resolutions, such as provinces, states, or countries. With such data we could explore whether unpredictability statistics look similar on different spatial resolutions. We may also study whether high unpredictability within a country shapes development in different ways than high unpredictability within a city.

Linking individual- and environment-level data

As already noted, our framework can be applied to individual-level data, as well as environment-level data (e.g., crime records). In our case study, we have shown how to compute environmental statistics for environment-level data, both on their own (NYC regions) and linked to individuals (simulated data). Combining individual- and environment-level data can paint a more precise picture of an individual's lived experience. Examples of integrating data of an individual's surrounding ecology into developmental studies already exist. For example, Snyder et al. explored whether the prevalence of crime in a woman's surrounding ecology predicts a preference for aggressive and physically strong men (Snyder et al., 2011). Current and childhood prevalence of crime were assessed by linking women's zip codes to local crime indices. Similarly, Miller et al. explored whether individual differences in brain connectivity moderate the relationship between cardiometabolic health (e.g. obesity or insulin resistance) and neighborhood violence (Miller et al., 2018). As a proxy for neighborhood violence, they computed a neighborhood murder index using public, local crime records. Both studies benefitted from the inclusion of public datasets to enrich individual-level measures of crime and violence. In these cases, environment-level data enriched answers to existing research questions. But environment-level data can do more

than that. They provide us with opportunities to ask new research questions. For example, a recent study investigated whether structural forms of stigma resulting from regional social policies, shape brain development in Black and Latinx youth (Hatzenbuehler et al., 2020). Without the availability of environment-level data we would not be able to ask this question.

Another benefit of using publicly available data is that they often provide a higher number of repeated measures than data collected in the laboratory (Kievit et al., 2021). Kievit et al. (2021), however, also pointed out challenges to using such data. Openly available data may have been collected with a different purpose in mind and may drastically vary in ease of handling, depending on the size of the dataset and the documentation. In addition, such data may be subject to bias. For example, the number of recorded assaults in NYC in our dataset may be biased by the police. Police officers may consciously or subconsciously under- or overreport assaults occurring in different demographic areas (Warren, Tomaskovic-Devey, Smith, Zingraff, & Mason, 2006). We were not able to determine for our data whether they contain bias. However, it is important to be aware of the possibility of bias in the data and to be cautious when interpreting findings.

Constructs, measurements, and statistical formalizations

We have focused on statistical definitions of environmental stability and change, using the construct of unpredictability as a case study. However, difficulties also arise at earlier stages when developing such constructs and their measures. Constructs in psychology often suffer from the 'jingle-jangle' fallacy: we consider constructs with similar names to be theoretically and empirically similar and those with dissimilar names to be different. Some might think this problem can be solved by giving the underlying constructs more distinctive names. However, the real problem lies deeper: because natural language is ambiguous, the definition of constructs is imprecise. This problem can be overcome through formalization. But even if we used theory to define constructs precisely, we still need to empirically verify that our constructs measure what they are supposed to. This involves checking whether indicators of our construct correlate with indicators of theoretically similar constructs ('convergent validity') but not with those of dissimilar constructs ('discriminant validity'). This step is often omitted when developing new constructs (Flake & Fried, 2020; Hodson, 2021). As a consequence, some psychological constructs that are considered theoretically distinct in the literature are actually highly correlated (e.g., burnout and depression; Schonfeld & Bianchi, 2016; Schonfeld & Verkuilen, 2019; grit and conscientiousness; Credé et al., 2017). Although their measurement instruments were designed to measure distinct constructs, they do not. We do not know whether the measurement instruments are ill-suited to capturing all (and only those) elements that are relevant for the constructs or whether these constructs are actually the same.

The problem of construct and measurement validity is complex and there exists no easy solution. As an example of steps taken to improve these issues, Flake and Fried developed a guide to identify 'questionable measurement practices', such as non-transparent reporting of the measurement procedure (Flake & Fried, 2020). They also discuss the use of formal theorizing as one way to develop more precise construct definitions. While our framework does not offer solutions to construct and measurement validity, it may aid the development of more precise definitions. Just like formal theorizing and modeling, our framework

invites researchers to be explicit about the conceptual and statistical definitions of their constructs (Borsboom et al., 2021; Frankenhuis & Tiokhin, 2018; Smaldino, 2020). When various statistical definitions are equally suited to capture the construct, this may indicate that its conceptual definition is not (yet) precise enough. Our framework encourages the exploration of different statistical definitions and the subsequent refinement of constructs and measurement instruments.

A database of environmental statistics

Applied across studies, our framework can contribute to building a database of environmental statistics for developmental research (Frankenhuis, Nettle, et al., 2019). Such a database would function as an open, shared platform for storing statistics computed for different environmental dimensions. Statistics may be computed across both time and space (e.g., autocorrelation across a child's first ten years of life and different residencies) and different resolutions (e.g., days or years, cities or countries) (see also Mendoza & Fausey, 2013). Knowing the range of values for different environmental statistics has the potential to inform empirical and theoretical work.

Empirical researchers can use the database to examine existing hypotheses or generate new ones. For example, the database could provide the necessary data to test the association between different statistical definitions of unpredictability and poverty. Theoretical modelers can use it to set model parameters based on empirical values. For instance, a modeler may want to explore to what extent environmental stability shapes organisms' ability to adjust development based on early experiences. She might use the database to set the autocorrelation parameters in her model to empirically occurring values. Modelers could also use the database to evaluate the plausibility of model outcomes. For instance, if modeling shows that it is only adaptive to use early experience to shape development if environmental autocorrelation is very high (Nettle et al., 2013), she might consult the database to evaluate how common such high values are.

Quantifying the statistical structure of the environment on shorter time scales has already proven to advance infancy research. For example, Warlaumont et al. (2021) documented everyday auditory experiences of infants. They found that infants seek out vocal responses from adults similar to how animals forage for resources. This parallel between vocal exploration and foraging dynamics offers opportunities to generate novel hypotheses about learning in infants. Similarly, Smith and colleagues used head cameras and eye trackers to document infants' everyday visual experiences (J. E. Smith & Pinter-Wollman, 2021). As sensorimotor development progresses, infants' interactions with their visual environment change, granting them access to novel experiences; referred to as 'curriculum for learning'. The authors hypothesize that infant learning is optimized for the continuously changing visual environment. Computing environmental statistics can advance developmental research in similar ways as documenting the early environment has advanced infancy research.

Bridging theoretically- and empirically-driven approaches to studying environmental exposures

Our framework supports both theoretically-driven (top-down) and empirically-driven (bottom-up) approaches to studying environmental exposures. Theoretically-driven approaches derive variables and constructs of environmental exposures from theory. Empirically-driven approaches derive them from empirical studies. Sometimes, theoretically derived constructs are too broad and unspecific to capture nuances of environmental exposures (Pollak & Smith, 2021). The extent to which this is the case depends on the specificity of the theory. Ideally, theory development and empirical studies take turns and refine each other. Bringing theoretically specified constructs and empirical studies exploring these constructs together can accelerate scientific progress. Our framework can help on both ends. It aids theory development by inviting precision about definitions of constructs and their formalizations. But it also offers tools to realize these different statistical definitions in empirical studies. With results from those studies, we can compare and refine existing constructs of environmental exposures.

Limitations

Our framework is only a starting point to compute environmental statistics of stability and change. In some cases, it might be difficult to interpret the numeric values of individual statistics. What does an entropy of 0.95 tell us about unpredictability without comparing it against other regions or individuals? To be able to interpret these statistics we first need to compute them more regularly and in different contexts (e.g., samples, time scales, countries). A database that collects the values of these statistics (see 'A database of environmental statistics') can helps us to organize them. Such a database would also make it easier to compare environmental statistics across contexts and to self-reported measures of relevant constructs (e.g., unpredictability). In the long-term this may allow us to interpret the values of individual statistics. Along the way, we might also gain some insight into how objective measures of environmental constructs, such as harshness and unpredictability, relate to subjective perceptions of these constructs (Bartels, 2002).

Regularly computing environmental statistics can also help us to explore the different adaptive challenges organisms face for different types of unpredictability. For example, we have considered high absolute autocorrelation and color of noise as more predictable than lower values. However, high negative autocorrelation and color of noise may influence individuals quite differently than high positive values. High negative values indicate predictable oscillations between high and low values of, for example, harshness. High positive values indicate long runs of above or below average conditions. An individual exposed to the former can predict environmental conditions in the near future well, experiencing little unpredictability. At the same time, she also has to cope with those drastic fluctuations in harshness. Being aware of these differences and empirically exploring how they shape development can deepen our understanding of how organisms adapt to unpredictability.

Our statistical formalizations are simple and do not cover more advanced time series modeling approaches. Rather than extracting individual statistics from time series data, as we do, these models can be used to test hypotheses about how the data change across time. For example, Jebb et al. offer a beginner-friendly tutorial on time series analysis in

psychological research and provide pointers to further, more advanced reading (Jebb et al., 2015). Applying time series modeling to developmental research may help us answer research questions that we cannot answer with our current methods. However, successfully applying time series modeling to developmental questions can be challenging. Haslbeck and Ryan address such challenges in the study of emotion development (Haslbeck & Ryan, 2021). They focus on model misspecification and sampling frequency. A misspecified model does not include all possible dynamics and variables that have produced the data. Every model is misspecified but the degree and consequences vary. A good model includes the most essential dynamics and variables and advances our understanding of the data generating process. Sampling frequency refers to the timescale at which the data are measured (e.g., seconds, minutes, hours, or days). The authors simulate individual trajectories of emotion development across the span of days and apply different sampling frequencies (ranging from 6 seconds to 90 minutes) and models. They show that even a misspecified model applied to undersampled data can recover global dynamics of emotion development. Future work may explore whether these insights also translate to data sampled over weeks, months, and years. Insights from such work may help us to assess some of the challenges of applying time series modeling to developmental data.

One of the biggest factors constraining the feasibility of our framework and time series models, is the length of available time series data. We have recommended a minimum of around 20 repeated measures to extract environmental statistics. However, reaching this number can be challenging. Although ideally practical limitations should not dictate how we conduct science, they present real barriers. Fortunately, with calls for more collaborative science on the rise, increasing incentives to share data collected from participants, and the availability of public environment-level data (e.g., crime records), we see exciting opportunities ahead for systematically quantifying environmental stability and change across human development.

Chapter 6

General discussion

6.1 Outlook

As part of this dissertation, I have presented central tenets, insights, and predictions from existing models of sensitive period evolution. I have contributed to this literature by developing two models of the evolution and development of sensitive periods. These models extend previous models that assume incremental and irreversible development in two ways: the first model explores the evolution of sensitive periods when organisms experience variation in the reliability of cues; the second model explores variation in the environmental state across ontogeny. To facilitate synergies between such models and empirical data, I have also developed a framework for studying environmental statistics across development. This is the third (and methodological) contribution of this dissertation.

As a whole, my dissertation contributes to scientific consilience. Consilience means the "linking of facts and facts-based theory across disciplines to create a common groundwork of explanation" (E. O. Wilson, 1999). The term 'consilience' dates back to the 19th century when it was coined by William Whewell (Whewell, 1840). But scientific inquiry that transcends disciplines is rare, even today. Being an interdisciplinary researcher is often not rewarded by existing incentive structures in science; journals and grant committees often expect highly specialized work within specific disciplines (Bromham et al., 2016). As a result we tend to conduct science in isolated silos, where the same 'facts' coexist within the confines of the ideas and languages of different disciplines, each with unique assumptions and methodologies (Lim, 2016). My work bridges different methodologies (theory and empirics) and different disciplines (biology and psychology).

The models presented here contribute towards an Integrative theoretical framework of the evolution and development of sensitive periods. They provide novel theoretical insights, generate empirical predictions, and pave the way for future models. My computational framework can help to consolidate these insights with empirical data. Patterns derived from models such as those developed in this dissertation can be compared against environmental statistics from different species and environmental dimensions. The framework also functions as a platform for organizing statistical definitions of environmental stability and change from different empirical studies and disciplines.

In what follows, I will discuss how evolutionary thinking can inform developmental research, what insights I have gained about the development of sensitive periods from modelling their evolution, and the potential of my computational framework to integrate ideas and findings across methods, data, and disciplines.

6.2 Why evolutionary thinking is useful

Dutch biologist and ethologist Niko Tinbergen proposed a foundational framework for explaining any behavior or trait (Tinbergen, 1963). He argued that a complete understanding of behavior requires asking four questions, each providing complementary insights: How does the behavior *develop*, what is its underlying *mechanism* (i.e., its physiological instantiation), why did it *evolve,* and why is it *adaptive*? Typically, answers related to

development and mechanism (questions 1 and 2) are referred to as proximate ('how') explanations, whereas answers to questions about evolution and function (questions 3 and 4) are considered ultimate ('why') explanations. However, the boundaries between these categories are blurry. For example, we know that evolution and development depend on each other. Over evolutionary timescales, natural selection shapes developmental systems, which allow organisms to adapt to their environment across development (Frankenhuis, Panchanathan, et al., 2019). This results in phenotypic variation, which functions as a basis for natural selection in subsequent generations.

A recent article has proposed to view evolution, development, and mechanism of a behavior or trait as a continuum of causes, which operate on different timescales (Bergman & Beehner, 2022). Adopting this updated framework, my models advance our understanding of how environmental conditions over evolutionary time have shaped levels of plasticity across development. The presence of sensitive periods indicates that over evolutionary timescales changes in plasticity across ontogeny were adaptive. Although models like mine typically do not directly provide insights into how plasticity is instantiated, they can inform mechanism.

Examples of this already exist for other models. For example, the idea of predictive adaptive responses ('external PAR') traditionally suggests that early-life stress predicts similarly harsh environmental conditions later in life and that organisms adjust their development accordingly. Evolutionary modelling has shown that under some conditions early-life stress more likely predicts an individual's future internal somatic decline and that the individual adjusts to this decline instead ('internal PAR') (Nettle et al., 2013, 2014). Specifically, we would expect external PARs when environmental conditions are highly correlated across development and internal PARs when they are not (Frankenhuis et al., 2018). These predictions have received empirical support across various species, thus providing insights into how animals respond to different conditions (Berghänel et al., 2016; Chua et al., 2017; Douhard et al., 2016; Hartman et al., 2017).

Sometimes evolutionary explanations and theory are disregarded because they are said to have limited applied value (see e.g., Sunstein, 2022). It is true that evolutionary theory does not provide us directly with solutions to undo trauma from early-life adversity or to avoid societally undesirable behaviors, such as risk-taking or impulsivity. However, evolutionary theories highlight the places where we can look for such solutions. For different environmental conditions and assumptions, they describe the patterns that we would expect to observe in behavior and development (Frankenhuis & Walasek, 2020). For example, evolutionary modelling has identified conditions in which we might expect individuals to act impulsively (Fenneman & Frankenhuis, 2020). Such work has applied value. To design appropriate interventions, we first need to know when and why such behaviors occur.

6.3 Insights from modelling the evolution and development of sensitive periods

My models have provided novel insights into the evolution and development of sensitive periods. My first model shows that, when the reliability of cues increases across

ontogeny, sensitive periods can evolve at later developmental stages (Chapter 3) (Gee, 2022; Walasek et al., 2021). When cue reliability decreases across ontogeny, sensitive periods only evolve at the onset of development. Overall, natural selection appears to have adapted levels of plasticity to track the reliability of cues. These patterns are qualitatively similar across different simulated study paradigms for quantifying plasticity. This suggests that empirical patterns of plasticity might be comparable across different experimental designs. My second model shows that, when cue reliability is constant across ontogeny but the environmental state fluctuates, sensitive periods can occur at the onset, midway through, and even towards the end of ontogeny (Chapter 4) (Walasek et al., 2022). This finding contrasts findings from previous models of sensitive period evolution in which plasticity often reaches zero, and never increases towards the end of ontogeny. Regardless of when during ontogeny plasticity peaks, organisms always retain residual plasticity late in ontogeny when the environment fluctuates. My results thus suggest, that critical periods, after which plasticity reaches zero, are unlikely to be favored in fluctuating environments.

The models presented here and their predecessor (Panchanathan & Frankenhuis, 2016) form a family of models of incremental and irreversible development. Together, they establish links between an organism's evolutionary ecology (e.g., constant or varying environmental states) and expected patterns of sensitive periods (Table 6.1). Each model on its own explores sensitive periods when organisms can only gradually and incrementally develop phenotypes. In a way, the whole family of models also takes an incremental approach to studying sensitive periods, as each model only differs from the others in one aspect. This allowed me to identify how each additional assumption shapes patterns of sensitive periods. Currently, I have not yet developed a model in which both cue reliability and the environmental state vary across ontogeny. This model would fill the empty cell in Table 6.1. This is a direction for further research.

I can aggregate findings across models to derive predictions about sensitive periods. For example, Table 6.1 suggests under what environmental conditions we may expect to observe peaks in plasticity at later developmental stages, as well as critical periods. When environmental conditions only vary between generations, and both the environmental state and cue reliability are constant across an organism's lifespan, we likely will not observe peaks in plasticity beyond early life. Across models, peaks in plasticity at later developmental stages are common. However, the reasons for why they occur differ. We may observe peak-plasticity later in development when organisms use cues that increase in reliability. Alternatively, plasticity may peak at later stages when the environmental state fluctuates and early experiences violate organisms' expectations at birth. In Chapter 4, I have discussed other models that produce peaks in plasticity at later stages when assuming unconstrained and reversible development.

Environmental state	Cue reliability: Constant	Cue reliability: Varying
Constant	Sensitive periods at the onset of ontogeny (Frankenhuis & Panchanathan, 2011b; Panchanathan & Frankenhuis, 2016) Often critical periods	Sensitive periods at the onset and midway through ontogeny (Chapter 3; Walasek et al., 2021) Only critical periods
Varying	Sensitive periods the onset, midway through, and even towards the end of ontogeny (Chapter 4; Walasek et al., 2022) No critical periods	

Table 6.1 Overview over patterns of sensitive periods. Individual cells describe patterns of sensitive periods that emerged from models assuming incremental and irreversible development. Rows indicate whether the model assumes a constant or varying environmental state across ontogeny. Columns indicate whether the models assume constant or varying cue reliability across ontogeny.

Across models, we find that critical periods are often favored when environmental conditions are stable within an organism's lifespan but not when they fluctuate. In stable environments, the prevalence of critical over sensitive periods may depend on whether organisms have access to highly reliable cues at any point during ontogeny. When this is the case, plasticity will likely always reach zero by the end of development, as there is no need for maintaining it. These insights produce several testable, empirical predictions which I have discussed in detail in Chapters 3 and 4.

More generally, my findings contribute to two current topics in developmental research and biology. First, my models enrich our understanding of sensitive periods in adolescence. Empirical work has uncovered different aspects of brain and behavioral development that are shaped during adolescence (Blakemore & Mills, 2014; Fuhrmann et al., 2015; Knoll et al., 2016). Despite this progress we know little about the exact onset, duration, and offset of sensitive periods during this window (Fuhrmann et al., 2015; Gee, 2022). It remains an open question whether plasticity is only heightened during adolescence, constantly elevated across childhood and adolescence, or continuously decreasing across childhood and adolescence (Fuhrmann et al., 2015). For specific traits we may be able to distinguish these three models with controlled experiments, measuring plasticity in the same individuals in childhood, adolescence, and adulthood. Still, this would tell us little about why different traits show different patterns of plasticity. My models can identify conditions that may produce these different patterns, thus steering future empirical studies. They may do so by generating new empirical hypotheses or by providing insights into existing, empirical patterns. For example, a continuous decline of plasticity across childhood and adolescence may be likely in traits for which the brain expects highly reliable inputs early in life.

Second, my models may provide insights into how organisms cope with fluctuating environmental conditions. This topic is of great relevance given the challenges organisms face due to climate change. One of the most urgent questions is how organisms adapt to novel environmental conditions, as this will likely determine their chances of survival

(Jury et al., 2019; Snell-Rood et al., 2018). My model of fluctuating environments cannot answer this question. I have not explored conditions that are unknown to developing organisms. However, my model does provide insights into organisms' capacity to adjust to different rates of environmental fluctuations (Trimmer et al., 2019; Winkler et al., 2014). If the adult lifespan is long relative to ontogeny and one state of the environment is more likely, plasticity is zero across all of ontogeny. This is true for all rates of environmental fluctuations. However, populations whose adult lifespans are short or moderately long relative to ontogeny, are highly plastic at the end of ontogeny when environments fluctuate. If we assume that phenotypic adaptations to known environments are partially useful in novel environments, my model could provide coarse insights into how different species may deal with such novel conditions. Populations with a long adult lifespan may do well if the adaptions to the more common environmental state are also useful for novel states, and poorly otherwise. Populations with shorter adult lifespans may benefit from enhanced plasticity if environmental conditions late in ontogeny predict the adult environment.

6.4 Future ideas for modelling the evolution and development of sensitive periods

Across this dissertation, I have noted various limitations of evolutionary models in general but also of my specific models. Here, I will discuss future ideas for synthesizing existing findings and extending current models. To start with, I would like to synthesize current findings from studying the evolution and development of sensitive periods. There already exist review papers that provide an excellent overview of these models (Fawcett & Frankenhuis, 2015; Frankenhuis & Fraley, 2017; Walasek et al., 2021). What I have in mind is a higher resolution version of Table 6.1 that incorporates the models assuming incremental and irreversible development, as well as models that do not make these assumptions (e.g., Fischer et al., 2014; Stamps & Krishnan, 2014b). By higher resolution I do not mean a higher quantity of models reviewed. Rather, I would like to closely look at the consequences of different assumptions and explored parameter ranges. In its current form, Table 6.1 masks the underlying parameter ranges of each of the models, such as the range of explored priors, cue reliability values, cue reliability patterns, adult lifespans, and rates of environmental fluctuations. A higher resolution version of Table 6.1, which incorporates these nuances, may be useful to identify patterns across models and highlight connections that were previously not visible. Such an overview could provide a valuable resource for generating predictions for subsequent testing in empirical studies, for instance, of experimental evolution in insects (Dunlap & Stephens, 2009, 2016; English & Barreaux, 2020).

The next step would be to extend my current models. In Chapters 3 and 4 I have discussed various possibilities, such as exploring a variable length of ontogeny, reversible development, a continuum of environmental states (opposed to two discrete states), or fitness as a function of fertility and mortality (opposed to just fertility). I have also discussed the idea to explicitly incorporate known proximate mechanisms of plasticity. At present, there are two specific future directions that I find most interesting. First, I would like to explore how sensitive my findings are to the specific assumptions I have made. For each model in Table 6.1, I would like to explore patterns of sensitive periods as a result of changing

one assumption at a time. For example, how would the cells of Table 6.1 differ if all models assumed reversible development? How would they differ, if organisms might die throughout ontogeny when they are poorly matched to their environments? Exploring the consequences of these assumptions would be in line with my incremental approach to studying sensitive periods. Knowing when my findings break down and when they are robust can deepen our understanding of sensitive periods across species with different life histories. The second direction involves more drastic changes in the assumptions of my models.

All of the previous models have assumed that organisms are born 'knowing' the optimal phenotypic response to different environmental states. This implies that natural selection has equipped organisms with phenotypic responses for different environmental conditions. These responses may be genetically encoded and may themselves be the result of natural selection over previous generations. I will refer to these responses as 'innate' in the sense of being present at birth and encoded via genes (see Mameli & Bateson, 2011; Samuels, 2004 for different meanings of 'innateness'). Such innate responses can be adaptive if the range of possible environmental states is small or fixed across generations. If instead this range is very large or variable across generations, organisms may need to learn the adaptive response based on the consequences of their past actions. Learning can broadly be defined as the acquisition of knowledge, abilities, or skills as a result of experience (Frankenhuis, Panchanathan, et al., 2019). Trial-and-error or reinforcement learning represents one type of learning by which organisms may acquire the adaptive phenotypic response (Frankenhuis, Panchanathan, et al., 2019; Moczek et al., 2010; Snell-Rood, 2012). Across all animal taxa, such learning from successes and failures of past behaviors is common in the development of many behavioral traits (Snell-Rood, 2012). Future modeling could explore under which environmental conditions natural selection might favor such learning of adaptive phenotypic responses versus equipping organisms with innate, adaptive responses.

The extent to which traits develop through innate or learned responses may depend on the range and variability of environmental conditions across generations, as well as conditions within an organism's lifespan. If the environment is stable (relative to the organism's lifespan) and cues are reliable, phenotypic responses might be instantiated through innate responses. If the environment fluctuates frequently, natural selection might favor learning of adaptive responses (Fawcett et al., 2014). Such learning may be especially beneficial when organisms encounter novel environmental conditions not experienced by previous generations. Learning likely increases the capacity for adapting to these novel conditions, benefitting organisms' fitness. Another possibility is a combination of innate and learned responses. Both learned and innate phenotypic responses may depend on the organism's ability to accurately sense cues in its environment. However, innate responses might have a smaller margin of error. Suppose an organism is developing phenotypic specializations for a world in which predator density varies across ontogeny. If cues reliably predict predator density, they may induce adaptive innate phenotypic adjustments for different density levels. If cues are poor and it is costly to induce the wrong adjustments, plasticity may not be favored across ontogeny (e.g., Panchanathan & Frankenhuis, 2016). However, if organisms can learn the adaptive phenotypic response through interactions with their environment, they may develop an appropriate phenotype despite poor cues. Thus, organisms might fall back on learning when they fail to infer current conditions based on cues. Although learning

may increase phenotype-environment match, it comes with costs (Snell-Rood & Steck, 2019). Exploring different phenotypes and processing environmental feedback costs time and energy, increasing, for example, the risk of predation. Using modeling, we can explore under which conditions the benefits of learning outweigh these costs.

6.5 Bridges across methods, data, and disciplines

In Chapter 5, I have presented a computational framework for studying environmental statistics across development. The framework highlights different ways in which patterns of environmental stability and change may be defined and computed. In this way it can foster integration and comparability of findings across studies. I have also outlined several ways in which my framework can contribute towards consilience, the integration of all sciences. The framework offers a platform for integrating methods from various fields, such as biology, ecology, or psychology to quantify stability and change. Housing these disciplines under one roof will make it easier to learn from and with each other. The resulting toolkit of methods can be applied to different types of data, such as repeated measures collected from individuals ('individual-level' data) or longitudinal data from their surrounding ecology, such as public crime records or access to health care facilities ('environment-level' data). I have also shown how environmental statistics computed for environment-level data may be linked to individuals through the locations they have lived in. The framework may also strengthen synergies between empirical and theoretical work. Suppose I have already synthesized patterns of sensitive periods from existing models in a version of Table 6.1 that also includes the explored parameter ranges. I could use the values of environmental statistics computed with my framework to highlight specific cells in that table. In this way I could explore patterns of sensitive periods in environmental conditions that match those of the environmental dimension, timescale, and species of interest. From these theoretical patterns I can derive empirical predictions tailored to a specific study organism and subject.

I concluded Chapter 5 by outlining opportunities for systematically quantifying environmental stability and change across human development. I hope that my framework will be part of this endeavor. A distant vision that I have for my framework, is for it to become a public tool, space, and resource for developmental research. My framework could grow into a platform that functions as a tool for researchers to compute environmental statistics of stability and change and as a space to publicly store the resulting values. Ideally, the platform would offer ways to visualize the resulting statistics and to aggregate them across different samples which quantify the same environmental dimension. Other researchers could use the platform as a resource to look up environmental statistics relevant for their own research questions. To facilitate the linking of environment-level and individual-level data, I envision that those statistics should be connected to a map. Suppose that you have already filtered the database based on keywords related to your specific environmental dimension of interest and species. Now imagine, viewing a map of the world. You can click on a specific country and select statistics of stability and change for that country. You can select a time window and zoom in further to view statistics for individual cities. Within cities you can filter neighborhoods based on zip codes. The resulting statistics can then be directly linked to individuals based on the locations they have lived in, allowing us to quantify their

environmental exposures across time and space. The idea for such a platform is inspired by existing repositories, such as the NYC Open Data project[4]. Developing this platform would take years and it would take even longer to fill it with data. However, I am convinced that its benefits would greatly outweigh its costs. Such a platform has the potential to be a tool for connecting different disciplines and methodologies, a space for sharing data and findings, and an open resource for researchers and the public.

6.6 Concluding remarks

"The belief in the possibility of consilience beyond science and across the great branches of learning is not yet science. It is a meta-physical world view, and a minority one at that, shared by only a few scientists and philosophers. It cannot be proved with logic from first principles or grounded in any definitive set of empirical tests, at least not by any yet conceived. It's best support is no more than an extrapolation of the consistent past success of the natural sciences. Its surest test will be its effectiveness in the social sciences and humanities. The strongest appeal of consilience is in the prospect of intellectual adventure and, given even modest success, the value of understanding the human condition with a higher degree of certainty." Edward O. Wilson (1999) (1999). *Consilience: The Unity of Knowledge.*

I chose this brief excerpt from Wilson's book 'Consilience: The Unity of Knowledge' (E. O. Wilson, 1999) because it captures what I tried to convey in this Discussion. I hope that, nowadays, the belief in the possibility of consilience is not only held by a minority of scientists. The past years have seen immense technological progress in our ability to acquire, store, and handle large scale data, as well as gradual reforms in science, promoting theory development and collaborative research. These circumstances make now a unique time to do interdisciplinary research. The studies presented in this dissertation attempt to connect different disciplines and methodologies. I often relied on methods and ideas that are more common in the life or natural sciences than in the social sciences. Based on my experiences of applying these methods and ideas within the social sciences, I found that integrating social and natural sciences adds novel perspectives to understanding development. Most importantly, it creates opportunities for science to progress and consilience to grow.

4 https://opendata.cityofnewyork.us/

Bibliography

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology*, *70*(1), 141–164. https//doi.org/10.1146/annurev-psych-010418-102836

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*(3), 527–543. https://doi.org/10.1037/0033-295X.115.3.527

Agrawal, A. A., Laforsch, C., & Tollrian, R. (1999). Transgenerational induction of defences in animals and plants. *Nature*, *401*(6748), 60–63. https://doi.org/10.1038/43425

Allen, J. P., Chango, J., & Szwedo, D. (2014). The adolescent relational dialectic and the peer roots of adult social functioning. *Child Development*, *85*(1), 192–204. https://doi.org/10.1111/cdev.12106

Auld, J. R., Agrawal, A. A., & Relyea, R. A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 503–511. https://doi.org/10.1098/rspb.2009.1355

Barone, G., & Mocetti, S. (2016). Inequality and trust: New evidence from panel data. *Economic Inquiry*, *54*(2), 794–809. https://doi.org/10.1111/ecin.12309

Barrett, H. C. (2014). *The shape of thought: How mental adaptations evolve*. Oxford University Press.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117–150. https://doi.org/10.1023/A:1021226224601

Bateson, P. (1979). How do sensitive periods arise and what are they for? *Animal Behaviour*, *27*, 470–486. https://doi.org/10.1016/0003-3472(79)90184-2

Bateson, P., & Martin, P. (1999). *Design for a life: How behaviour develops*. London, UK: Jonathan Cape.

Bath, K. G., Manzano-Nieves, G., & Goodwill, H. (2016). Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Hormones and Behavior*, *82*, 64–71. https://doi.org/10.1016/j.yhbeh.2016.04.010

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, *20*(3), 143–149. https://doi.org/10.1016/j.tree.2005.01.004

Behrens, T. E. J. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Belsky, J., & Pluess, M. (2009). Beyond Diathesis Stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*(6), 885–908. https://doi.org/10.1037/a0017376

Berghänel, A., Heistermann, M., Schülke, O., & Ostner, J. (2016). Prenatal stress effects in a wild, long-lived primate: Predictive adaptive responses in an unpredictable environment. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1839), 20161304. https://doi.org/10.1098/rspb.2016.1304

Bergman, T. J., & Beehner, J. C. (2022). Leveling with Tinbergen: Four levels simplified to causes and consequences. *Evolutionary Anthropology: Issues, News, and Reviews*, *31*(1), 12–19. https://doi.org/10.1002/evan.21931

Bernardi, L., & Hutter, I. (2007). The anthropological demography of Europe. *Demographic Research*, *17*(December), 541–566. https://doi.org/10.4054/DemRes.2007.17.18

Bierbach, D., Laskowski, K. L., & Wolf, M. (2017). Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nature Communications*, *8*(1), 15361. https://doi.org/10.1038/ncomms15361

Bitter, M. C., Wong, J. M., Dam, H. G., Donelan, S. C., Kenkel, C. D., Komoroske, L. M., Nickols, K. J., Rivest, E. B., Salinas, S., Burgess, S. C., & Lotterhos, K. E. (2021). Fluctuating selection and global change: A synthesis and review on disentangling the roles of climate amplitude, predictability and novelty. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1957), 20210727. https://doi.org/10.1098/rspb.2021.0727

Bjorklund, D. F. (1997). The role of immaturity in human development. *Psychological Bulletin*, *122*(2), 153–169. https://doi.org/10.1037/0033-2909.122.2.153

Bjorklund, D. F., & Ellis, B. J. (2014). Children, childhood, and development in evolutionary perspective. *Developmental Review*, *34*(3), 225–264. https://doi.org/10.1016/j.dr.2014.05.005

Blakemore, S.-J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology*, *65*(1), 187–207. https://doi.org/10.1146/annurev-psych-010213-115202

Bolhuis, J. J. (1991). Mechanisms of avian imprinting: A review. *Biological Reviews*, *66*(4), 303–345. https://doi.org/10.1111/j.1469-185X.1991.tb01145.x

Bornstein, M. H. (1989). *Sensitive periods in development: Structural characteristics and causal interpretations*. *105*(2), 179–197.

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620969647

Botero, C. A., Dor, R., McCain, C. M., & Safran, R. J. (2014). Environmental harshness is positively correlated with intraspecific divergence in mammals and birds. *Molecular Ecology*, *23*(2), 259–268. https://doi.org/10.1111/mec.12572

Botero, C. A., Weissing, F. J., Wright, J., & Rubenstein, D. R. (2015). Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(1), 184–189. https://doi.org/10.1073/pnas.1408589111

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Bradshaw, A. D. (1965). Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, *13*(1), 115–155.

Breed, M. D., & Moore, J. (2015). Chapter 3-Behavioral Genetics. *Animal Behavior. Academic Press. Https://Doi. Org/10.1016/C2013-0-14008-1*.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687. https://doi.org/10.1038/nature18315

Brumbach, B. H., Figueredo, A. J., & Ellis, B. J. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies. *Human Nature*, *20*(1), 25–51. https://doi.org/10.1007/s12110-009-9059-3

Bullinaria, J. A. (2003). From biological models to the evolution of robot control systems. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *361*(1811), 2145–2164. https://doi.org/10.1098/rsta.2003.1249

Burgess, S. C., & Marshall, D. J. (2011). Temperature-induced maternal effects and environmental predictability. *Journal of Experimental Biology*, *214*(14), 2329–2336. https://doi.org/10.1242/jeb.054718

Burgess, S. C., & Marshall, D. J. (2014). Adaptive parental effects: The importance of estimating environmental predictability and offspring fitness appropriately. *Oikos*, *123*(7), 769–776. https://doi.org/10.1111/oik.01235

Buwalda, B., Geerdink, M., Vidal, J., & Koolhaas, J. M. (2011). Social behavior and social stress in adolescence: A focus on animal models. *Neuroscience & Biobehavioral Reviews*, *35*(8), 1713–1721. https://doi.org/10.1016/j.neubiorev.2010.10.004

Callaghan, B. L., & Tottenham, N. (2016). The stress acceleration hypothesis: Effects of early-life adversity on emotion circuits and behavior. *Current Opinion in Behavioral Sciences*, *7*, 76–81. https://doi.org/10.1016/j.cobeha.2015.11.018

Callahan, H. S., Pigliucci, M., & Schlichting, C. D. (1997). Developmental phenotypic plasticity: Where ecology and evolution meet molecular biology. *BioEssays*, *19*(13), 519–525.

Carroll, S. P., & Corneli, P. S. (1995). Divergence in male mating tactics between two populations of the soapberry bug: II. Genetic change and the evolution of a plastic reaction norm in a variable social environment. *Behavioral Ecology*, *6*(1), 46–56. https://doi.org/10.1093/beheco/6.1.46

Chang, E. F., & Merzenich, M. M. (2003). Environmental noise retards auditory cortical development. *Science*, *300*(5618), 498–502. https://doi.org/10.1126/science.1082163

Chen, A. N., Meliza, X. C. D., & Daniel Meliza, C. (2020). Experience- and sex-dependent intrinsic plasticity in the zebra finch auditory cortex during song memorization. *Journal of Neuroscience*, *40*(10), 2047–2055. https://doi.org/10.1523/JNEUROSCI.2137-19.2019

Chevin, L. M., & Lande, R. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology*, *24*(7), 1462–1476. https://doi.org/10.1111/j.1420-9101.2011.02279.x

Chua, K. J., Lukaszewski, A. W., Grant, D. M. M., & Sng, O. (2017). Human life history strategies: Calibrated to external or internal cues? *Evolutionary Psychology*, *15*(1), 1–16. https://doi.org/10.1177/1474704916677342

Chumbley, J., & Steinhoff, A. (2019). A computational perspective on social attachment. *Infant Behavior and Development*, *54*(July 2018), 85–98. https://doi.org/10.1016/j.infbeh.2018.12.001

Conroy-Beam, D., Goetz, C. D., & Buss, D. M. (2016). What predicts romantic relationship satisfaction and mate retention intensity: Mate preference fulfillment or mate value discrepancies? *Evolution and Human Behavior*, *37*(6), 440–448. https://doi.org/10.1016/j.evolhumbehav.2016.04.003

Creanza, N., Fogarty, L., & Feldman, M. W. (2016). Cultural niche construction of repertoire size and learning strategies in songbirds. *Evolutionary Ecology*, *30*(2), 285–305. https://doi.org/10.1007/s10682-015-9796-1

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Dahl, R. E. (2004). Adolescent brain development: A period of vulnerabilities and opportunities. Keynote address. *Annals of the New York Academy of Sciences*, *1021*(1), 1–22.

Dall, S. R. X., Giraldeau, L. A., Olsson, O., McNamara, J. M., & Stephens, D. W. (2005). Information and its use by animals in evolutionary ecology. *Trends in Ecology and Evolution*, *20*(4), 187–193. https://doi.org/10.1016/j.tree.2005.01.010

Dall, S. R. X., McNamara, J. M., & Leimar, O. (2015). Genes as cues: Phenotypic integration of genetic and epigenetic information from a Darwinian perspective. *Trends in Ecology and Evolution*, *30*(6), 327–333. https://doi.org/10.1016/j.tree.2015.04.002

De Courson, B., & Nettle, D. (2021). Why do inequality and deprivation produce high crime and low trust? *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-020-80897-8

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*(OCT), 1–19. https://doi.org/10.3389/fpsyg.2017.01849

Del Giudice, M. (2014). Life history plasticity in humans: The predictive value of early cues depends on the temporal structure of the environment. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780). https://doi.org/10.1098/rspb.2013.2222

Del Giudice, M., & Belsky, J. (2010). The development of life history strategies: Toward a multi-stage theory. In *The Evolution of Personality and Individual Differences* (pp. 154–176). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195372090.003.0006

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience and Biobehavioral Reviews*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

DePasquale, C. E., Donzella, B., & Gunnar, M. R. (2018). Pubertal recalibration of cortisol reactivity following early life stress: a cross-sectional analysis. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.12992

DeWitt, T. J., & Scheiner, S. M. (2004). *Phenotypic plasticity: Functional and conceptual approaches*. Oxford University Press.

DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Cost and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, *13*(97), 77–81. https://doi.org/10.1111/j.1558-5646.2009.00647.x

Dezhbakhsh, H., & Levy, D. (1994). Periodic properties of interpolated time series. *Economics Letters*, *44*(3), 221–228.

Douhard, M., Loe, L. E., Stien, A., Bonenfant, C., Irvine, R. J., Veiberg, V., Ropstad, E., & Albon, S. (2016). The influence of weather conditions during gestation on life histories in a wild Arctic ungulate. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1841), 20161760. https://doi.org/10.1098/rspb.2016.1760

Dridi, S., & Lehmann, L. (2016). Environmental complexity favors the evolution of learning. *Behavioral Ecology*, *27*(3), 842–850. https://doi.org/10.1093/beheco/arv184

Dunlap, A. S., & Stephens, D. W. (2009). Components of change in the evolution of learning and unlearned preference. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1670), 3201–3208. https://doi.org/10.1098/rspb.2009.0602

Dunlap, A. S., & Stephens, D. W. (2016). Reliability, uncertainty, and costs in the evolution of animal learning. *Current Opinion in Behavioral Sciences*, *12*, 73–79. https://doi.org/10.1016/j.cobeha.2016.09.010

Dunn, E. C., Soare, T. W., Zhu, Y., Simpkin, A. J., Suderman, M. J., Klengel, T., Smith, A. D. A. C. A. C., Ressler, K. J., & Relton, C. L. (2019). Sensitive periods for the effect of childhood adversity on dna methylation: Results from a prospective, longitudinal study. *Biological Psychiatry*, *85*(10), 838–849. https://doi.org/10.1016/j.biopsych.2018.12.023

Ellefsen, K. O. (2013). *Evolved Sensitive Periods in Learning*. *2005*, 409–416. https://doi.org/10.7551/978-0-262-31709-2-ch059

Ellis, B. J., & Boyce, W. T. (2008). Biological sensitivity to context. *Current Directions in Psychological Science*, *17*(3), 183–187. https://doi.org/10.1111/j.1467-8721.2008.00571.x

Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary- neurodevelopmental theory. *Development and Psychopathology*, *23*(1), 7–28. https://doi.org/10.1017/S0954579410000611

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk: The impact of harsh versus unpredictable environments on the evolution and development of life history strategies. In *Human Nature* (Vol. 20, Issue 2). https://doi.org/10.1007/s12110-009-9063-7

English, S., & Barreaux, A. M. (2020). The evolution of sensitive periods in development: Insights from insects. *Current Opinion in Behavioral Sciences*, *36*, 71–78. https://doi.org/10.1016/j.cobeha.2020.07.009

English, S., Fawcett, T. W., Higginson, A. D., Trimmer, P. C., & Uller, T. (2016). Adaptive use of information during growth can explain long-term effects of early life experiences. *American Naturalist*, *187*(5), 620–632. https://doi.org/10.1086/685644

Enquist, M., Lind, J., & Ghirlanda, S. (2016). The power of associative learning and the ontogeny of optimal behaviour. *Royal Society Open Science*, *3*(11), 160734. https://doi.org/10.1098/rsos.160734

Epstein, J. M. (2008). Why Model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12. https://www.jasss.org/11/4/12.html

Erdogan, E., Ma, S., Beygelzimer, A., & Rish, I. (2005). Statistical models for unequally spaced time series. *Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005*, 626–630. https://doi.org/10.1137/1.9781611972757.74

Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E. W., Trimmer, P. C., & McNamara, J. M. (2014). The evolution of decision rules in complex environments. *Trends in Cognitive Sciences*, *18*(3), 153–161. https://doi.org/10.1016/j.tics.2013.12.012

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1), S3. https://doi.org/10.1186/1742-9994-12-S1-S3

Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology*, *24*(1), 2–11. https://doi.org/10.1093/beheco/ars085

Fawcett, T. W., & Johnstone, R. A. (2003). Optimal assessment of multiple cues. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1524), 1637–1643. https://doi.org/10.1098/rspb.2003.2328

Fenneman, J., & Frankenhuis, W. E. (2020). Is impulsive behavior adaptive in harsh and unpredictable environments? A formal model. *Evolution and Human Behavior*, *41*(4), 261–273. https://doi.org/10.1016/j.evolhumbehav.2020.02.005

Ferrari, M. C. O., Wisenden, B. D., & Chivers, D. P. (2010). Chemical ecology of predator-prey interactions in aquatic ecosystems: A review and prospectus. *Canadian Journal of Zoology*, *88*(7), 698–724. https://doi.org/10.1139/Z10-029

Fischer, B., van Doorn, G. S., Dieckmann, U., & Taborsky, B. (2014). The evolution of age-dependent plasticity. *The American Naturalist*, *183*(1), 108–125. https://doi.org/10.1086/674008

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Forbes, E. E., & Dahl, R. E. (2010). Pubertal development and behavior: Hormonal activation of social and motivational tendencies. *Brain and Cognition*, *72*(1), 66–72. https://doi.org/10.1016/j.bandc.2009.10.007

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review*, *112*(1), 60–74. https://doi.org/10.1037/0033-295X.112.1.60

Frankenhuis, W. E., & Del Giudice, M. (2012). When do adaptive developmental mechanisms yield maladaptive outcomes? *Developmental Psychology*, *48*(3), 628–642. https://doi.org/10.1037/a0025629

Frankenhuis, W. E., & Fraley, R. C. (2017). What do evolutionary models teach us about sensitive periods in psychological development? *European Psychologist*, *22*(3), 141–150. https://doi.org/10.1027/1016-9040/a000265

Frankenhuis, W. E., & Nettle, D. (2020). Integration of plasticity research across disciplines. *Current Opinion in Behavioral Sciences*, *36*, 157–162. https://doi.org/10.1016/j.cobeha.2020.10.012

Frankenhuis, W. E., Nettle, D., Dall, S. R. X., Nettle, D., & Frankenhuis, W. E. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Frankenhuis, W. E., Nettle, D., & Mcnamara, J. M. (2018). Echoes of early life: Recent insights from mathematical modeling. *Child Development*, *89*(5), 1504–1518. https://doi.org/10.1111/cdev.13108

Frankenhuis, W. E., & Panchanathan, K. (2011a). Individual differences in developmental plasticity may result from stochastic sampling. *Perspectives on Psychological Science*, *6*(4), 336–347. https://doi.org/10.1177/1745691611412602

Frankenhuis, W. E., & Panchanathan, K. (2011b). Balancing sampling and specialization: An adaptationist model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1724), 3558–3565. https://doi.org/10.1098/rspb.2011.0055

Frankenhuis, W. E., Panchanathan, K., & Barto, A. G. (2019). Enriching behavioral ecology with reinforcement learning methods. *Behavioural Processes*, *161*(August 2017), 94–100. https://doi.org/10.1016/j.beproc.2018.01.008

Frankenhuis, W. E., Panchanathan, K., & Belsky, J. (2016). A mathematical model of the evolution of individual differences in developmental plasticity arising through parental bet-hedging. *Developmental Science*, *19*(2), 251–274. https://doi.org/10.1111/desc.12309

Frankenhuis, W. E., Panchanathan, K., & Clark Barrett, H. (2013). Bridging developmental systems theory and evolutionary psychology using dynamic optimization. *Developmental Science*, *16*(4), 584–598. https://doi.org/10.1111/desc.12053

Frankenhuis, W. E., & Tiokhin, L. (2018). Bridging evolutionary biology and developmental psychology: Toward an enduring theoretical infrastructure. *Child Development*, *89*(6), 2303–2306. https://doi.org/10.1111/cdev.13021

Frankenhuis, W. E., & Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience*, *41*(April 2019), 100715. https://doi.org/10.1016/j.dcn.2019.100715

French, R. M., Mermillod, M., Chauvin, A., Quinn, P. C., & Mareschal, D. (2002). The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity. In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 322–327). Routledge. https://doi.org/10.4324/9781315782379-92

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Kruger, A., Sachser, N., Lindenberger, U., & Kempermann, G. (2013). Emergence of individuality in genetically identical mice. *Science*, *340*(6133), 756–759. https://doi.org/10.1126/science.1235294

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Fuhrmann, D., Knoll, L. J., & Blakemore, S. J. (2015). Adolescence as a sensitive period of brain development. *Trends in Cognitive Sciences*, *19*(10), 558–566. https://doi.org/10.1016/j.tics.2015.07.008

Gabard-Durnam, L. J., & McLaughlin, K. A. (2019). Do sensitive periods exist for exposure to adversity? *Biological Psychiatry*, *85*(10), 789–791. https://doi.org/10.1016/j.biopsych.2019.03.975

Gabard-durnam, L., & Mclaughlin, K. A. (2020). Sensitive periods in human development: Charting a course for the future. *Current Opinion in Behavioral Sciences*, *36*, 120–128. https://doi.org/10.1016/j.cobeha.2020.09.003

Galván, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping*, *31*(6), 879–890. https://doi.org/10.1002/hbm.21029

Gee, D. G. (2022). When do sensitive periods emerge later in development? *Trends in Cognitive Sciences*, *26*(2), 97–98. https://doi.org/10.1016/j.tics.2021.12.001

Gee, D. G., Gabard-Durnam, L. J., Flannery, J., Goff, B., Humphreys, K. L., Telzer, E. H., Hare, T. A., Bookheimer, S. Y., & Tottenham, N. (2013). Early developmental emergence of human amygdala-prefrontal connectivity after maternal deprivation. *Proceedings of the National Academy of Sciences*, *110*(39), 15638–15643. https://doi.org/10.1073/pnas.1307893110

Gibson, E. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*(1), 1–41. https://doi.org/10.1146/annurev.psych.39.1.1

Gilbert, S. F. (2003). *Developmental biology 7th edition.* Sinauer associates inc. Sunderland, Massachusetts.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1803), 20190502. https://doi.org/10.1098/rstb.2019.0502

Gottlieb, G. (1991). Experiential canalization of behavioral development: Results. *Developmental Psychology*, *27*(1), 35–39. https://doi.org/10.1037/0012-1649.27.1.35

Grainger, T. N., Senthilnathan, A., Ke, P. J., Barbour, M. A., Jones, N. T., Delong, J. P., Otto, S. P., O'connor, M. I., Coblentz, K. E., Goel, N., Sakarchi, J., Szojka, M. C., Levine, J. M., & Germain, R. M. (2021). An empiricist's guide to using ecological theory. *American Naturalist*, *199*(1). https://doi.org/10.1086/717206

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development*, *58*(3), 539. https://doi.org/10.2307/1130197

Groothuis, T. G. G., & Taborsky, B. (2015). Introducing biological realism into the study of developmental plasticity in behaviour. *Frontiers in Zoology*, *12*(1), S6. https://doi.org/10.1186/1742-9994-12-S1-S6

Grosbois, V., Gimenez, O., Gaillard, J.-M., Pradel, R., Barbraud, C., Clobert, J., Møller, A. P., & Weimerskirch, H. (2008). Assessing the impact of climate variation on survival in vertebrate populations. *Biological Reviews*, *83*(3), 357–399. https://doi.org/10.1111/j.1469-185X.2008.00047.x

Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, *9*(6), 1614–1625. https://doi.org/10.1111/2041-210X.12983

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620970585

Haltigan, J. D., Roisman, G. I., & Fraley, R. C. (2013). The predictive significance of early caregiving experiences for symptoms of psychopathology through midadolescence: Enduring or transient effects? *Development and Psychopathology*, *25*(1), 209–221. https://doi.org/10.1017/S0954579412000260

Hammel, E. A. (2005). Demographic dynamics and kinship in anthropological populations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 2248–2253. https://doi.org/10.1073/pnas.0409762102

Hartman, S., Li, Z., Nettle, D., & Belsky, J. (2017). External-environmental and internal-health early life predictors of adolescent development. *Development and Psychopathology*, *29*(5), 1839–1849. https://doi.org/10.1017/S0954579417001432

Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*, *0*(0), 1–32. https://doi.org/10.1080/00273171.2021.1896353

Hatzenbuehler, M. L., Weissman, D. G., McKetta, S., Lattanner, M. R., Ford, J. V., Barch, D. M., & McLaughlin, K. A. (2021). Smaller hippocampal volume among Black and Latinx youth living in high-stigma contexts. *Journal of the American Academy of Child & Adolescent Psychiatry*, 1–11. https://doi.org/10.1016/j.jaac.2021.08.017

Haynes, K., Killick, R., Fearnhead, P., & Eckley, I. (2016). changepoint. np: Methods for nonparametric changepoint detection. *R Package Version 0.0*, *2*.

Hensch, T. K. (2004). Critical period regulation. *Annual Review of Neuroscience*, *27*(1), 549–579. https://doi.org/10.1146/annurev.neuro.27.070203.144327

Hensch, T. K., & Bilimoria, P. M. (2012). Re-opening windows: Manipulating critical periods for brain development. *Cerebrum : The Dana Forum on Brain Science*, *2012*(August), 11.

Hirano, Y., Masuda, T., Naganos, S., Matsuno, M., Ueno, K., Miyashita, T., Horiuchi, J., & Saitoe, M. (2013). Fasting launches crtc to facilitate long-term memory formation in Drosophila. *Science*, *339*(6118), 443–446. https://doi.org/10.1126/science.1227170

Hodson, G. (2021). Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *Journal of Theoretical Social Psychology*, *00*(August), 1–15. https://doi.org/10.1002/jts5.120

Hopwood, C. J. ., Bleidorn, W., & Wright, A. G. C. (2021). Connecting theory to methods in longitudinal research. *Perspectives on Psychological Science*, 174569162110084. https://doi.org/10.1177/17456916211008407

Horn, M. E., Ferrari, M. C. O., Douglas, P., Chivers, D. P., Douglas, P., Chivers, D. P., & Douglas, P. (2019). Retention of learned predator recognition in embryonic and juvenile rainbow trout. *Behavioral Ecology*, 1–8. https://doi.org/10.1093/beheco/arz116

Houston, A., Clark, C., Mcnamara, J. M., & Mangel, M. (1988). Dynamic models in behavioural and evolutionary ecology. *Nature*, *332*(6159), 29–34. https://doi.org/10.1038/332029a0

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *The Journal of Physiology*, *206*(2), 419–436.

Humphreys, K. L., King, L. S., Sacchet, M. D., Camacho, C., Colich, N. L., Ordaz, S. J., Ho, T. C., Gotlib, I. H., Camacho, M. C., Colich, N. L., Ordaz, S. J., Ho, T. C., Gotlib, I. H., Camacho, C., Colich, N. L., Ordaz, S. J., Ho, T. C., Gotlib, I. H., Camacho, M. C., … Gotlib, I. H. (2019). Evidence for a sensitive period in the effects of early life stress on hippocampal volume. *Developmental Science*, *22*(3), 1–10. https://doi.org/10.1111/desc.12775

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Innes-Gold, A. A., Zuczek, N. Y., & Touchon, J. C. (2019). Right phenotype, wrong place: Predator-induced plasticity is costly in a mismatched environment. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1916), 20192347. https://doi.org/10.1098/rspb.2019.2347

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, *345*, 12–26. https://doi.org/10.1016/j.neuroscience.2016.03.021

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*(JUN), 1–24. https://doi.org/10.3389/fpsyg.2015.00727

Johnson, M. H. (2005). Sensitive periods in functional brain development: Problems and prospects. *Developmental Psychobiology*, *46*(3), 287–292. https://doi.org/10.1002/dev.20057

Jones, R. H. (1984). Fitting multivariate models to unequally spaced data. In *Time series analysis of irregularly observed data* (pp. 158–188). Springer.

Jury, C. P., Toonen, R. J., & Jury, C. P. (2019). Adaptive responses and local stressor mitigation drive coral resilience in warmer, more acidic oceans. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1902). https://doi.org/10.1098/rspb.2019.0614

Kacelnik, A. (2012). Putting mechanisms into behavioral ecology. In P. Hammerstein & R. Stevens (Eds.), *Evolution and the mechanisms of decision making* (pp. 21–38). Cambridge, MA: MIT Press.

Kaler, S. R., & Freeman, B. J. (1994). Analysis of environmental deprivation: Cognitive and social development in Romanian orphans. *Journal of Child Psychology and Psychiatry*, *35*(4), 769–781.

Kelly, T. K., Ahmadiantehrani, S., Blattler, A., & London, S. E. (2018). Epigenetic regulation of transcriptional plasticity associated with developmental song learning. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1878), 20180160. https://doi.org/10.1098/rspb.2018.0160

Kievit, R. A., McCormick, E. M., Fuhrmann, D., Deserno, M. K., & Orben, A. (2021). Using large, publicly available datasets to study adolescent development: Opportunities and challenges. *Current Opinion in Psychology*, *44*, 303–308. https://doi.org/10.1016/j.copsyc.2021.10.003

Killick, R., & Eckley, I. A. (2014). changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, *58*(3).

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S. J. (2016). A window of opportunity for cognitive training in adolescence. *Psychological Science*, *27*(12), 1620–1631. https://doi.org/10.1177/0956797616671327

Knudsen, E. I. (2004). *Sensitive Periods in the Development of the Brain and Behavior*. 1412–1425. https://doi.org/10.1162/0898929042304796

Kok, E. M. A., Burant, J. B., Dekinga, A., Manche, P., Saintonge, D., Piersma, T., & Mathot, K. J. (2019). Within-individual canalization contributes to age-related increases in trait repeatability: A longitudinal experiment in red knots. *The American Naturalist*, *194*(4), 000–000. https://doi.org/10.1086/704593

Kuijper, B., Johnstone, R. A., & Kuijper, B. (2019). The evolution of early-life effects on social behaviour—why should social adversity carry over to the future? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 1–18. https://doi.org/10.1098/rstb.2018.0111

Kuzawa, C. W. (2005). Fetal origins of developmental plasticity: Are fetal cues reliable predictors of future nutritional environments? *American Journal of Human Biology*, *17*(1), 5–21. https://doi.org/10.1002/ajhb.20091

Lande, R. (2014). Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative character in a fluctuating environment. *Journal of Evolutionary Biology*, *27*(5), 866–875. https://doi.org/10.1111/jeb.12360

Lande, R. (2019). Developmental integration and evolution of labile plasticity in a complex quantitative character in a multiperiodic environment. *Proceedings of the National Academy of Sciences of the United States of America*, *166*(23), 11361–11369. https://doi.org/10.1073/pnas.1900528116

Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews*, *94*(March), 179–195. https://doi.org/10.1016/J.NEUBIOREV.2018.09.005

Law, R. (1979). Optimal life histories under age-specific predation. *The American Naturalist*, *114*(3), 399–417.

Leung, C., Rescan, M., Grulois, D., & Chevin, L. (2020). Reduced phenotypic plasticity evolves in less predictable environments. *Ecology Letters*, *23*(11), 1664–1672. https://doi.org/10.1111/ele.13598

Levins, R. (1966). The strategy of model building in population biology. In *American Naturalist* (Vol. 54, Issue 4, pp. 421–431). https://doi.org/10.2307/27836590

Li, Z., Liu, S., Hartman, S., & Belsky, J. (2018). Interactive effects of early-life income harshness and unpredictability on children's socioemotional and academic functioning in kindergarten and adolescence. *Developmental Psychology*. https://doi.org/10.1037/dev0000601

Lickliter, R., & Honeycutt, H. (2003). *Developmental Dynamics: Toward a Biologically Plausible Evolutionary Psychology*. *129*(6), 819–835. https://doi.org/10.1037/0033-2909.129.6.819

Lim, S.-S. (2016). Towards consilience in science of learning: Data as currency for collaboration. In *npj Science of Learning* (Vol. 1, Issue 1). https://doi.org/10.1038/npjscilearn.2016.9

Lin, W. C., Delevich, K., & Wilbrecht, L. (2020). A role for adaptive developmental plasticity in learning and decision making. *Current Opinion in Behavioral Sciences*, *36*, 48–54. https://doi.org/10.1016/j.cobeha.2020.07.010

Luby, J. L., Barch, D. M., Belden, A., Gaffrey, M. S., Tillman, R., Babb, C., Nishino, T., Suzuki, H., & Botteron, K. N. (2012). Maternal support in early childhood predicts larger hippocampal volumes at school age. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(8), 2854–2859. https://doi.org/10.1073/pnas.1118003109

Luby, J. L., Belden, A., Harms, M. P., Tillman, R., & Barch, D. M. (2016). Preschool is a sensitive period for the influence of maternal support on the trajectory of hippocampal development. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(20), 5742–5747. https://doi.org/10.1073/pnas.1601443113

Mameli, M., & Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1563), 436–443. https://doi.org/10.1098/rstb.2010.0174

Mangel, M. (1990). Dynamic information in uncertain and changing worlds. *Journal of Theoretical Biology*, *146*(3), 317–332.

Mangel, M., & Clark, C. W. (2019). *Dynamic modeling in behavioral ecology*. Princeton University Press.

Marcus, G. (2009). *Kluge: The haphazard evolution of the human mind*. Houghton Mifflin Harcourt.

Mariette, M. M. (2020). Acoustic developmental programming: Implications for adaptive plasticity and the evolution of sensitive periods. *Current Opinion in Behavioral Sciences*, *36*, 129–134. https://doi.org/10.1016/j.cobeha.2020.09.010

Marshall, D. J., & Burgess, S. C. (2015). Deconstructing environmental predictability: Seasonality, environmental colour and the biogeography of marine life histories. *Ecology Letters*, *18*(2), 174–181. https://doi.org/10.1111/ele.12402

Martínez, A., & Mammola, S. (2021). Specialized terminology reduces the number of citations of scientific papers. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1948), rspb.2020.2581. https://doi.org/10.1098/rspb.2020.2581

Mascie-Taylor, C. G. N., & Little, M. A. (2004). History of migration studies in biological anthropology. *American Journal of Human Biology*, *16*(4), 365–378. https://doi.org/10.1002/ajhb.20046

Matsumura, K., Yumise, K., Fujii, Y., Hayashi, T., & Miyatake, T. (2020). Anti-predator behaviour depends on male weapon size. *Biology Letters*, *16*(12), 20200601. https://doi.org/10.1098/rsbl.2020.0601

Maugarny-Calès, A., & Laufs, P. (2018). Getting leaves into shape: A molecular, cellular, environmental and evolutionary view. *Development*, *145*(13). https://doi.org/10.1242/dev.161646

McCullough, E. L., & Simmons, L. W. (2016). Selection on male physical performance during male–male competition and female choice. *Behavioral Ecology*, *27*(5), 1288–1295. https://doi.org/10.1093/beheco/arw033

McLaughlin, K. A., Sheridan, M. A., & Nelson, C. A. (2017). Neglect as a violation of species-expectant experience: Neurodevelopmental consequences. *Biological Psychiatry*, *82*(7), 462–471. https://doi.org/10.1016/j.biopsych.2017.02.1096

McNamara, J. M., Dall, S. R. X., Hammerstein, P., & Leimar, O. (2016). Detection vs. selection: Integration of genetic, epigenetic and environmental cues in fluctuating environments. *Ecology Letters*, *19*(10), 1267–1276. https://doi.org/10.1111/ele.12663

McNamara, J. M., Green, R. F., & Olsson, O. (2006). Bayes' theorem and its applications in animal behaviour. *Oikos*, *112*(2), 243–251. https://doi.org/10.1111/j.0030-1299.2006.14228.x

Mcnamara, J. M., & Houston, A. (1980). The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, *85*(4), 673–690.

McNamara, J. M., & Houston, A. I. (2009). Integrating function and mechanism. *Trends in Ecology and Evolution*, *24*(12), 670–675. https://doi.org/10.1016/j.tree.2009.05.011

Mendoza, J. K., & Fausey, C. M. (2013). Quanitfying everyday ecologies: Principles for manual annotation of many hours of infants' lives. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.3389/fpsyg.2021.710636

Mery, F., & Kawecki, T. J. (2003). A fitness cost of learning ability in Drosophila melanogaster. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1532), 2465–2469. https://doi.org/10.1098/rspb.2003.2548

Mery, F., & Kawecki, T. J. (2004). An operating cost of learning in Drosophila melanogaster. *Animal Behaviour*, *68*(3), 589–598. https://doi.org/10.1016/j.anbehav.2003.12.005

Mery, F., & Kawecki, T. J. (2005). A cost of long-term memory in Drosophila. *Science*, *308*(5725), 1148–1148. https://doi.org/10.1126/science.1111331

Michel, G. F., & Tyler, A. N. (2005). Critical period: A history of the transition from questions of when, to what, to how. *Developmental Psychobiology*, *46*(3), 156–162. https://doi.org/10.1002/dev.20058

Miller, G. E., Chen, E., Armstrong, C. C., Carroll, A. L., Ozturk, S., Rydland, K. J., Brody, G. H., Parrish, T. B., & Nusslock, R. (2018). Functional connectivity in central executive network protects youth against cardiometabolic risks linked with neighborhood violence. *Proceedings of the National Academy of Sciences*, *115*(47), 12063–12068. https://doi.org/10.1073/pnas.1810067115

Moczek, E. C. S.-R. J. D. V. D. T. C. M. J. W. A. P., Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J., & Moczek, A. P. (2010). Toward a population genetic framework of developmental evolution: The costs, limits, and consequences of phenotypic plasticity. *BioEssays*, *32*(1), 71–81. https://doi.org/10.1002/bies.200900132

Moscarello, J. M., & Hartley, C. A. (2017). Agency and the calibration of motivated behavior. *Trends in Cognitive Sciences*, *21*(10), 725–735. https://doi.org/10.1016/j.tics.2017.06.008

Mountjoy, J. D., & Lemon, R. E. (1995). Extended song learning in wild European starlings. *Animal Behaviour*, *49*(2), 357-366.

Mueller, C. A. (2018). Critical windows in animal development: Interactions between environment, phenotype, and time. *Development and Environment*, 41–72. https://doi.org/10.1007/978-3-319-75935-7_3

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Mutwill, A. M., Zimmermann, T. D., Hennicke, A., Richter, S. H., Kaiser, S., & Sachser, N. (2020). Adaptive reshaping of the hormonal phenotype after social niche transition in adulthood. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1928), 20200667. https://doi.org/10.1098/rspb.2020.0667

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society of London B: Biological Sciences*, *1*(2), 20131343. https://doi.org/Artn 20131343\rDoi 10.1098/Rspb.2013.1343

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2014). The evolution of predictive adaptive responses in humans: Response. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), 20132822. https://doi.org/10.1098/rspb.2013.2822

Nijhout, H. F. (2015). A developmental-physiological perspective on the development and evolution of phenotypic plasticity. In *Conceptual change in biology* (pp. 147–173). Springer.

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, *8*(1), 25–35. https://doi.org/10.1177/2167702619865969

Otto, S. P., & Rosales, A. (2020). Theory in service of narratives in evolution and ecology. *The American Naturalist, 195*(2), 290–299. https://doi.org/10.1086/705991

Pallier, C. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, *13*(2), 155–161. https://doi.org/10.1093/cercor/13.2.155

Panchanathan, K., & Frankenhuis, W. E. (2016). The evolution of sensitive periods in a model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152439. https://doi.org/10.1098/rspb.2015.2439

Pascalis, O., Fort, M., & Quinn, P. C. (2020). Development of face processing: Are there critical or sensitive periods? *Current Opinion in Behavioral Sciences*, *36*, 7–12. https://doi.org/10.1016/j.cobeha.2020.05.005

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the links among face processing, language processing, and narrowing during development. *Child Development Perspectives*, *8*(2), 65–70. https://doi.org/10.1111/cdep.12064

Pavía-Miralles, J. M. (2010). A survey of methods to interpolate, distribute and extrapolate time series. *Journal of Service Science and Management*, *03*(04), 449–463. https://doi.org/10.4236/jssm.2010.34051

Pfab, F., Gabriel, W., & Utz, M. (2016). Reversible phenotypic plasticity with continuous adaptation. *Journal of Mathematical Biology*, *72*(1–2), 435–466. https://doi.org/10.1007/s00285-015-0890-3

Pigliucci, M. (2005). Evolution of phenotypic plasticity: Where are we going now? *Trends in Ecology & Evolution*, *20*(9), 481–486. https://doi.org/10.1016/j.tree.2005.06.001

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Plaçais, P.-Y., & Preat, T. (2013). To favor survival under food shortage, the brain disables costly memory. *Science*, *339*(6118), 440–442. https://doi.org/10.1126/science.1226018

Pollak, S. D., & Smith, K. E. (2021). Thinking clearly about biology and childhood adversity: Next steps for continued progress. In *Perspectives on Psychological Science, 16*(6), 1473-1477. https://doi.org/10.1177/17456916211031539

Polverino, G., Palmas, B. M., Evans, J. P., & Gasparini, C. (2019). Individual plasticity in alternative reproductive tactics declines with social experience in male guppies. *Animal Behaviour*, *148*(January), 113–121. https://doi.org/10.1016/j.anbehav.2018.12.014

Raab, H. A., & Hartley, C. A. (2019). Adolescents exhibit reduced Pavlovian biases on instrumental learning. *Scientific Reports*, *0123456789*. https://doi.org/10.31234/osf.io/38vgr

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychological Aging*, *24*(4), 778–791. https://doi.org/10.1037/a0017915.Time-Structured

Ratikainen, I. I., & Kokko, H. (2019). The coevolution of lifespan and reversible plasticity. *Nature Communications*, *10*(1), 1–7. https://doi.org/10.1038/s41467-019-08502-9

Réale, D., & Dingemanse, N. J. (2012). *Animal personality*. In *eLS*. John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9780470015902.a0023570

Reh, R. K., Dias, B. G., Nelson, C. A., Kaufer, D., Werker, J. F., Kolb, B., Levine, J. D., & Hensch, T. K. (2020). Critical period regulation across multiple timescales. *Proceedings of the National Academy of Sciences*, 201820836. https://doi.org/10.1073/pnas.1820836117

Reimer, J. R., Mangel, M., Derocher, A. E., & Lewis, M. A. (2019). Matrix methods for stochastic dynamic programming in ecology and evolutionary biology. *Methods in Ecology and Evolution*, *10*(11), 1952–1961. https://doi.org/10.1111/2041-210X.13291

Relyea, R. A. (2002). Costs of phenotypic plasticity. *The American Naturalist*, *159*(3), 272–282. https://doi.org/10.1086/338540

Relyea, R. A. (2003). Predators come and predators go: The reversibility of predator-induced traits. *Ecology*, *84*(7), 1840–1848. https://doi.org/10.1890/0012-9658(2003)084[1840:PCAPGT]2.0.CO;2

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Americal Journal of Physiology Heart and Circulatory Physiology*, *278*(6), H2039–H2049.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037//0033-2909.126.1.3

Rufrancos, H., & Power, M. (2013). Income inequality and crime: A review and explanation of the time-series evidence. *Sociology and Criminology-Open Access*, *01*(01), 1–9. https://doi.org/10.4172/2375-4435.1000103

Ruokolainen, L., Lindén, A., Kaitala, V., & Fowler, M. S. (2009). Ecological and evolutionary dynamics under coloured environmental variation. *Trends in Ecology and Evolution*, *24*(10), 555–563. https://doi.org/10.1016/j.tree.2009.04.009

Sachser, N., Hennessy, M. B., & Kaiser, S. (2018). The adaptive shaping of social behavioural phenotypes during adolescence. *Biology Letters*, *14*(11), 20180536. https://doi.org/10.1098/rsbl.2018.0536

Sachser, N., Zimmermann, T. D., Hennessy, M. B., & Kaiser, S. (2020). Sensitive phases in the development of rodent social behavior. *Current Opinion in Behavioral Sciences*, *36*, 63–70. https://doi.org/10.1016/j.cobeha.2020.07.014

Samuels, R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences*, *8*(3), 136–141. https://doi.org/10.1016/j.tics.2004.01.010

Schlichting, C. D. (1986). The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics*, *17*(5), 667–693. https://doi.org/10.1093/cz/zow084

Schlichting, C. D., Pigliucci, M., & others. (1998). *Phenotypic evolution: a reaction norm perspective.* Sinauer Associates Incorporated.

Schonfeld, I. S., & Bianchi, R. (2016). Burnout and depression: Two entities or one? *Journal of Clinical Psychology*, *72*(1), 22–37. https://doi.org/10.1002/jclp.22229

Schonfeld, I. S., & Verkuilen, J. (2019). Inquiry into the correlation between burnout and depression. *Journal of Occupational Health Psychology*, *24*(6), 603–616. https://doi.org/10.1037/ocp0000151

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*(4), 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x

Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Van Cleve, J., & Yeh, D. J. (2014). Not just a theory - The utility of mathematical models in evolutionary biology. *PLoS Biology, 12*(12), e1002017. https://doi.org/10.1371/journal.pbio.1002017

Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. *Processes of Change in Brain and Cognitive Development: Attention and Performance XXI*, 315–347.

Shrager, J., & Johnson, M. H. (1996). Dynamic plasticity influences the emergence of function in a simple cortical array. *Neural Networks*, *9*(7), 1119–1129. https://doi.org/10.1016/0893-6080(96)00033-0

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: An integrative overview. *The Quarterly Review of Biology*, *79*(3), 241–277. https://doi.org/10.1086/422893

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: an evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70–82. https://doi.org/10.1109/TAMD.2010.2051031

Smaldino, P. E. (2017). Models are stupid, and we need more of them. *Computational Social Psychology*, 311–331. https://doi.org/10.4324/9781315173726

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207–218. https://doi.org/10.1027/1864-9335/a000425

Smallegange, I. M. (2011). Complex environmental effects on the expression of alternative reproductive phenotypes in the bulb mite. *Evolutionary Ecology*, *25*(4), 857–873. https://doi.org/10.1007/s10682-010-9446-6

Smith, J. E., & Pinter-Wollman, N. (2021). Observing the unwatchable: Integrating automated sensing, naturalistic observations and animal social network analysis in the age of big data. *Journal of Animal Ecology*, *90*(1), 62–75. https://doi.org/10.1111/1365-2656.13362

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Snell-Rood, E. C. (2012). Selective processes in development: Implications for the costs and benefits of phenotypic plasticity. *Integrative and Comparative Biology*, *52*(1), 31–42. https://doi.org/10.1093/icb/ics067

Snell-Rood, E. C., Kobiela, M. E., Sikkink, K. L., & Shephard, A. M. (2018). Mechanisms of plastic rescue in novel environments. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 331–354. https://doi.org/10.1146/annurev-ecolsys-110617-062622

Snell-Rood, E. C., & Steck, M. K. (2019). Behaviour shapes environmental variation and selection on learning and plasticity: Review of mechanisms and implications. *Animal Behaviour*, *147*, 147–156. https://doi.org/10.1016/j.anbehav.2018.08.007

Snyder, J. K., Fessler, D. M. T. T., Tiokhin, L., Frederick, D. A., Lee, S. W., Navarrete, C. D., Woo, S., David, C., Lee, S. W., & Navarrete, C. D. (2011). Trade-offs in a dangerous world: Women's fear of crime predicts preferences for aggressive and formidable mates. *Evolution and Human Behavior*, *32*(2), 127–137. https://doi.org/10.1016/j.evolhumbehav.2010.08.007

Sol, D., Sayol, F., Ducatez, S., & Lefebvre, L. (2016). The life-history basis of behavioural innovations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1690), 20150187. https://doi.org/10.1098/rstb.2015.0187

Sowersby, W., Eckerström-Liedholm, S., Kotrschal, A., Näslund, J., Rowiński, P., Gonzalez-Voyer, A., & Rogell, B. (2021). Fast life-histories are associated with larger brain size in killifishes. *Evolution*, *in press*, evo.14310. https://doi.org/10.1111/evo.14310

Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development predicts genotypic differences in plasticity. *Evolution*, *72*(10), 2167–2180. https://doi.org/10.1111/evo.13585

Stamps, J. A., & Frankenhuis, W. E. (2016). Bayesian models of development. *Trends in Ecology & Evolution*, *31*(4), 260–268. https://doi.org/10.1016/j.tree.2016.01.012

Stamps, J. A., & Krishnan, V. V. (2014a). Individual differences in the potential and realized developmental plasticity of personality traits. *Frontiers in Ecology and Evolution*, *2*(October). https://doi.org/10.3389/fevo.2014.00069

Stamps, J. A., & Krishnan, V. V. (2014b). Combining information from ancestors and personal experiences to predict individual differences in developmental trajectories. *The American Naturalist*, *184*(5), 647–657. https://doi.org/10.1086/678116

Stamps, J. A., & Krishnan, V. V. (2017). Age-dependent changes in behavioural plasticity: Insights from Bayesian models of development. *Animal Behaviour*, *126*, 53–67. https://doi.org/10.1016/j.anbehav.2017.01.013

Stearns, S. C. (1989). The evolutionary significance of phenotypic plasticity. *Bioscience*, *39*(7), 436–445.

Stephens, D. W. (1991). Change, regularity, and value in the evolution of animal learning. *Behavioral Ecology*, *2*(1), 77–89. https://doi.org/10.1093/beheco/2.1.77

Stimson, J. A. (1985). Regression in space and time: A statistical essay. *American Journal of Political Science*, *29*(4), 914. https://doi.org/10.2307/2111187

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences*, *105*(1), 394–398. https://doi.org/10.1073/pnas.0706079105

Sullivan, R. M., & Holman, P. J. (2010). Transitions in sensitive period attachment learning in infancy: The role of corticosterone. *Neuroscience & Biobehavioral Reviews*, *34*(6), 835–844. https://doi.org/10.1016/j.neubiorev.2009.11.010

Sunstein, C. R. (2022). On the limited policy relevance of evolutionary explanations. *SSRN Electronic Journal*, 1–12. https://doi.org/10.2139/ssrn.4003339

Taborsky, B., English, S., Fawcett, T. W., Kuijper, B., Leimar, O., McNamara, J. M., Ruuskanen, S., & Sandi, C. (2021). Towards an evolutionary theory of stress responses. *Trends in Ecology & Evolution*, *36*(1), 39–48. https://doi.org/10.1016/j.tree.2020.09.003

Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Progress in Brain Research*, *207*, 3–34.

Thomas, M. S. C., & Johnson, M. H. (2008). New advances in understanding sensitive periods in brain development. *Current Directions in Psychological Science*, *17*(1), 1–5. https://doi.org/10.1111/j.1467-8721.2008.00537.x

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift fur Tierpsychologie 20*, 410-433.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(5), 727–741. https://doi.org/10.1017/S0140525X00003447

Tooby, J., Cosmides, L., & Barrett, H. C. (2003). The second law of thermodynamics is the first law of psychology: Evolutionary developmental psychology and the theory of tandem, coordinated inheritances: Comment on lickliter and honeycutt (2003). *Psychological Bulletin*, *129*(6), 858–865. https://doi.org/10.1037/0033-2909.129.6.858

Tooley, U. A., Bassett, D. S., Mackey, A. P., & Mackey, A. (2021). Environmental influences on the pace of brain development. *Nature Reviews Neuroscience*. https://doi.org/10.1038/s41583-021-00457-5

Tottenham, N., Hare, T. A., Quinn, B. T., McCarry, T. W., Nurse, M., Gilhooly, T., Millner, A., Galvan, A., Davidson, M. C., Eigsti, I.-M., Thomas, K. M., Freed, P. J., Booma, E. S., Gunnar, M. R., Altemus, M., Aronson, J., & Casey, B. J. (2010). Prolonged institutional rearing is associated with atypically large amygdala volume and difficulties in emotion regulation. *Developmental Science*, *13*(1), 46–61. https://doi.org/10.1111/j.1467-7687.2009.00852.x

Tottenham, N., Shapiro, M., Flannery, J., Caldera, C., & Sullivan, R. M. (2019). Parental presence switches avoidance to attraction learning in children. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0656-9

Tramontin, A. D., & Brenowitz, E. A. (2000). Seasonal plasticity in the adult brain. In *Trends in Neurosciences* (Vol. 23, Issue 6, pp. 251–258). https://doi.org/10.1016/S0166-2236(00)01558-7

Trillmich, F., Müller, T., & Müller, C. (2018). Understanding the evolution of personality requires the study of mechanisms behind the development and life history of personality traits. *Biology Letters*, *14*(2), 20170740. https://doi.org/10.1098/rsbl.2017.0740

Trimmer, P. C., Barrett, B. J., McElreath, R., & Sih, A. (2019). Rapid environmental change in games: Complications and counter-intuitive outcomes. *Scientific Reports*, *9*(1), 1–6. https://doi.org/10.1038/s41598-019-43770-x

Trimmer, P. C., Houston, A. I., Marshall, J. A. R., Mendl, M. T., Paul, E. S., & McNamara, J. M. (2011). Decision-making under uncertainty: Biases and Bayesians. *Animal Cognition*, *14*(4), 465–476. https://doi.org/10.1007/s10071-011-0387-4

Trimmer, P. C., McNamara, J. M., Houston, A. I., & Marshall, J. A. R. (2012). Does natural selection favour the Rescorla–Wagner rule? *Journal of Theoretical Biology*, *302*, 39–52. https://doi.org/10.1016/j.jtbi.2012.02.014

Tufto, J. (2000). The evolution of plasticity and nonplastic spatial and temporal adaptations in the presence of imperfect environmental cues. *American Naturalist*, *156*(2), 121–130. https://doi.org/10.1086/303381

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *15*(4), 357–368.

Uller, T., English, S., & Pen, I. (2015). When is incomplete epigenetic resetting in germ cells favoured by natural selection? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1811). https://doi.org/10.1098/rspb.2015.0682

Utz, M., Jeschke, J. M., Loeschcke, V., & Gabriel, W. (2014). Phenotypic plasticity with instantaneous but delayed switches. *Journal of Theoretical Biology*, *340*, 60–72. https://doi.org/10.1016/j.jtbi.2013.08.038

van den Bos, W., & Eppinger, B. (2016). Developing developmental cognitive neuroscience: From agenda setting to hypothesis testing. *Developmental Cognitive Neuroscience*, *17*, 138–144. https://doi.org/10.1016/j.dcn.2015.12.011

van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, *16*(4), 682–697. https://doi.org/10.1177/1745691620970604

Vasseur, D. A., & Yodzis, P. (2004). The color of environmental noise. *Ecology*, *85*(4), 1146–1152. https://doi.org/10.1890/02-3122

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Van Tienderen, P. H. (1995). Adaptive phenotypic plasticity: Consensus and controversy. *Trends in Ecology & Evolution*, *10*(5), 212–217. https://doi.org/10.1016/S0169-5347(00)89061-8

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2021). An evolutionary model of sensitive periods when the reliability of cues varies across ontogeny. *Behavioral Ecology*, 33(1), 1–14. https://doi.org/10.1093/beheco/arab113

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2022). Sensitive periods, but not critical periods, evolve in a fluctuating environment: A model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1969), 20212623. https://doi.org/10.1098/rspb.2021.2623

Warlaumont, A. S., Sobowale, K., & Fausey, C. M. (2021). Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. *Current Directions in Psychological Science*, 096372142110581. https://doi.org/10.1177/09637214211058166

Warren, P., Tomaskovic-Devey, D., Smith, W., Zingraff, M., & Mason, M. (2006). Driving while black: Bias processes and racial disparity in police stops. *Criminology, 44*(3), 709–738. https://doi.org/10.1111/j.1745-9125.2006.00061.x

Weiss, L. C. (2019). Sensory ecology of predator-induced phenotypic plasticity. *Frontiers in Behavioral Neuroscience*, *12*(January), 1–12. https://doi.org/10.3389/fnbeh.2018.00330

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Whewell, W. (1840). *The Philosophy of the Inductive Sciences, Founded Ou Their History*.

Whitman, D. W., Agrawal, A. A., & others. (2009). What is phenotypic plasticity and why is it important. *Phenotypic Plasticity of Insects: Mechanisms and Consequences*, 1–63.

Wilbrecht, L. (2006). High levels of new neuron addition persist when the sensitive period for song learning is experimentally prolonged. *Journal of Neuroscience*, *26*(36), 9135–9141. https://doi.org/10.1523/JNEUROSCI.4869-05.2006

Wilson, E. O. (1999). *Consilience: The unity of knowledge* (Vol. 31). Vintage.

Wilson, S. P., & Prescott, T. J. (2022). Scaffolding layered control architectures through constraint closure: Insights into brain evolution and development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844). https://doi.org/10.1098/rstb.2020.0519

Winkler, D. W., Jørgensen, C., Both, C., Houston, A. I., Mcnamara, J. M., Levey, D. J., Partecke, J., Fudickar, A., Kacelnik, A., Roshier, D., & Piersma, T. (2014). Cues, strategies, and outcomes: How migrating vertebrates track environmental change. *Movement Ecology*, *2*(1), 1–15. https://doi.org/10.1186/2051-3933-2-10

Wolf, D. L. (2007). *Beyond Anne Frank*. University of California Press.

Wuerz, Y., & Krüger, O. (2015). Personality over ontogeny in zebra finches: Long-term repeatable traits but unstable behavioural syndromes. *Frontiers in Zoology*, *12*(1). https://doi.org/10.1186/1742-9994-12-S1-S9

Yeh, D. J., Boughman, J. W., Sætre, G. P., & Servedio, M. R. (2018). The evolution of sexual imprinting through reinforcement*. *Evolution*, *72*(7), 1336–1349. https://doi.org/10.1111/evo.13500

Young, E. S., Frankenhuis, W. E., & Ellis, B. J. (2020). Theory and measurement of environmental unpredictability. *Evolution and Human Behavior*, *February*, 0–1. https://doi.org/10.1016/j.evolhumbehav.2020.08.006

Zeanah, C. H., Gunnar, M. R., McCall, R. B., Kreppner, J. M., & Fox, N. A. (2011). VI. Sensitive periods. *Monographs of the Society for Research in Child Development*, *76*(4), 147–162.

Zhang, N. R., & Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*(1), 22–32. https://doi.org/10.1111/j.1541-0420.2006.00662.x

Supplementary materials

Supplementary materials for Chapters 3-5

Appendix 1 – Chapter 3

Additional plots for 20 time periods

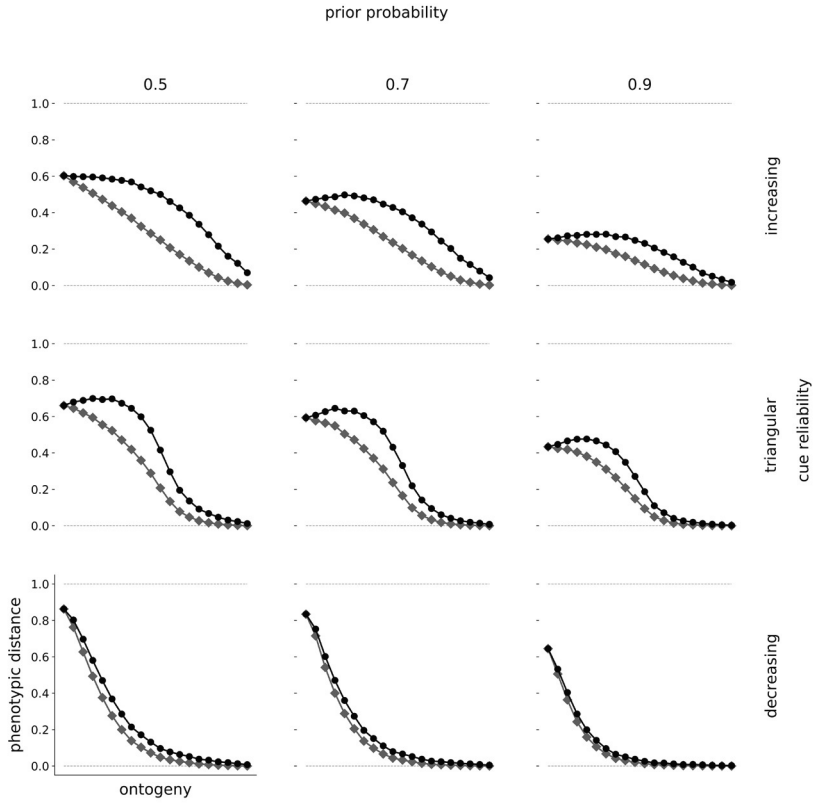


Figure A1.1 Plasticity across ontogeny – maximal cue reliability of 0.75. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance at the end of ontogeny between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance following separation.

Here, we additionally show distributions of mature phenotypes and compare the fitness of the optimal policy with the fitness of two non-plastic strategies, a generalist and a specialist, to get a sense of whether and by how much the optimal policy outperforms simpler strategies. We do not discuss results from these analyses in the main text, as the results are qualitatively similar to those of a model with fixed cue reliabilities (Panchanathan & Frankenhuis, 2016). The main text focuses on those results that are qualitatively different when cue reliabilities are variable rather than fixed across ontogeny.

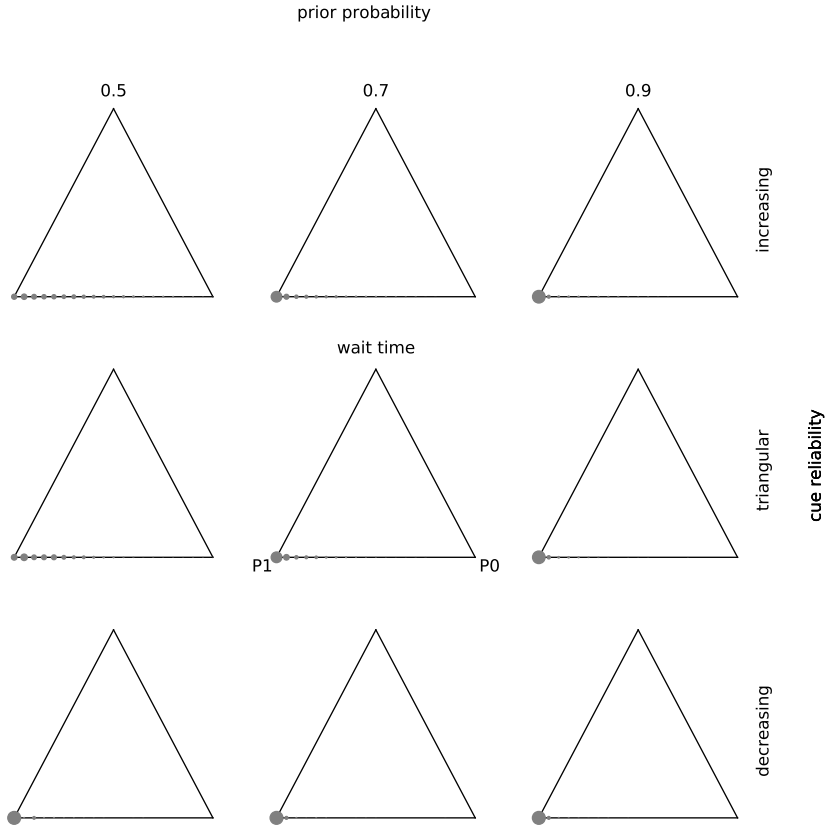


Figure A1.2 Distributions of mature phenotypes. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents a simulation study. For each study we simulate 10,000 organisms who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. Each triangle plots the distribution of phenotypes at the end of ontogeny. The number of time periods waited, time periods specialized towards P_1 and time periods specialized towards P_0 make up a phenotype. The position of a circle indicates the composition of mature phenotypes. The left and right vertices represent organisms that only specialized towards P_1 and P_0, respectively. The top vertex represents organisms that only waited. Circles on the outer boundary indicate a mixture of two phenotypic decisions, while circles within the triangle indicate a mixture of all three decisions. The area of a circle is proportional to the fraction of simulated organisms that developed the same phenotype.

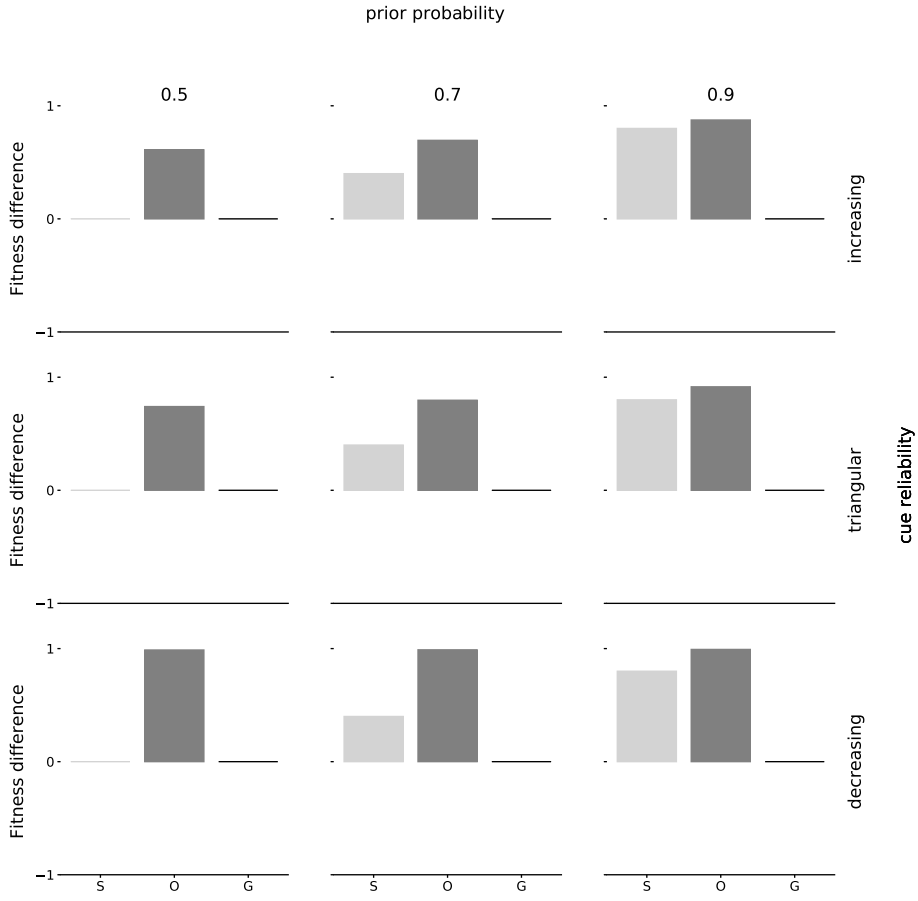


Figure A1.3 Fitness of mature phenotypes. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents a simulation study. For each study we simulate 10,000 organisms who follow the optimal policy and track their development across ontogeny. For half of the population the environmental state is fixed to E_0 and for the other half to E_1 (5,000 organisms per environment).We then compare the average fitness across organisms following the optimal policy ('O'; center, dark-grey bar) to two non-plastic strategies: generalists ('G'; right, black bar) and specialists ('S'; left, light-grey bar). Generalists always specialize halfway towards each phenotypic target, while specialists specialize towards the phenotypic target that is more likely according to the prior. If the prior is 0.5 organisms choose a target at random. Bars indicate the expected fitness difference from baseline (marked as 0) of the three strategies, normalized to range between -1 and 1 with 1 indicating a perfect match to the environment. Fitness differences can be negative when mismatch penalties exceed rewards for correct matches.

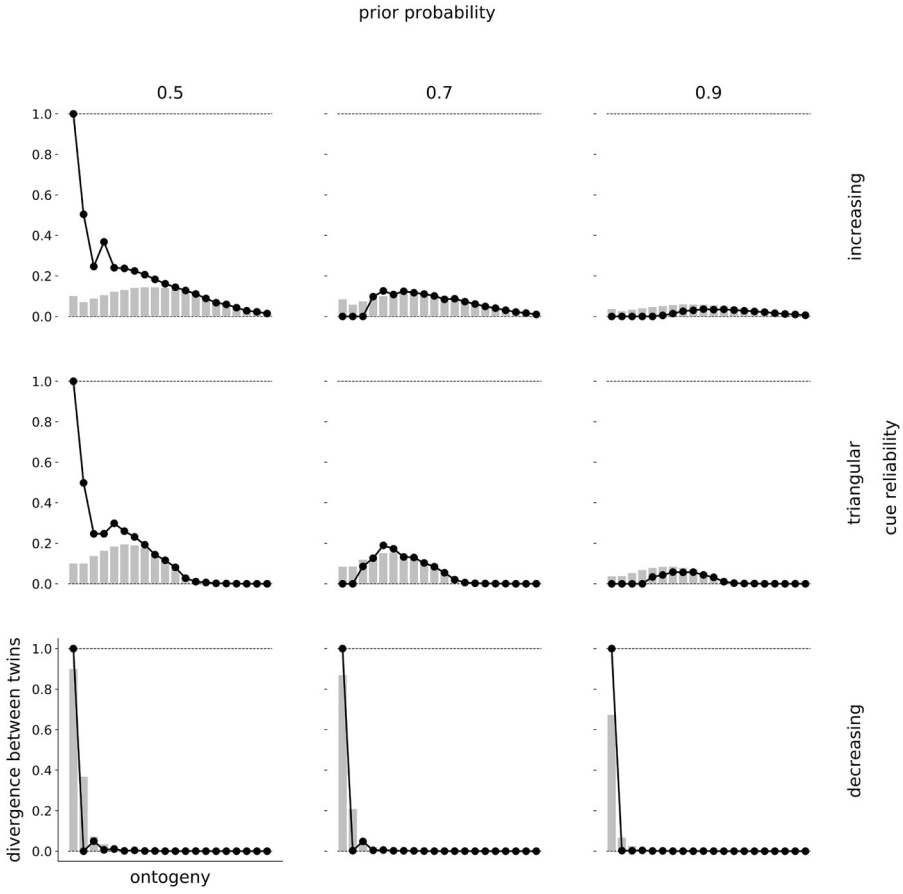


Figure A1.4 Plasticity in phenotype and posterior estimate. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Black circles correspond to the phenotype-and-posterior and gray bars to the posterior-only model. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it receives one reciprocal, opposite cue, and then continues normal development with its twin until the end of ontogeny. The vertical axis within each panel depicts the difference between focal individuals and their clones in the phenotype-and-posterior model and the posterior-only model. The horizontal axis depicts the time period in which pairs of twins were separated. Black lines and circles depict the average Euclidean distance between the 10,000 focal individuals and their clones after their separation (scaled to a 0 to 1 range), divided by the maximum possible distance attainable within one time period. Gray bars correspond the average absolute distance in posteriors between those same simulated organisms after their separation.

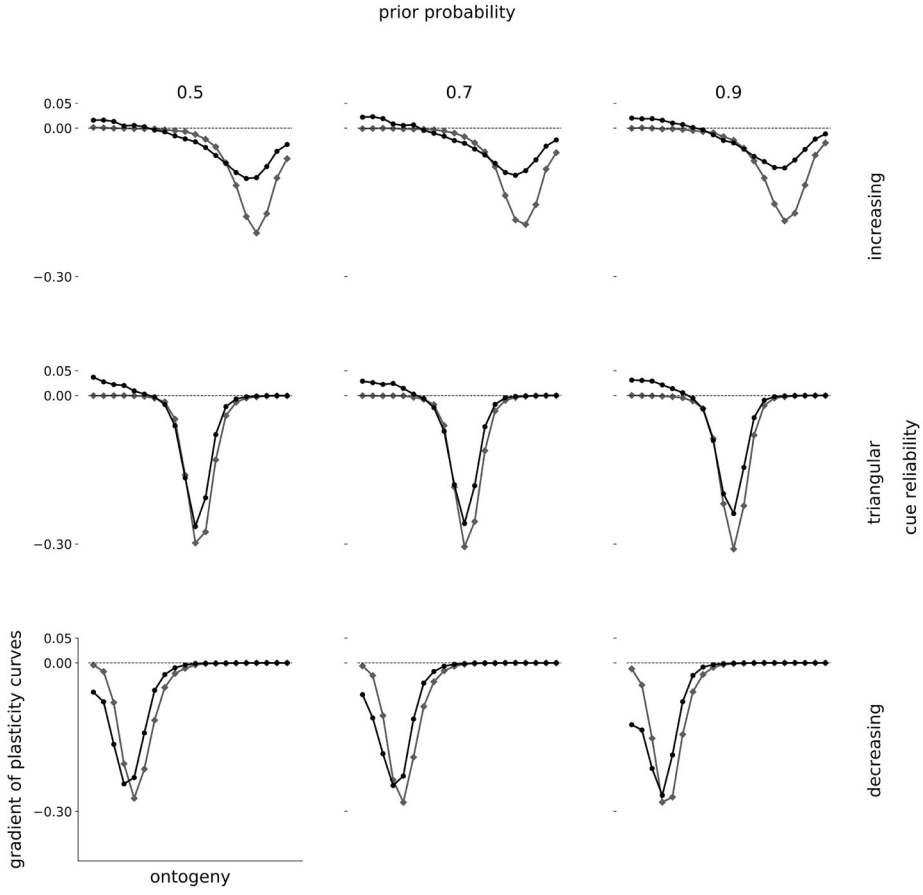


Figure A1.5 Gradients of phenotypic plasticity and plasticity in posterior estimates. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Black lines and circles correspond to the phenotype-and-posterior and gray lines and diamonds to the posterior-only model. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts until the end of ontogeny. Within each panel we compute the difference between focal individuals and their clones in the phenotype-and-posterior model and the posterior-only model. In the phenotype-and-posterior model this difference is the average Euclidean distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range), divided by the maximum possible distance following separation. In the posterior-only model this difference is the average absolute distance in posteriors between those same simulated organisms at the end of ontogeny. The horizontal axis depicts the time period in which pairs of twins were separated. The vertical axis within each panel depicts the gradients of the resulting plasticity trajectories from the phenotype-and-posterior and the posterior-only model.

Dynamic programming equations

a. Environmental variables and state of an organism

Environmental variable	Explanation
E_0	Environment 0
E_1	Environment 1
P_0	Optimal phenotype for E_0
P_1	Optimal phenotype for E_1
C_0	Cue indicating E_0
C_1	Cue indicating E_1
T	End of ontogeny, i.e., 20 for the results in the main text

The state of an organism is characterized by a 5-tuple (D_t, y_0, y_1, y_w, t). In each time step (from 1 until T) organisms first sample a cue and then make a phenotypic decision. D_t denotes the sequence of cues that an organism has sampled by time period t.

Variable	Explanation
D_t	$D_t = \{x_1, x_2, \ldots x_t\}$, where x_1, x_2 up until x_t denote the kind of cue (C_0 or C_1) received in each time period
y_0	Number of specialization steps towards P_0
y_1	Number of specialization steps towards P_1
y_w	Number of time steps spent waiting
t	Current time period in ontogeny

b. Bayesian inference

Organisms use Bayesian inference to update their initial prior estimate of the environmental state based on the sampled cues.

Parameters for Bayesian inference	Explanation
$P(E_0)$	Prior probability of E_0
$P(E_1)$	Prior probability of E_1
$P(C_{0,t}\|E_0)$	Cue reliability; conditional probability of receiving C_0in E_0 at t
$P(C_{1,t}\|E_1)$	Cue reliability; conditional probability of receiving C_1 in E_1 at t
$P(E_0\|D_t)$	Posterior probability of E_0 after having sampled D_t
$P(E_1\|D_t)$	Posterior probability of E_1 after having sampled D_t

According to the laws of probability it holds that:

$P(E_0) + P(E_1) = 1$

$P(E_0|D_t) + P(E_1|D_t) = 1$

$P(C_{0,t}|E_0) + P(C_{1,t}|E_0) = 1$

$P(C_{1,t}|E_1) + P(C_{0,t}|E_1) = 1$

Further, we assume that $P(C_{0,t}|E_0) = P(C_{1,t}|E_1)$.

We assume that organisms are Bayesian learners, using the fixed distribution of patches as the prior estimate of the environmental state and the time-dependent cue reliabilities to update these estimates. To see how this works, suppose an organism has sampled a specific sequence of cues $D_{t=3} = \{x_1 = C_0, x_2 = C_1, \ldots x_3 = C_0\}$.

According to Bayes' theorem, its posterior estimate after the first cue is:

$$P(E_0|C_0) = \frac{P(C_0|E_0) \bullet P(E_0)}{P(C_0|E_0) \bullet P(E_0) + P(C_0|E_1) \bullet P(E_1)}$$
$$P(E_1|C_0) = 1 - P(E_0|C_0)$$

To compute the posteriors $P(E_0|D_t)$ and $P(E_1|D_t)$ after the whole sequence of cues, we have to reapply Bayes' theorem for each cue using the previous posterior as the new prior.

c. Fitness functions

We denote the mature phenotype at the end of ontogeny by $Y_{mat} = (y_0, y_1, T)$.

Functions and constants	Explanation
$\phi(Y_{mat})$	Expected, additive fitness reward at the end of ontogeny
$\psi(Y_{mat})$	Expected, additive fitness penalty at the end of ontogeny
$\pi(Y_{mat})$	Expected fitness at the end of ontogeny
π_0	Baseline fitness
$f(y)$	Mapping between phenotypic increments and fitness rewards (or penalties)

Fitness consequences of phenotypic decisions are not accrued throughout ontogeny but only at the end of ontogeny. The fitness difference from baseline at the end of ontogeny corresponds to the total rewards for correct specializations minus penalties from incorrect specializations, where each correct increment results in a marginal gain and each incorrect increment results in a marginal penalty. We studied three mappings between correct (or incorrect) phenotypic development and fitness rewards (or penalties).

Suppose a mature organism is in the following state at the end of ontogeny (D_t, y_0, y_1, y_w, T) having sampled the cue sequence D_t and developed the mature phenotype $Y_{mat} = \{y_0, y_1, T\}$. Developing organisms aim to maximize expected fitness at the end of ontogeny. Expected fitness $\pi(Y_{mat})$ corresponds to the sum of expected rewards and penalties, in addition to the baseline fitness:

$$\pi(Y_{mat}) = \pi_0 + \phi(Y_{mat}) + \psi(Y_{mat}).$$

Suppose that an organism has sampled a specific sequence of cues, D_t, throughout ontogeny. Its posterior estimates $P(E_0|D_{t=T})$ and $P(E_1|D_{t=T})$ reflect the probabilities of being in either environmental state at the end of ontogeny. Thus, to compute rewards and penalties, we need to compute the expectation across both environmental states, weighted by how likely each state is as indicated by the posterior estimates at the end of ontogeny. We denote the mapping from phenotypic increments to rewards and penalties by $f(y)$, where y can refer to both y_0 and y_1, and derive the following expressions for expected rewards and penalties:

$$\phi(Y_{mat}) = P(E_0|D_{t=T}) \bullet f(y_0) + P(E_1|D_{t=T}) \bullet f(y_1)$$
$$\psi(Y_{mat}) = -\big(P(E_0|D_{t=T}) \bullet f(y_1) + P(E_1|D_{t=T}) \bullet f(y_0)\big)$$

Lastly, we present the three functional mappings between the realized phenotype and fitness rewards and penalties:

Returns on fitness - $f(y)$	Formula	Parameter settings to ensure that maximal rewards and penalties correspond to T - 1
linear	$f(y) = y$	-
diminishing	$f(y) = \alpha(1 - e^{-\beta y})$	$\beta = 0.2, \alpha = \frac{T-1}{1-e^{-\beta(T-1)}}$
increasing	$f(y) = \alpha(e^{\beta y} - 1)$	$\beta = 0.2, \alpha = \frac{T-1}{e^{\beta(T-1)}-1}$

d. Optimal decisions

In each time period, a developing organism can choose one of three options: increment one step on P_0, one step on P_1 or wait and forgo specialization. It chooses the option with the highest expected fitness at the end of ontogeny at $t = T$. In the event of a tie between two or all of the options the organism chooses amongst the current alternatives with equal probability.

$F(D_t, y_0, y_1, y_w, t, T)$ denotes the maximum expected fitness that can be attained as a result of decisions made between t and T, when the organism's current state after the last cue sampled is (D_t, y_0, y_1, y_w, T) and the organisms chooses option α, so that:

$$F(D_t, y_0, y_1, y_w, t, T) = \max_{\alpha \epsilon \{0,1,w\}} F_a \text{, where}$$

$$F_0 = F(D_t, y_0 + 1, y_1, y_w, t + 1, T),$$
$$F_1 = F(D_t, y_0, y_1 + 1, y_w, t + 1, T),$$
$$F_w = F(D_t, y_0, y_1, y_w + 1, t + 1, T).$$

We apply backwards induction to solve the dynamic programming equation $F(D_t, y_0, y_1, y_w, t, T)$ for all t. We start with $t = T$: $F(D_t, y_0, y_1, y_w, T, T) = \pi(y_0, y_1, T)$.

After calculating expected fitness at the end of ontogeny we continue by decrementing t. For each $t < T$ we compute the α, which maximizes $F(D_t, y_0, y_1, y_w, t + 1, T)$ in time period t.

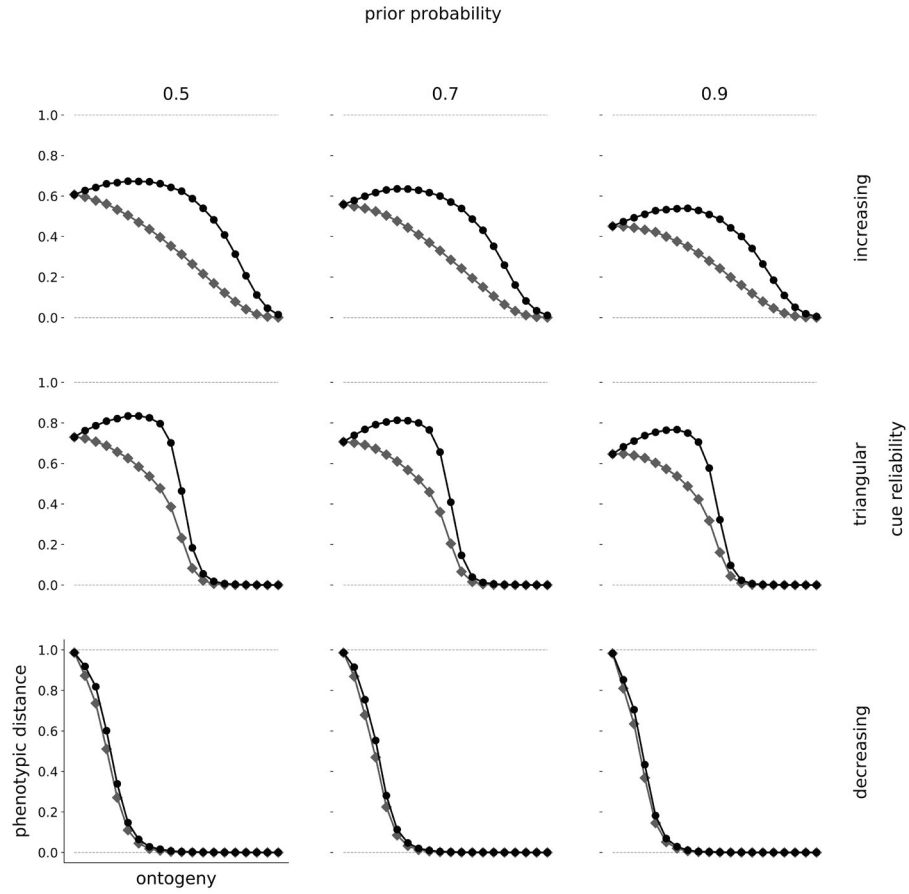


Figure A1.6 Plasticity across ontogeny – opposite patch cues. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive cues indicating the opposite patch, which lasts until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance at the end of ontogeny between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance following separation.

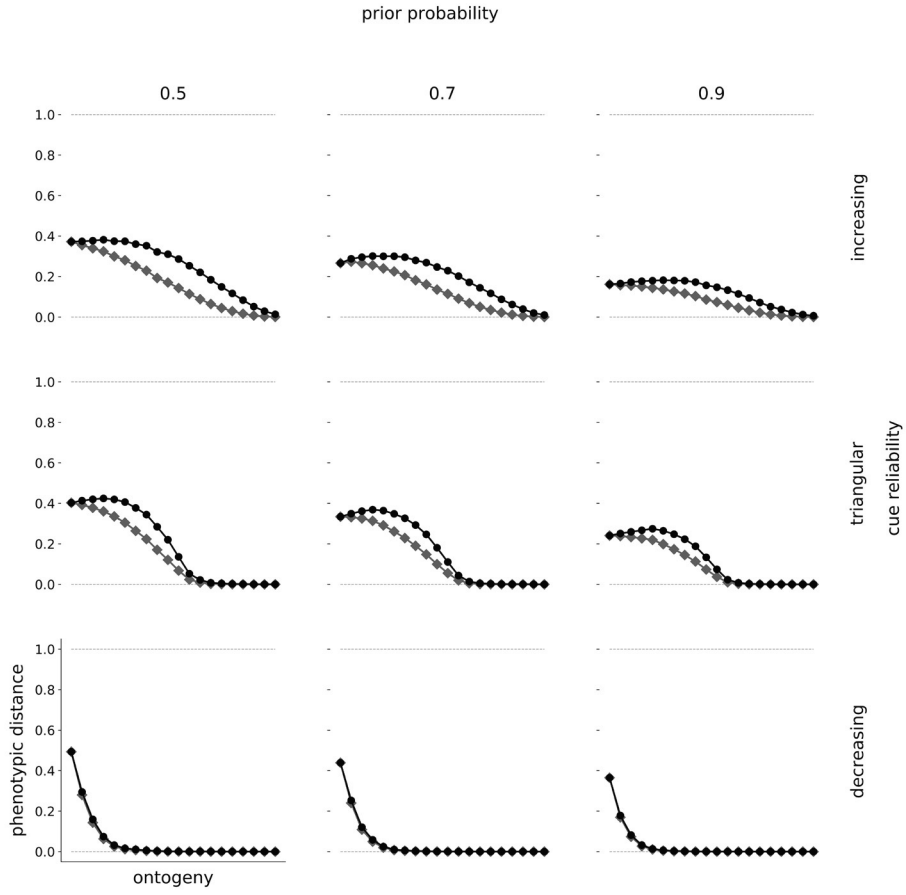


Figure A1.7 Plasticity across ontogeny - deprivation. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to experience deprivation, meaning that cues from both environmental states are equally likely, which lasts until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance at the end of ontogeny between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance following separation.

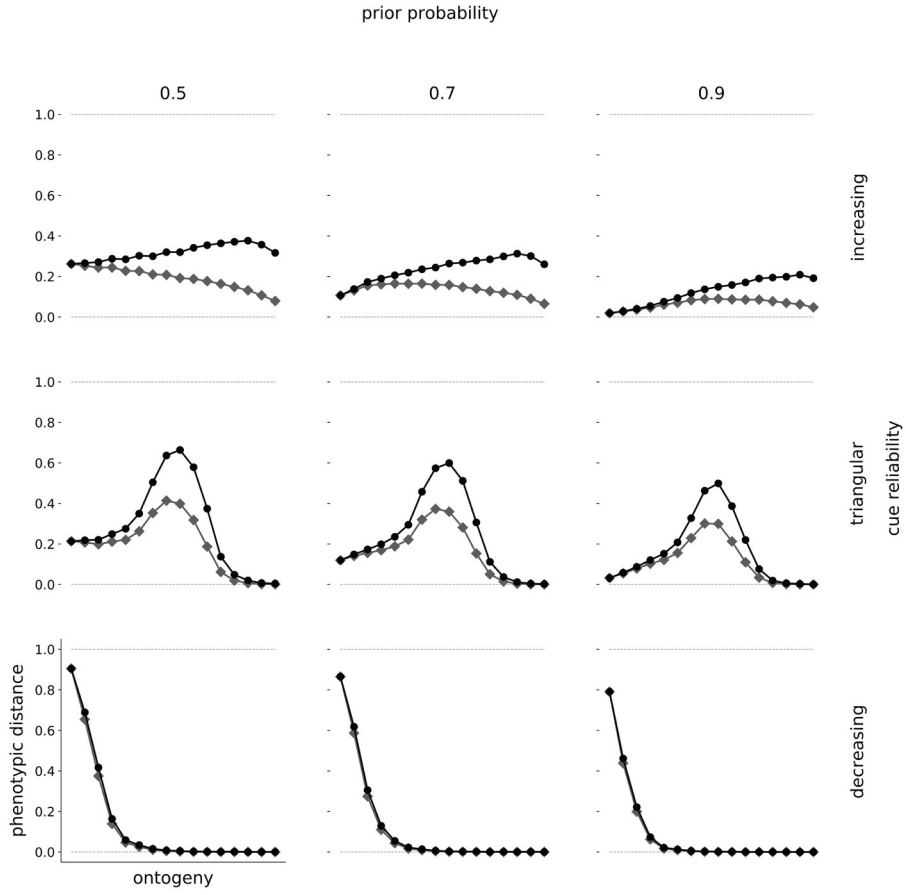


Figure A1.8 Plasticity across ontogeny – temporary separation and time of measurement is at the end of ontogeny. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts temporarily (for 5 discrete time points) before twins continue development together until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance at the end of ontogeny between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones at the end of ontogeny (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance following separation.

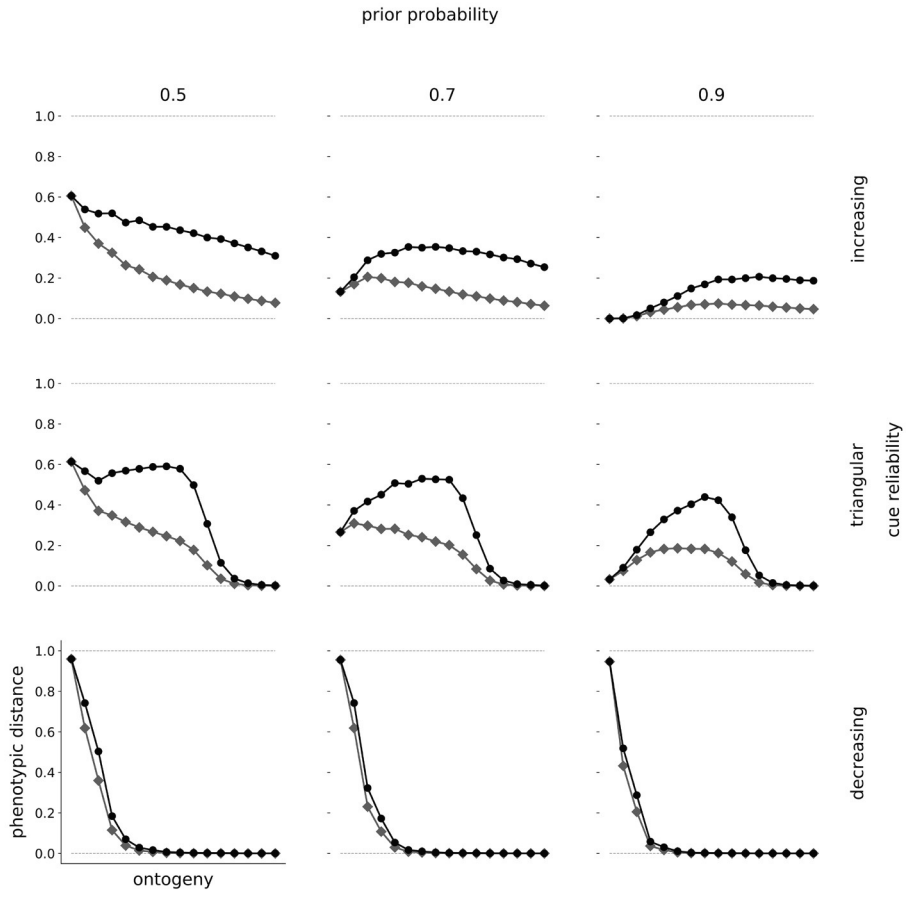


Figure A1.9 Plasticity across ontogeny – temporary separation and time of measurement is at the end of the separation. The fitness rewards for correct specializations and fitness penalties for incorrect specializations are linear across all panels. The prior probability of E_1 varies across columns and the cue reliability pattern varies across rows. Each panel represents T experimental 'twin studies', one for each $t \in \{1, T\}$. Outcomes of each twin study are marked by a grey diamond and a black circle. For each study we simulate 10,000 pairs of identical twins who follow the optimal policy and track their development across ontogeny. The environmental state is fixed to E_1. For each pair of twins, one individual (the 'focal') receives a set of environmental cues across ontogeny simulated from the prior probability and cue reliability pattern. Its clone receives the same cues until the moment of separation in time period t after which it begins to receive reciprocal, opposite cues, which lasts temporarily (for 5 discrete time points) before twins continue development together until the end of ontogeny. The vertical axis within each panel depicts the phenotypic distance between focal individuals and their clones. The horizontal axis depicts the time period in which pairs of twins were separated. The phenotypic distance after the separation between a focal individual and its clone corresponds to the Euclidean distance between their phenotypes. Grey lines and diamonds depict 'absolute' phenotypic distance, the average distance between the 10,000 focal individuals and their clones after the separation (ranging from 0 to 20√2, scaled to a 0 to 1 range). Black lines and circles depict 'proportional' distance, the average absolute distance divided by the maximum possible distance during the separation.

Appendix 2 – Chapter 4

Results for 15 time periods of ontogeny

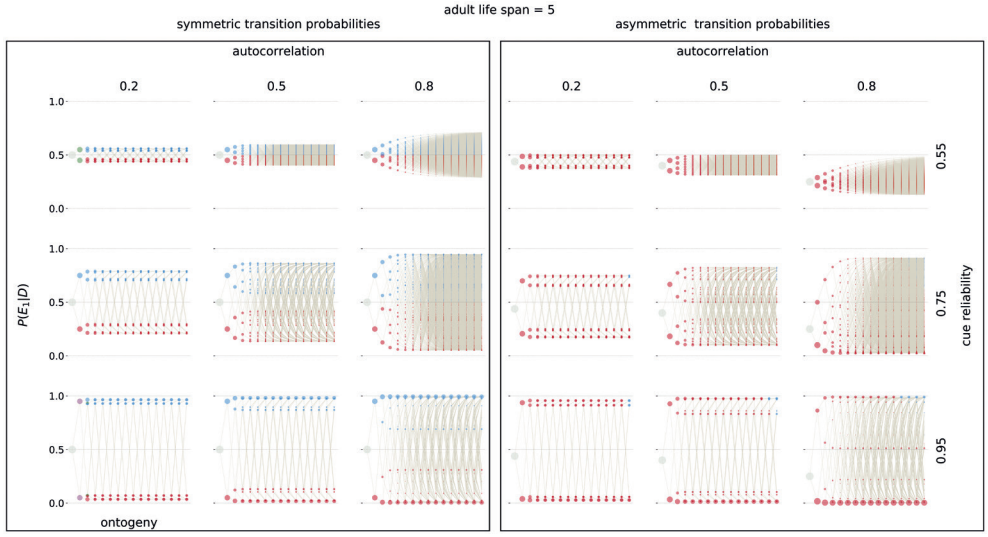


Figure A2.1 Optimal policies. Optimal policies are shown for linear rewards and linear penalties, T_{adult} = 5, and symmetric (left panel) and asymmetric (right panel) transition probabilities. Within each panel, columns indicate different autocorrelation levels and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays an organism's posterior estimate of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P_1, red to specializing towards P_0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. Per time step these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

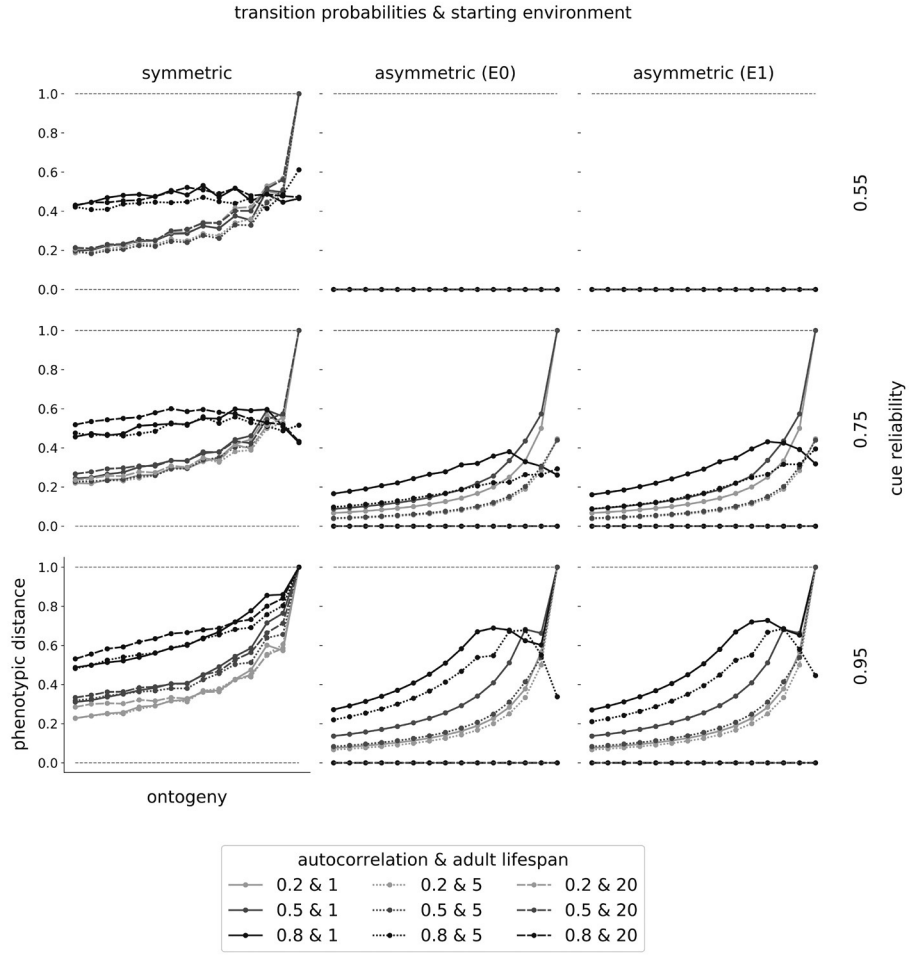


Figure A2.2 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. Rows indicate different cue reliabilities. Within each panel we show separate lines for different autocorrelation levels (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its counterpart. We assign weights to each sequence of cues depending on how likely it is for the respective Markov process. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of adoption. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

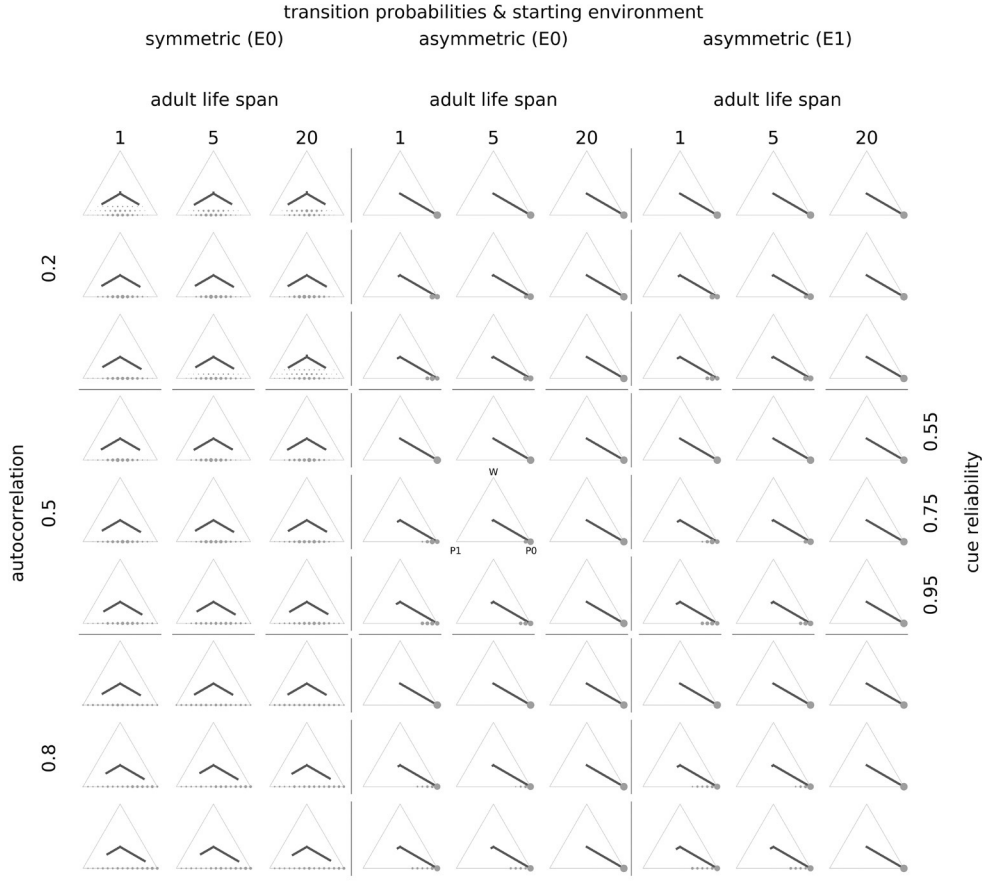


Figure A2.3 Distributions of mature phenotypes. Distributions of mature phenotypes are shown for linear rewards and linear penalties. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. When transition probabilities are symmetric, we only show distributions for E_0, since results for E_1 are mirrored across the vertical axis in the triangular plot. Rows (on the left side of the plot) indicate different autocorrelations. Transition probabilities and autocorrelations result in a 3 x 3 grid with nine panels. Within each panel, columns indicate different adult lifespans and rows (right side of the plot) indicate different cue reliability levels. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we simulate organisms following the optimal policy until maturation. Each light grey circle within a triangle identifies a unique phenotype at the end of ontogeny. A phenotype is specified by three numbers: the number of time steps waited (top vertex), the number of time steps specialized towards P_0 (right vertex), and the number of time steps specialized towards P_1 (left vertex). The closer a circle is towards one of the vertices, the higher the number of specializations towards that respective phenotype dimension. The area of a circle is proportional to the number of organisms developing a specific phenotype. Additionally, we show the average phenotype within each triangle (dark grey three-pronged star). The length of each prong pointing towards a vertex, represents the average number of specialization steps towards that vertex among all mature phenotypes.

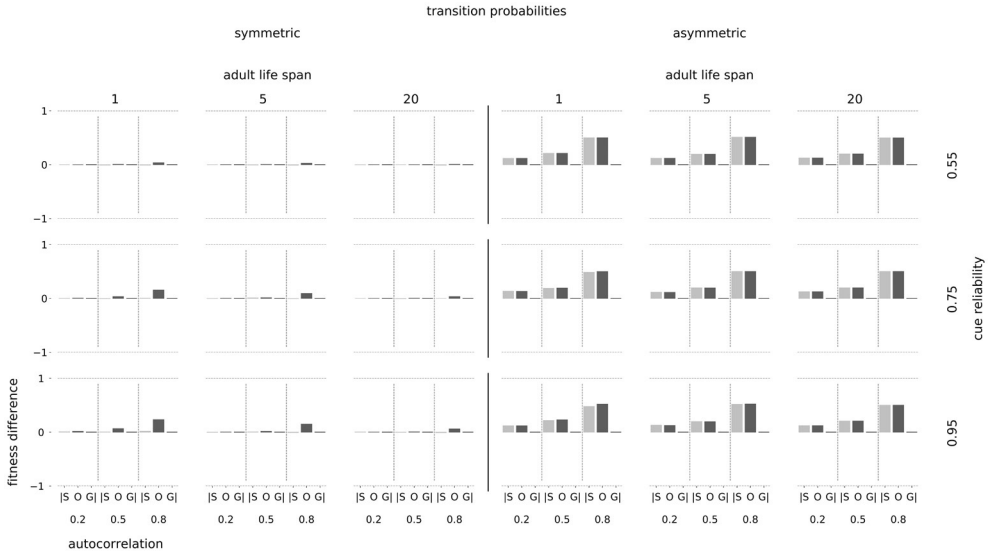


Figure A2.4 Fitness of the optimal policy and two fixed, non-plastic strategies. We compute fitness differences from baseline of specialized organisms. Fitness differences are shown for linear rewards and penalties. The left panel shows fitness differences for symmetric transition probabilities and the right panel for asymmetric ones. Within each panel columns indicate different adult lifespans and rows indicate different cue reliabilities. Each plot shows three different autocorrelation levels separated by dashed vertical lines. For each parameter combination, we compute fitness differences of three different strategies: The optimal policy (dark grey bars, indicated by "O"), a generalist strategy (black bars, indicated by "G"), and a specialist strategy (light grey bars, indicated by "S"). Generalists always specialize halfway towards P_0 and halfway towards P_1. Specialists fully specialize according to their prior (e.g., P_0 if $\pi(E_0) > \pi(E_1)$). If $\pi(E_1) = \nu(E_0) = 0.5$, one half of the population will fully specialize towards P_0 and the other half towards P_1. Total fitness corresponds to the average fit with the environment across adulthood. At each time period during adulthood fitness is the sum of rewards for correct specializations and penalties for incorrect ones.

Comparison 10 and 20 time periods of the adult lifespan

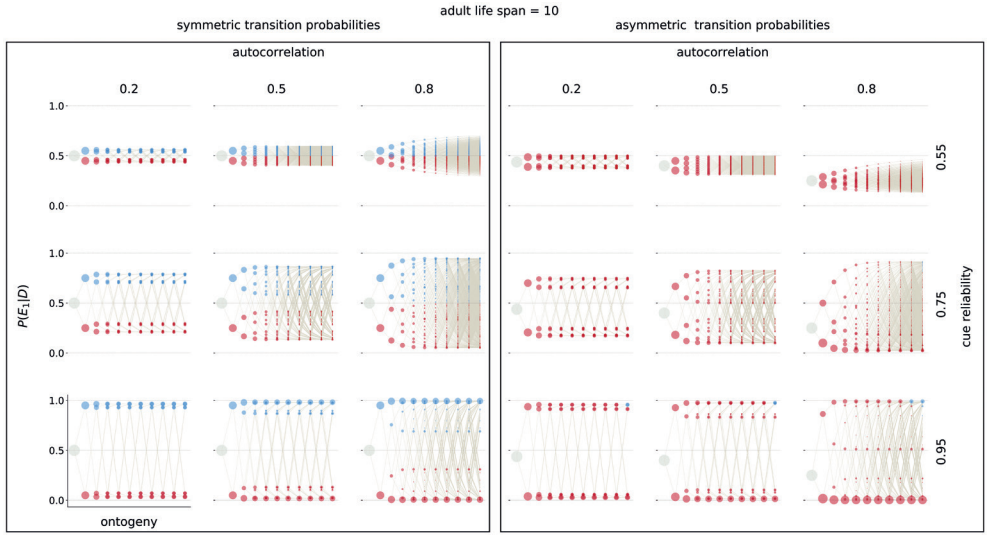


Figure A2.5 Optimal policies. Optimal policies are shown for linear rewards and linear penalties, T_{adult} = 10, and symmetric (left panel) and asymmetric (right panel) transition probabilities. Within each panel, columns indicate different autocorrelation levels and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays an organism's posterior estimate of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P_1, red to specializing towards P_0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. Per time step these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

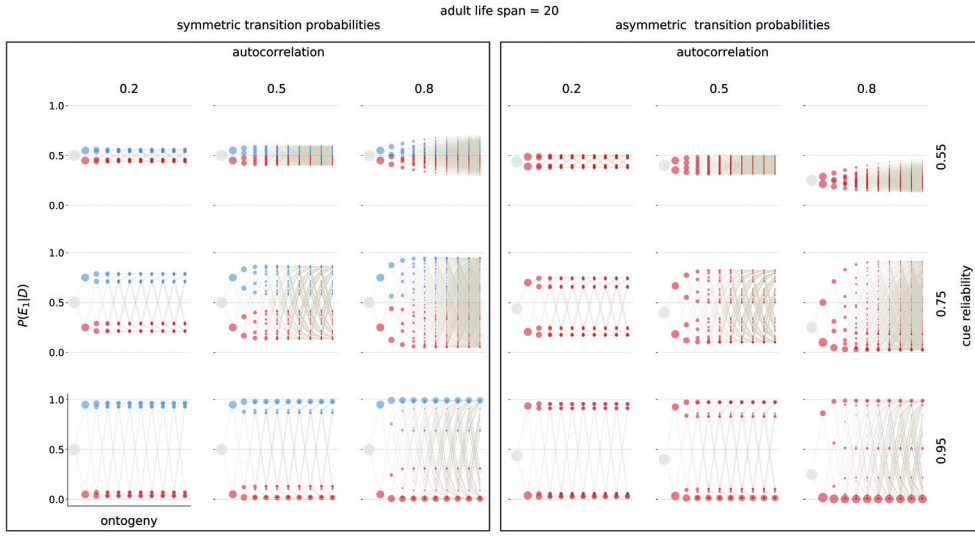


Figure A2.6 Optimal policies. Optimal policies are shown for linear rewards and linear penalties, T_{adult} = 20, and symmetric (left panel) and asymmetric (right panel) transition probabilities. Within each panel, columns indicate different autocorrelation levels and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays an organism's posterior estimate of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P_1, red to specializing towards P_0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. Per time step these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

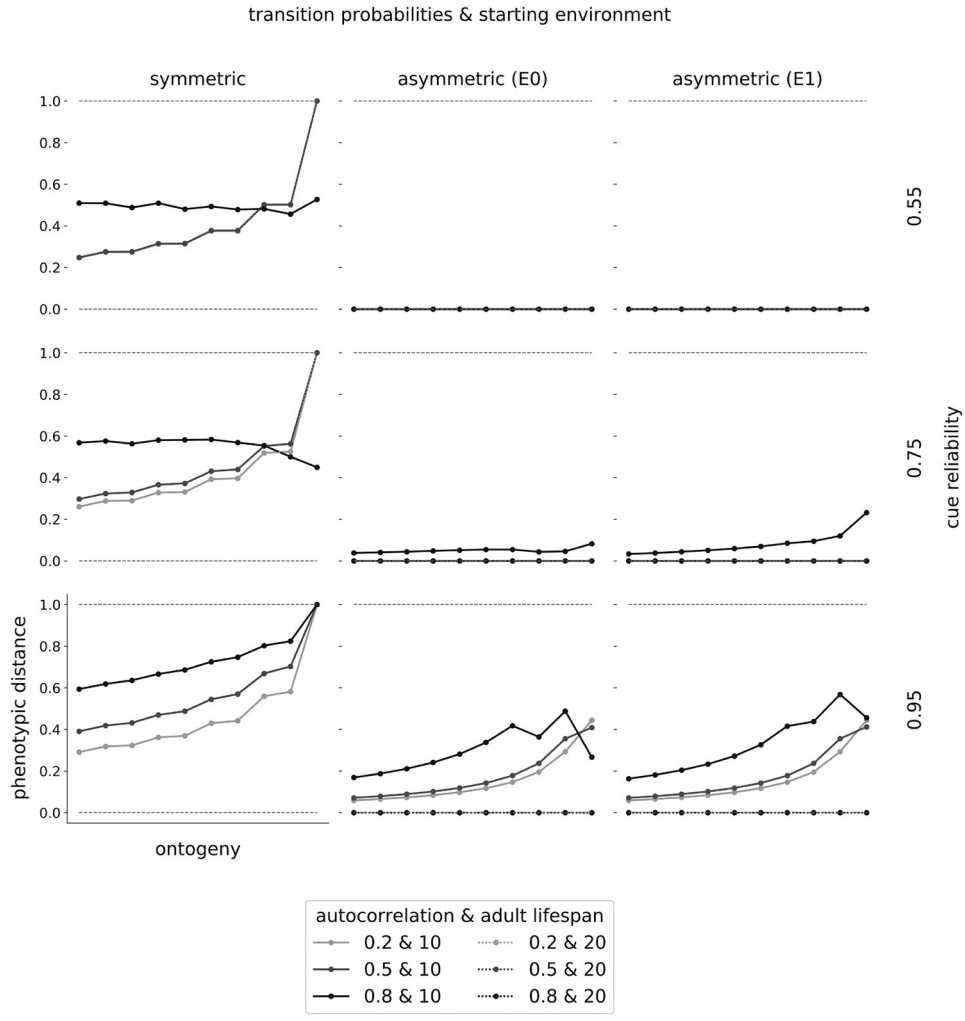


Figure A2.7 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. Rows indicate different cue reliabilities. Within each panel we show separate lines for different autocorrelation levels (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its counterpart. We assign weights to each sequence of cues depending on how likely it is for the respective Markov process. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of adoption. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

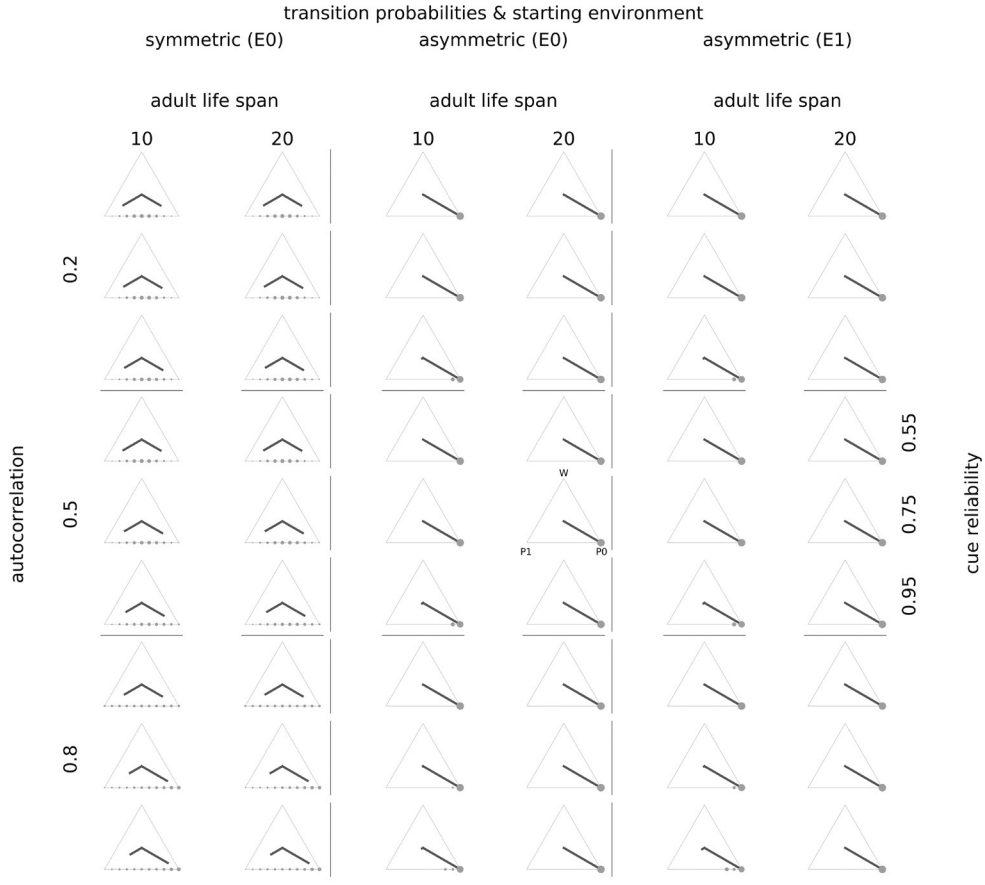


Figure A2.8 Distributions of mature phenotypes. Distributions of mature phenotypes are shown for linear rewards and linear penalties. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. When transition probabilities are symmetric, we only show distributions for E_0, since results for E_1 are mirrored across the vertical axis in the triangular plot. Rows (on the left side of the plot) indicate different autocorrelations. Transition probabilities and autocorrelations result in a 3 x 3 grid with nine panels. Within each panel, columns indicate different adult lifespans and rows (right side of the plot) indicate different cue reliability levels. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we simulate organisms following the optimal policy until maturation. Each light grey circle within a triangle identifies a unique phenotype at the end of ontogeny. A phenotype is specified by three numbers: the number of time steps waited (top vertex), the number of time steps specialized towards P_0 (right vertex), and the number of time steps specialized towards P_1 (left vertex). The closer a circle is towards one of the vertices, the higher the number of specializations towards that respective phenotype dimension. The area of a circle is proportional to the number of organisms developing a specific phenotype. Additionally, we show the average phenotype within each triangle (dark grey three-pronged star). The length of each prong pointing towards a vertex, represents the average number of specialization steps towards that vertex among all mature phenotypes.

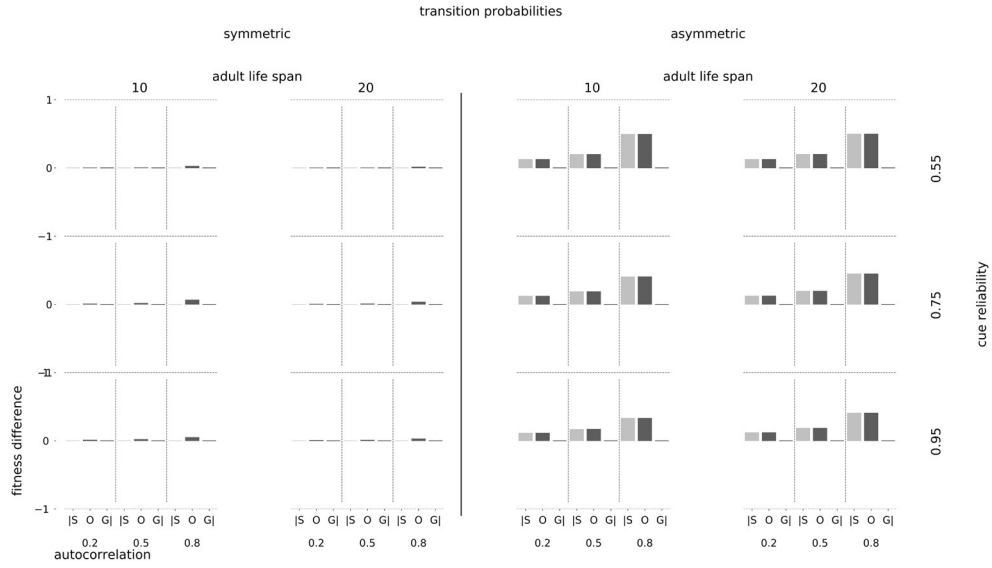


Figure A2.9 Fitness of the optimal policy and two fixed, non-plastic strategies. We compute fitness differences from baseline of specialized organisms. Fitness differences are shown for linear rewards and penalties. The left panel shows fitness differences for symmetric transition probabilities and the right panel for asymmetric ones. Within each panel columns indicate different adult lifespans and rows indicate different cue reliabilities. Each plot shows three different autocorrelation levels separated by dashed vertical lines. For each parameter combination, we compute fitness differences of three different strategies: The optimal policy (dark grey bars, indicated by "O"), a generalist strategy (black bars, indicated by "G"), and a specialist strategy (light grey bars, indicated by "S"). Generalists always specialize halfway towards P_0 and halfway towards P_1. Specialists fully specialize according to their prior (e.g., P_0 if $\pi(E_0) > \pi(E_1)$. If $\pi(E_1) = \nu(E_0) = 0.5$, one half of the population will fully specialize towards P_0 and the other half towards P_1. Total fitness corresponds to the average fit with the environment across adulthood. At each time period during adulthood fitness is the sum of rewards for correct specializations and penalties for incorrect ones.

Dynamic programming

a. Variables and explanation

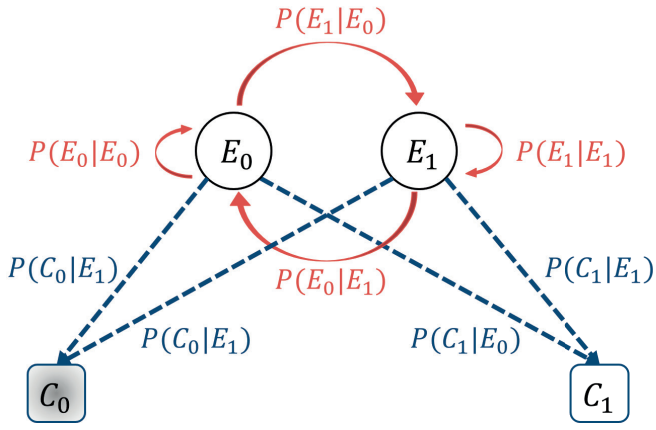
Environmental variable	Explanation
E_0	Environment 0
E_1	Environment 1
P_0	Optimal phenotype for E_0
P_1	Optimal phenotype for E_1
C_0	Cue indicating E_0
C_1	Cue indicating E_1
D_t	$D_t = \{x_1, x_2, \ldots x_t,\}$, denotes the sequence of cues until time period t where x_1, x_2, etc. until x_t denote the kind of cue (C_0 or C_1) sampled in each time period
t	Current time period ranges from $t = 0$ (birth) until T_{end} (the end of the reproductive cycle). It holds that $T_{end} = T_{ont} + T_{adult}$.
T_{ont}	Duration of ontogeny, i.e. ontogeny lasts for 10 time periods
T_{adult}	Duration of adulthood, i.e. adulthood lasts for 1, 5 or 10 time periods

In each time period (from 1 until T_{end}) the state of the environment may change according to the transition probabilities $P(E_0|E_1)$ and $P(E_1|E_0)$. Organisms first sample a cue and then make a phenotypic decision, i.e. increasing y_0, y_1 or y_w by 1 increment. $D_t = \{x_1, x_2, \ldots x_t,\}$ denotes the sequence of cues that an organism has sampled by time period t. The state of an organism is characterized by a 5-tuple (D_t, y_0, y_1, y_w, t).

Variable	Explanation
y_0	Number of specialization steps towards P_0
y_1	Number of specialization steps towards P_1
y_w	Number of time steps spent waiting

b. Hidden Markov Model and forward algorithm

We model ontogeny as a Hidden Markov Model in which the state of the environment is hidden or unobserved and cues correspond to the observed states. In this example the observed cue C_0 could have either been the outcome of the hidden state E_0 or E_1.



Organisms use Bayesian inference to update their initial prior estimate of the environmental state based on the sampled cues. The forward algorithm can be used to compute the posterior probability of each state given the history of sampled cues in a hidden Markov model.

Parameters for Bayesian inference	Explanation
$P(E_0)$	Prior probability of E_0
$P(E_1)$	Prior probability of E_1
$P(E_0\|E_1)$	Transition probability from E_1 to E_0
$P(E_1\|E_0)$	Transition probability from E_0 to E_1
$P(C_0\|E_0)$	Cue reliability; conditional probability of receiving C_0 in E_0
$P(C_1\|E_1)$	Cue reliability; conditional probability of receiving C_1 in E_1
$P(E_0\|D_t)$	Posterior probability of E_0 after having sampled D_t
$P(E_1\|D_t)$	Posterior probability of E_1 after having sampled D_t

According to the laws of probability it holds that:

$$P(E_0) + P(E_1) = 1$$

$$P(E_0|D_t) + P(E_1|D_t) = 1$$

$$P(E_0, D_t) + P(E_1, D_t) = P(D_t)$$

$$P(C_0|E_0) + P(C_1|E_0) = 1$$

$$P(C_1|E_1) + P(C_0|E_1) = 1$$

$$P(E_0|E_0) + P(E_1|E_0) = 1$$

$$P(E_1|E_1) + P(E_0|E_1) = 1$$

Further, we assume that $P(C_0|E_0) = P(C_1|E_1)$.

The forward algorithm allows us to compute the joint probabilities $P(E_0|D_t)$ and $P(E_1|D_t)$ for any sequence D_t.

For each possible sequence $D_t = \{x_1, x_2, \ldots x_t\}$ of cues C_0 and C_1, we apply the forward algorithm as defined by the following recursions running from 1 until the current time period t during ontogeny:

$$P(E_{0,t}, D_t) = P(x_t|E_{0,t}) \bullet \sum_{i \in \{0,1\}} P(E_{i,t-1}, D_{t-1}) \bullet P(E_{0,t}|E_{i,t-1})$$

$$P(E_{1,t}, D_t) = P(x_t|E_{1,t}) \bullet \sum_{i \in \{0,1\}} P(E_{i,t-1}, D_{t-1}) \bullet P(E_{1,t}|E_{i,t-1})$$

$P(x_t|E_{0,t})$ and $P(x_t|E_{1,t})$ are specified by the cue reliabilities $P(C_0|E_0)$ and $P(C_1|E_1)$, depending on whether the t^{th} cue in the sequence corresponded to C_0 or C_1. $P(E_{0,t}|E_{i,t-1})$ and $P(E_{1,t}|E_{i,t-1})$ are specified by the transition probabilities between states, where $E_{i,t-1}$ can correspond to either E_0 or E_1. The outcome for the recursion will be the joint probabilities of any of the two environmental states and the specific sequence of cues, i.e., $P(E_0|D_t)$ and $P(E_0|D_t)$.

Lastly, we compute the posterior probabilities $P(E_0|D_t)$ and $P(E_1|D_t)$ according to:

$$P(E_0|D_t) = \frac{P(E_0, D_t)}{P(E_0, D_t) + P(E_1, D_t)}$$

$$P(E_1|D_t) = \frac{P(E_1, D_t)}{P(E_0, D_t) + P(E_1, D_t)}$$

c. Fitness functions

We denote the mature phenotype at the end of ontogeny by $Y_{mat} = (y_0, y_1, T_{ont})$.

Functions and constants	Explanation
$\phi(Y_{mat}, t)$	Expected, fitness reward at time period t during adulthood ($t > T_{ont}$)
$\psi(Y_{mat}, t)$	Expected, fitness penalty at time period t during adulthood ($t > T_{ont}$)
$\pi(Y_{mat}, t)$	Expected fitness at time period t during adulthood ($t > T_{ont}$)
$\pi_{Total}(Y_{mat})$	Expected fitness across adulthood
π_0	Baseline fitness
$f(y)$	Mapping between phenotypic increments and fitness rewards (or penalties)
μ	Penalty weight

Fitness consequences of phenotypic decisions are not accrued throughout ontogeny but only during adulthood. At each time period during adulthood the fitness difference from baseline corresponds to the total rewards for correct specializations minus penalties from incorrect specializations, where each correct increment results in a marginal gain and each incorrect increment results in a marginal penalty. We studied three mappings between correct (or incorrect) phenotypic development and fitness rewards (or penalties).

Suppose a mature organism is in the following state at the end of ontogeny ($D_{Tont}, y_0, y_1, y_w, T_{ont}$) having sampled the sequence of cues D_{Tont} and developed the mature phenotype $Y_{mat} = \{y_0, y_1, T_{ont}\}$. Its posterior estimates $P(E_{0,Tont}|D_{Tont})$ and $P(E_{1,Tont}|D_{Tont})$ at the end of ontogeny reflect the probabilities of being in either environmental state at the onset of adulthood. Using these two probabilities as the starting distribution and the transition matrix P we are able to compute the probabilities $P(E_{0,Tont}|D_{Tont})$ and $P(E_{1,Tont}|D_{Tont})$ for each time period t in adulthood:

$$\left(P\left(E_{0,t}\middle|D_{T_{ont}}\right), P\left(E_{1,t}\middle|D_{T_{ont}}\right)\right) = \left(P\left(E_{0,T_{ont}}\middle|D_{T_{ont}}\right), P\left(E_{1,T_{ont}}\middle|D_{T_{ont}}\right)\right) \bullet P^{T_{End}-T_{ont}},$$
$$where\ P = \begin{pmatrix} P(E_0|E_0) & P(E_1|E_0) \\ P(E_0|E_1) & P(E_1|E_1) \end{pmatrix}.$$

To compute rewards and penalties at each adult time period, we need to compute the expectation across both environmental states, weighted by how likely each state is as indicated by the posterior estimates. We denote the mapping from phenotypic increments to rewards and penalties by $f(y)$, where y can refer to both y_0 and y_1, and derive the following expressions for expected rewards and penalties at each adult time period:

$$\phi(Y_{mat}, t) = P\left(E_{0,t}\middle|D_{T_{ont}}\right) \bullet f(y_0) + P\left(E_{1,t}\middle|D_{T_{ont}}\right) \bullet f(y_1)$$
$$\psi(Y_{mat}, t) = -\left(P\left(E_{0,t}\middle|D_{T_{ont}}\right) \bullet f(y_1) + P\left(E_{1,t}\middle|D_{T_{ont}}\right) \bullet f(y_0)\right)$$

At any time period in adulthood expected fitness $\pi(Y_{mat}, t)$ corresponds to the sum of expected rewards and penalties, in addition to the baseline fitness:

$$\pi(Y_{mat}, t) = \pi_0 + \phi(Y_{mat}, t) + \mu \bullet \psi(Y_{mat}, t).$$

Total fitness across adulthood then corresponds to the average fitness across adulthood:

$$\pi_{Total}(Y_{mat}) = \frac{1}{T_{adult}} \bullet \sum_t \pi(Y_{mat}, t).$$

Lastly, we present the three functional mappings between the realized phenotype and fitness rewards and penalties:

Returns on fitness - $f(y)$	**Formula**	**Parameter settings to ensure that maximal rewards and penalties correspond to** T_{ont}
linear	$f(y) = y$	-
diminishing	$f(y) = \alpha(1 - e^{-\beta y})$	$\beta = 0.2, \alpha = \frac{T_{ont}}{1 - e^{-\beta(T_{ont})}}$
increasing	$f(y) = \alpha(e^{\beta y} - 1)$	$\beta = 0.2, \alpha = \frac{T_{ont}}{e^{\beta(T_{ont})} - 1}$

d. Optimal decisions

In each time period, a developing organism can choose one of three options: increment one step on P_0, one step on P_1 or wait and forgo specialization. It chooses the option with the highest expected fitness across adulthood. In the event of a tie between two or all of the options the organism chooses amongst the current alternatives with equal probability.

$F(D_t, y_0, y_1, y_w, t, T_{ont})$ denotes the maximum expected fitness that can be attained across adulthood as a result of decisions made between t and T_{ont}, when the organism's current state after the last cue sampled is $(D_{Tont}, y_0, y_1, y_w, T_{ont})$ and the organisms chooses option a, so that:

$$F(D_t, y_0, y_1, y_w, t, T_{ont}) = \max_{a\epsilon\{0,1,w\}} F_a \text{, where}$$

$$F_0 = F(D_{t+1}, y_0 + 1, y_1, y_w, t + 1, T_{ont}),$$
$$F_1 = F(D_{t+1}, y_0, y_1 + 1, y_w, t + 1, T_{ont}),$$
$$F_w = F(D_{t+1}, y_0, y_1, y_w + 1, t + 1, T_{ont}).$$

We apply backwards induction to solve the dynamic programming equation $F(D_t, y_0, y_1, y_w, t, T_{ont})$ for all t. We start with $t = T_{ont}$:

$$F\left(D_{T_{ont}}, y_0, y_1, y_w, T_{ont}, T_{ont}\right) = \pi_{Total}(Y_{mat}).$$

After calculating expected fitness at the end of ontogeny we continue by decrementing t. For each $t < T_{ont}$ we compute the a, which maximizes $F(D_{t+1}, y_0, y_1, y_w, t + 1, T_{ont})$ in time period t.

From transition probabilities to autocorrelations

For any transition probabilities the autocorrelation can be computed from the transition matrix $\begin{pmatrix} P(E_0|E_0) & P(E_1|E_0) \\ P(E_0|E_1) & P(E_1|E_1) \end{pmatrix}$, using the following formula:

$$r = 1 - (P(E_1|E_0) + P(E_0|E_1)) \quad (1)$$

The following figure displays how transition probabilities translate to autocorrelation values and stationary distributions.

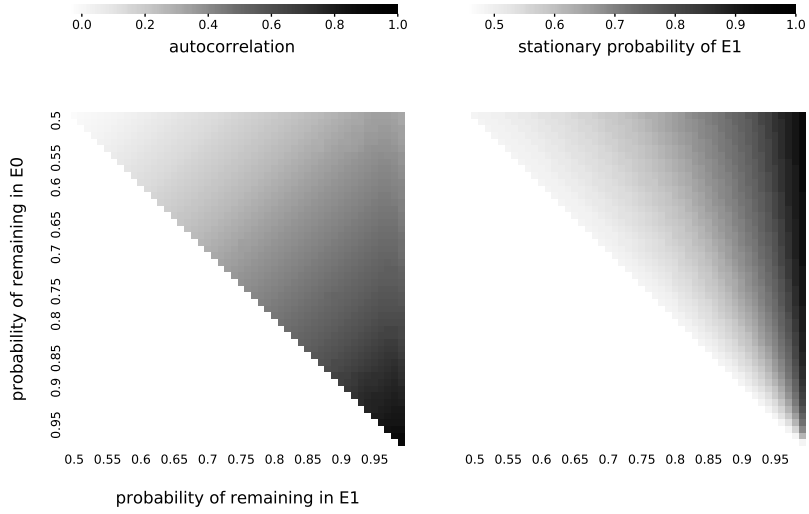


Figure A2.10 From transition probabilities to autocorrelations and stationary distributions. The left panel shows autocorrelations and the right panel shows the stationary probability of being in E_1. Within each panel the x-axis displays the probability $P(E_1|E_1)$ of remaining in E_1 and the y-axis displays the probability $P(E_0|E_0)$ of remaining in E_0. Autocorrelations and stationary probabilities have been computed for a continuous range of transition probabilities. Only a few discrete values are displayed on each axis for readability.

Comparison of different asymmetries

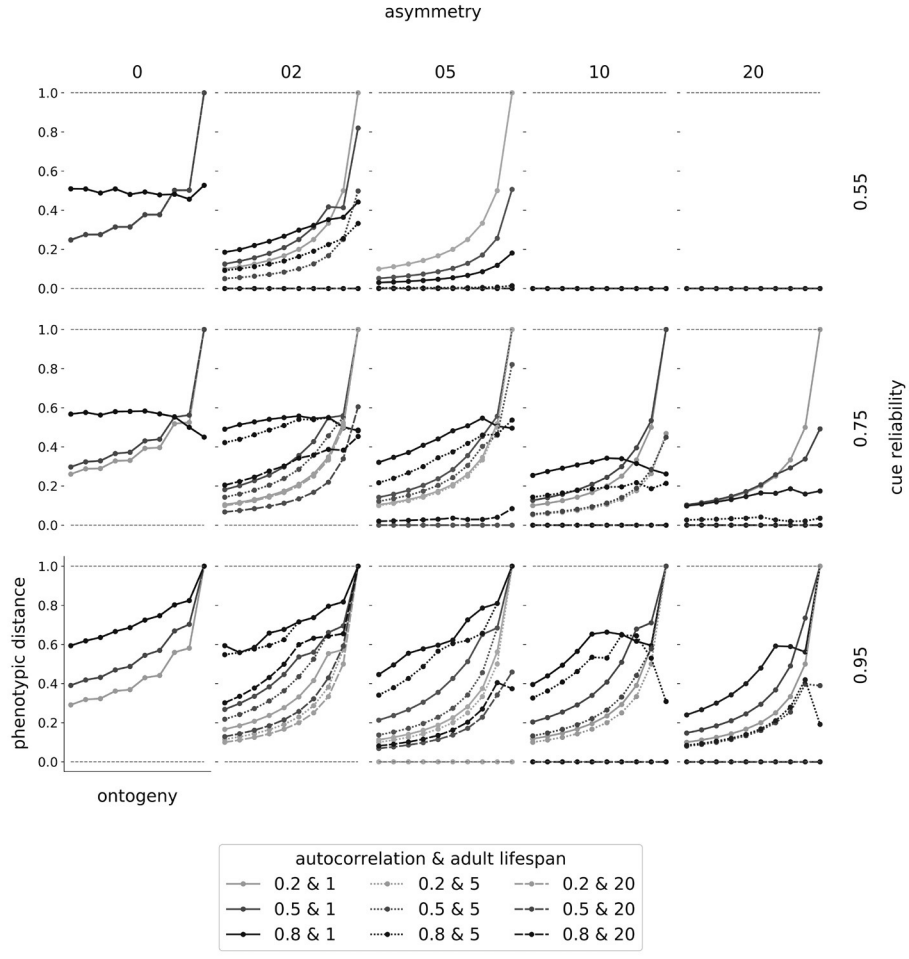


Figure A2.11 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties. Columns indicate the asymmetry level $P(E_0|E_1)$ - $P(E_1|E_0)$. An asymmetry of 0 implies symmetric transition probabilities. Pairs of clones started in the more likely environment, E_0. Rows indicate different cue reliabilities. Within each panel we show separate lines for different autocorrelation levels (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its counterpart. We assign weights to each sequence of cues depending on how likely it is for the respective Markov process. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of adoption. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

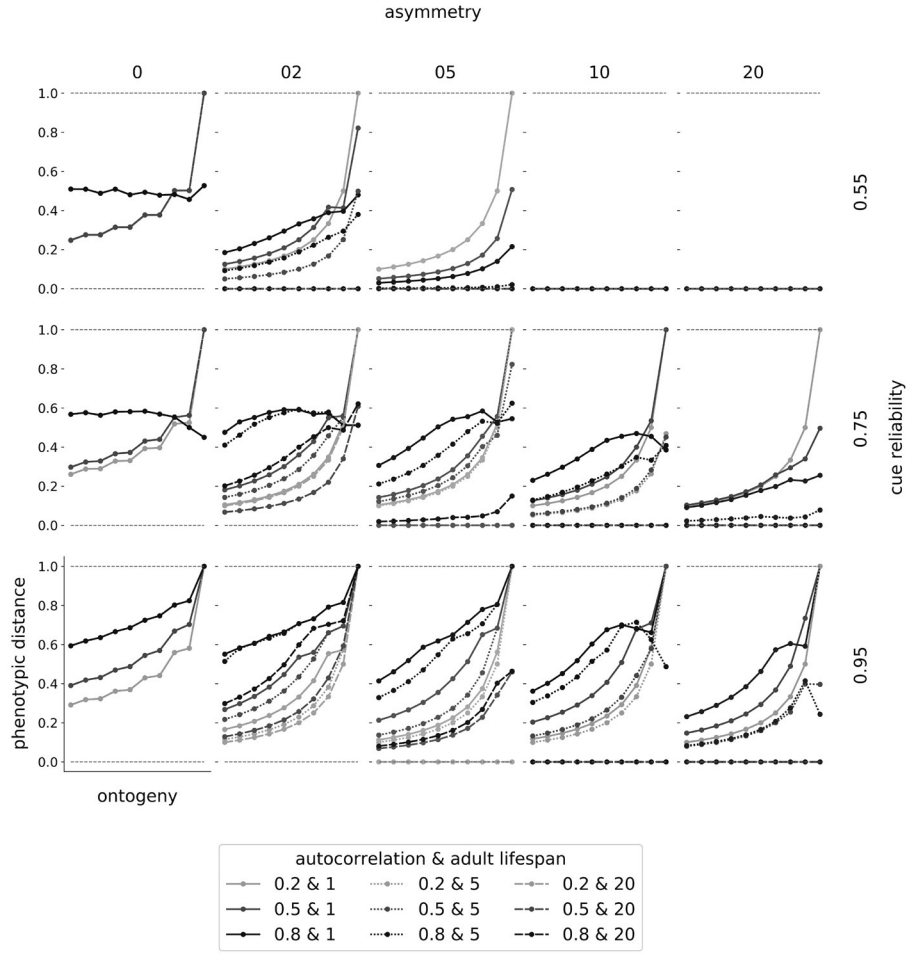


Figure A2.12 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties. Columns indicate the asymmetry level $P(E_0|E_1)$ - $P(E_1|E_0)$. An asymmetry of 0 implies symmetric transition probabilities. Pairs of clones started in the less likely environment, E_1. Rows indicate different cue reliabilities. Within each panel we show separate lines for different autocorrelation levels (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its counterpart. We assign weights to each sequence of cues depending on how likely it is for the respective Markov process. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of adoption. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

Adoption study paradigm and distributions of posterior estimates during ontogeny and adulthood

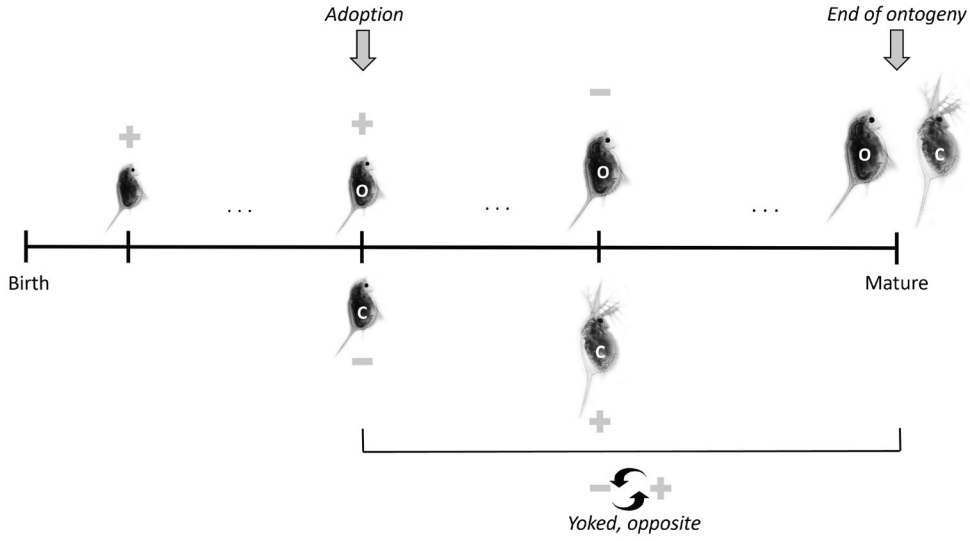


Figure A2.13 Adoption study paradigm. Clones develop together until adoption, experiencing the same sequence of cues and making the same phenotypic decisions, resulting in identical phenotypes. At this point, the clones are separated, with one (the focal, denoted by 'O') remaining in the original patch and the other (the copy, denoted by 'C') developing in a mirror patch. The sequence of environmental states in the mirror patch is the same as in the original patch. However, the cues in the mirror patch are opposite those in the original patch. Whenever the focal individual samples a cue to one state (e.g. '+'), the copy samples a cue to the other state (e.g. '-'), and vice versa. Focal-and-copy pairs continue development until maturation. At the end of ontogeny, we compute the average difference between the pairs of simulated twins. The larger the difference between mature phenotypes, the higher the levels of plasticity at the onset of adoption. Copyright: this figure has been adapted from Frankenhuis and Walasek (2020) and we have used the images of Daphnia with permission from Dr. Weiss (2019).

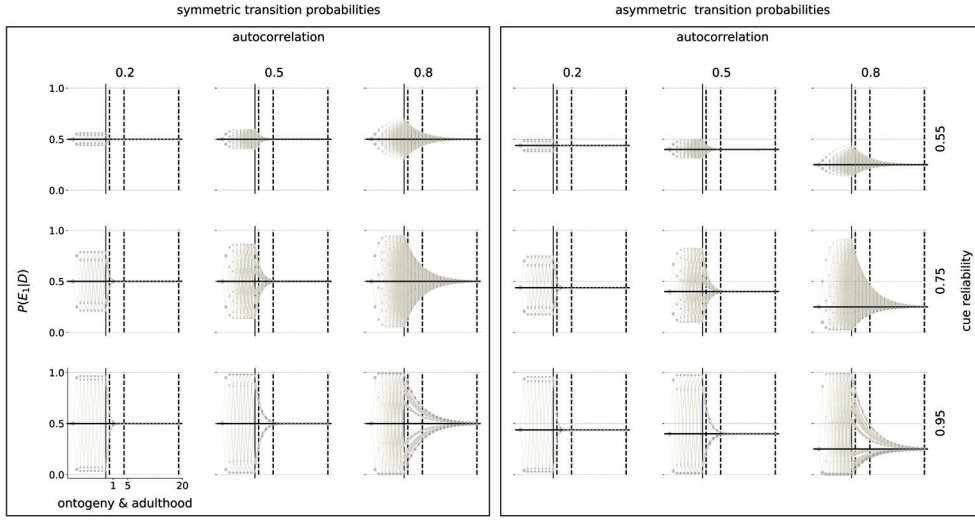


Figure A2.14 Distributions of posterior estimates during ontogeny and adulthood. Distributions are shown for symmetric (left panel) and asymmetric (right panel) transition probabilities. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny and adulthood. The solid, black horizontal line indicates the stationary, long-term probability of being in E_1, $\pi(E_1)$, for each Markov process. The vertical solid, black line marks the final time period in ontogeny. Finally, the dashed, black vertical lines mark possible end points of adulthood. Each filled grey circle indicates a possible posterior estimate throughout an organism's lifespan. The area of a circle is proportional to the probability of reaching that estimate. Grey lines between posteriors depict developmental trajectories.

Additional results for 10 time periods

Figure A2.15 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, symmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P1, red to specializing towards P0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

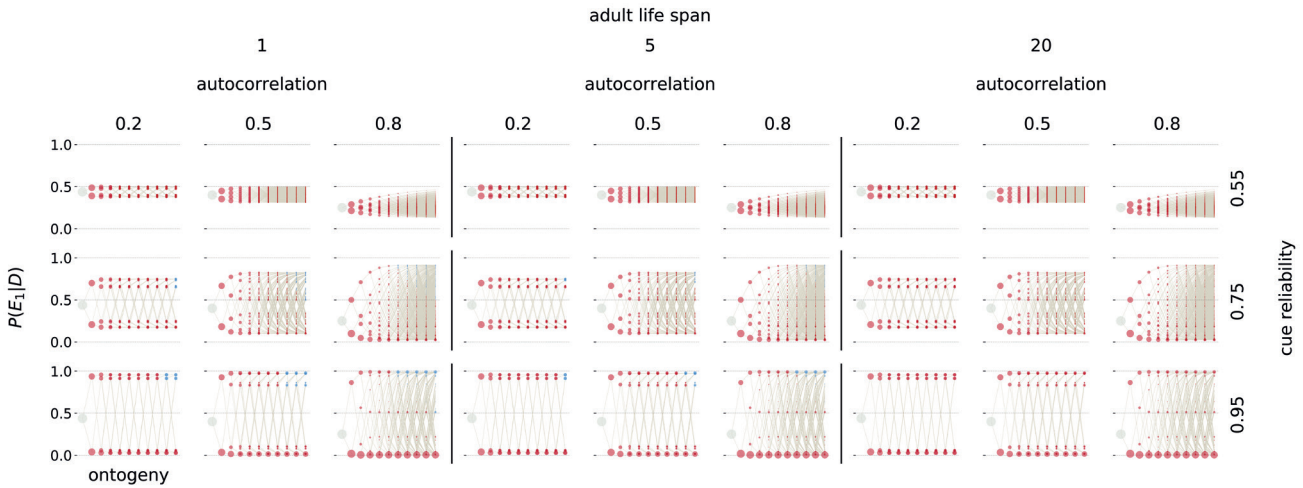


Figure A2.16 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, asymmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P1, red to specializing towards P0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

Results for different penalty weights

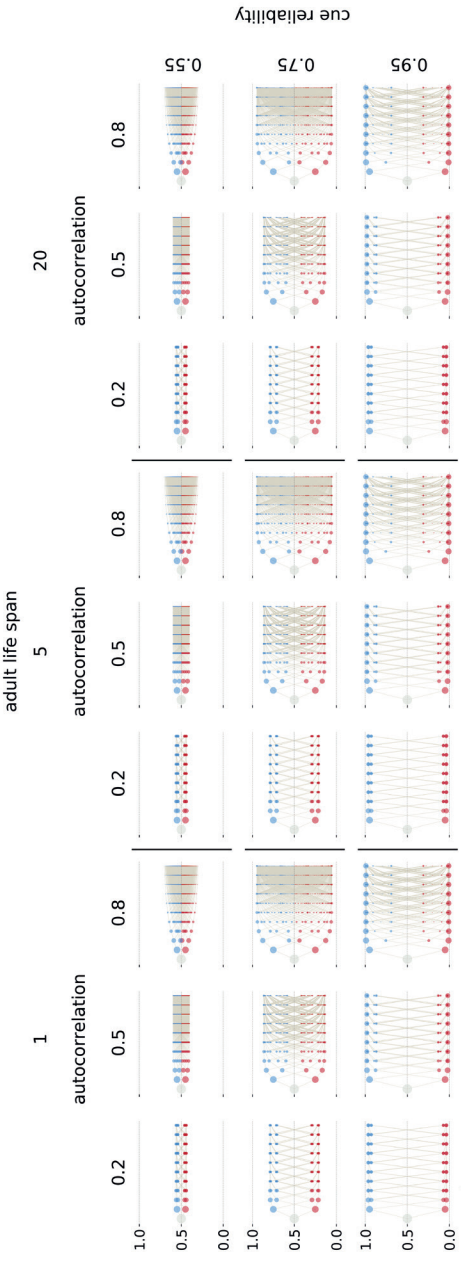


Figure A2.17 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, a penalty weight of 0.5, symmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P1, red to specializing towards P0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

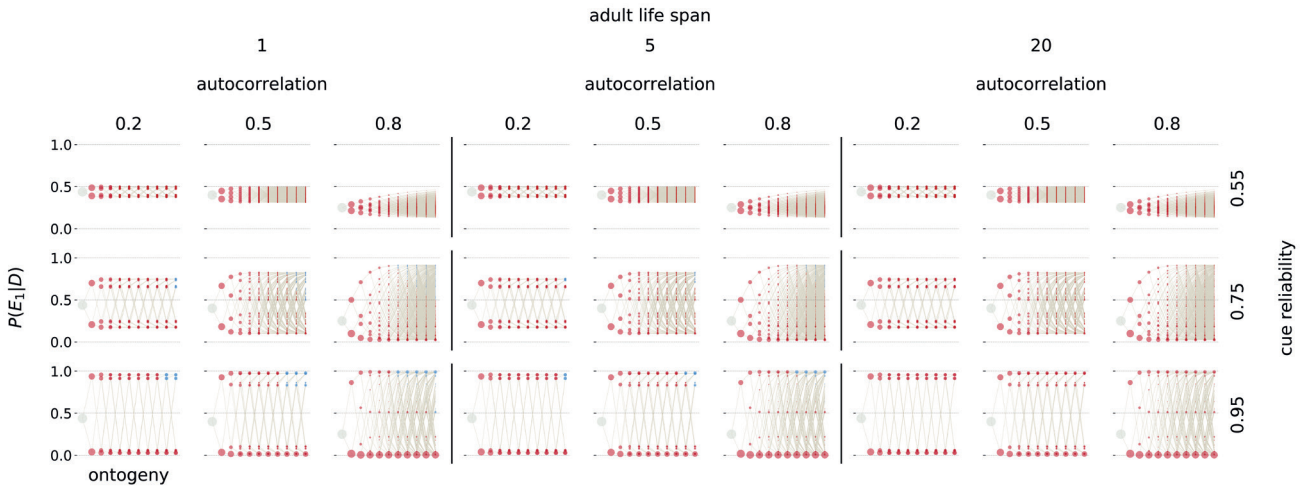


Figure A2.18 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, a penalty weight of 0.5, asymmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P_1, red to specializing towards P_0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

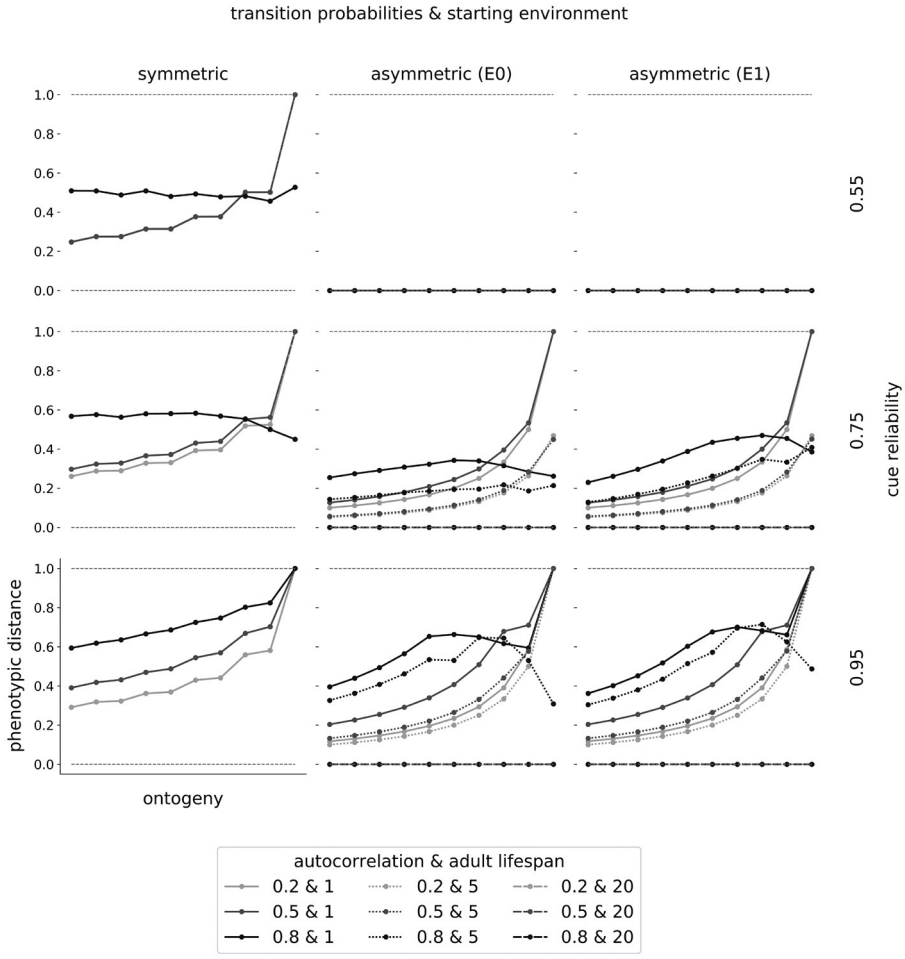


Figure A2.19 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties, and a penalty weight of 0.5. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. Rows indicate different cue reliabilities. Within each panel, we show separate lines for different levels of autocorrelation (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its focal individual. We assign weights to each sequence of cues depending on its likelihood of occurring. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of separation. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

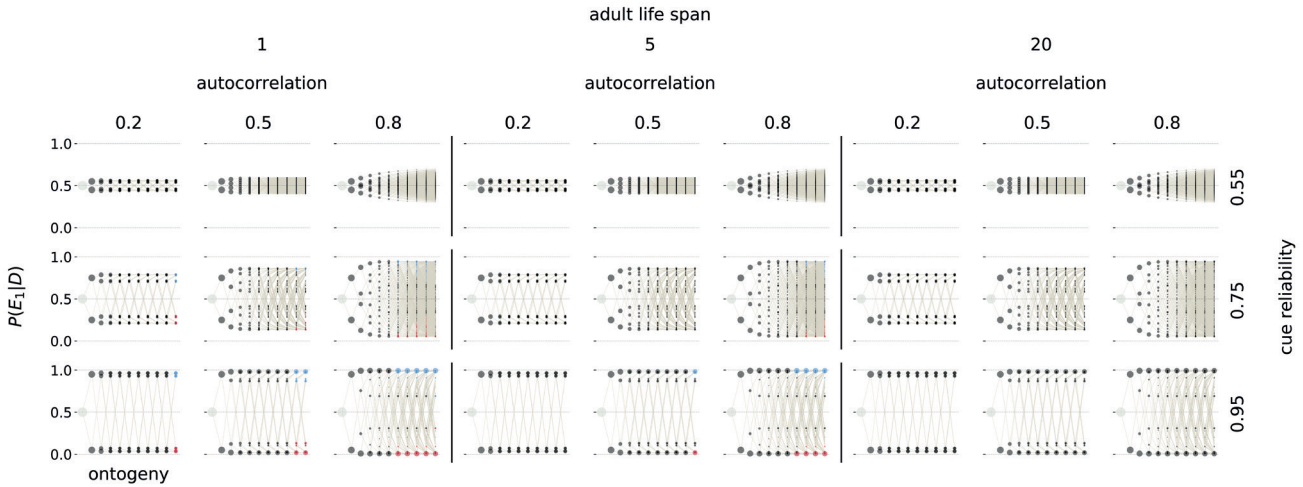


Figure A2.20 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, a penalty weight of 2, symmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P1, red to specializing towards P0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

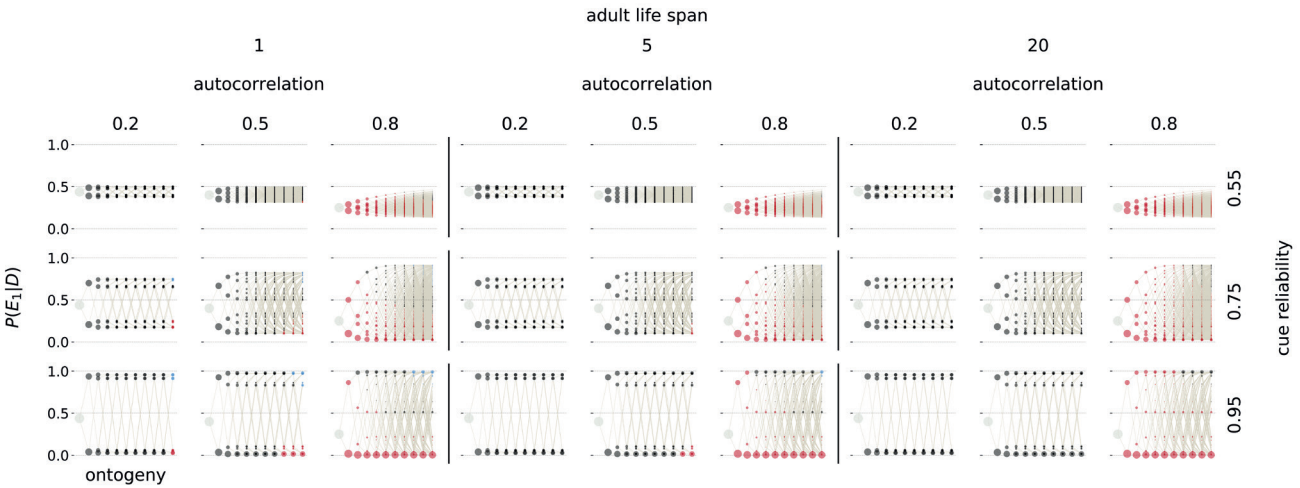


Figure A2.21 Optimal policies across adult life spans. Optimal policies are shown for linear rewards and linear penalties, a penalty weight of 2, asymmetric transition probabilities and T_{adult} = 5, 10 and 20 time periods; one for adult life span per panel. Within each panel, columns indicate different levels of autocorrelation and rows indicate different cue reliabilities. Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The vertical axis displays posterior estimates of being in E_1 and the horizontal axis displays time during ontogeny. At the onset of ontogeny, all organisms start with a prior according to the stationary distribution (large grey circle). Throughout ontogeny organisms sample cues and update their posteriors, resulting in the coloured circles. Colours indicate the optimal, fitness-maximizing phenotypic decision in each state. Pies highlight cases in which organisms with the same estimates make different phenotypic decisions. Black corresponds to waiting, blue to specializing towards P1, red to specializing towards P0, green to ties between all three, and purple to ties between specializing towards the two phenotypic targets. The area of a circle (or pie piece) is proportional to the probability of reaching each state. In each time period, these probabilities sum to 1. Beige lines between states depict possible developmental trajectories.

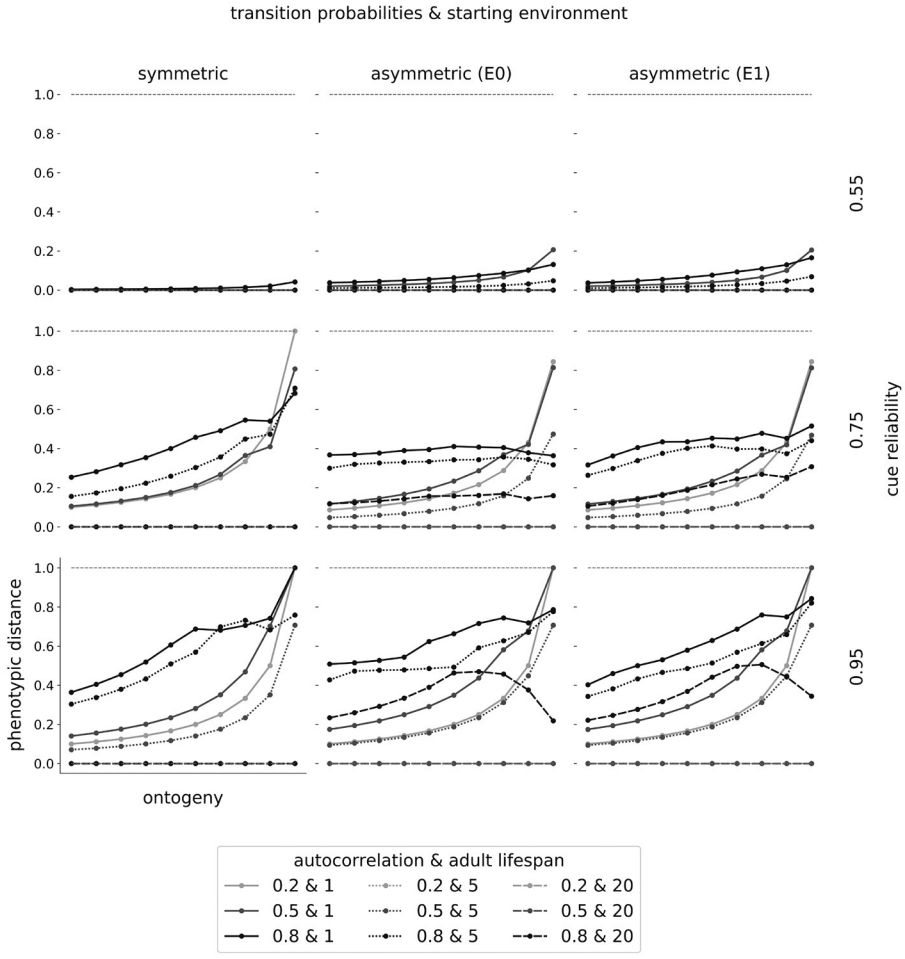


Figure A2.22 Phenotypic plasticity across ontogeny. Phenotypic plasticity across ontogeny is shown for linear rewards and linear penalties, and a penalty weight of 2. Columns indicate whether transition probabilities are symmetric or asymmetric and in the latter case, whether organisms start development in the more (E_0) or less likely (E_1) environmental state. Rows indicate different cue reliabilities. Within each panel, we show separate lines for different levels of autocorrelation (indicated by the colour) and different adult lifespans (indicated by the type of line). Each combination of asymmetry, autocorrelation level, and cue reliability results in a unique Markov process. The different starting environments and adult lifespans represent different possible start- and endpoints of each Markov process. For each combination of a unique Markov process, starting environment and adult lifespan, we conduct T_{ont} = 10 experimental twin studies, one for each $t \in \{1, T_{ont}\}$. We simulate 2^{10} pairs of clones (one for each possible sequence of cues), who follow the optimal policy and get separated at time point t during ontogeny (horizontal axis). After separation one of the clones of each pair, receives reciprocal, opposite cues compared to its focal individual. We assign weights to each sequence of cues depending on its likelihood of occurring. We compute phenotypic distance (vertical axis) as the average, weighted Euclidean distance of all pairs of clones at the end of ontogeny and plot it against the time of separation. Phenotypic distance is normalized by dividing it by the maximally attainable Euclidean distance.

Appendix 3 – Chapter 5

Crimes related to assault (NYC crime data January 2006 – December 2020)

PD description	n	Explanation
Assault 2,1, Peace Officer	17,301	First- and second-degree assault of a police officer or other public safety first responders who are performing their official duties. First degree assault can seriously harm or fatally injure a person. Second- and third-degree assault are less threatening forms of assault.
Assault 2,1, Unclassified	172,380	Unclassified assault is any assault that does not fit with predefined categories.
Assault 3	418,094	Third-degree assault.
Assault Police/Peace Officer	2,183	Assault of unspecified degree.
Rape 1	7,659	First-degree rape involves forcible compulsion and may occur with the use or threatened use of a deadly weapon, kidnapping, infliction of serious bodily injury to the alleged victim, or burglary. It is a class A felony, which allows for up to life imprisonment.
Rape 2	1,823	Second-degree rape occurs in cases that involve forcible compulsion but do not rise to the level of a first-degree offense. It may also occur when an alleged victim is mentally incapacitated or physically helpless, or when the defendant is in a position of authority over the alleged victim, such as in a healthcare or eldercare facility. It is also a class A felony.
Rape 3	3,000	If an alleged offense does not meet the definition of first- or second- degree rape, but still involves "clearly-expressed" lack of consent or threat of harm to the alleged victim's property, the state may charge it as third-degree rape.
Sexual Abuse	556	The infliction of sexual contact upon a person by forcible compulsion.
Sexual Abuse 1	4,303	First-degree sexual abuse
Sexual Abuse 3,2	16,395	Second- and third-degree sexual abuse.
Vehicular Assault (intox. driver)	990	Vehicular assault is defined as causing substantial bodily harm to another person while 1) driving a car under the influence of alcohol or any drug, or 2) driving a car in a reckless manner; or 3) driving a car with disregard for the safety of others.

Table A3.1 Internal police department descriptions of offenses included under 'assault', their frequency of occurrence between January 2006 and December 2020, and explanations.

Boroughs of NYC – population density

Borough	n
The Bronx	2.504.700
Brooklyn	1.385.108
Manhatten	1.585.873
Queens	2.230.722
Staten Island	468.730

Table A3.2 Population densities across NYC boroughs. Assaults occurring in a 5km radius around our regions of interest (i.e. Morrisania, Brownsville, Ozone Park, Upper East Side, and Tottenville) fall into different boroughs. We corrected the assault rates for each assault based on the population density of the borough they belong to.

Regions of interest – different resolution

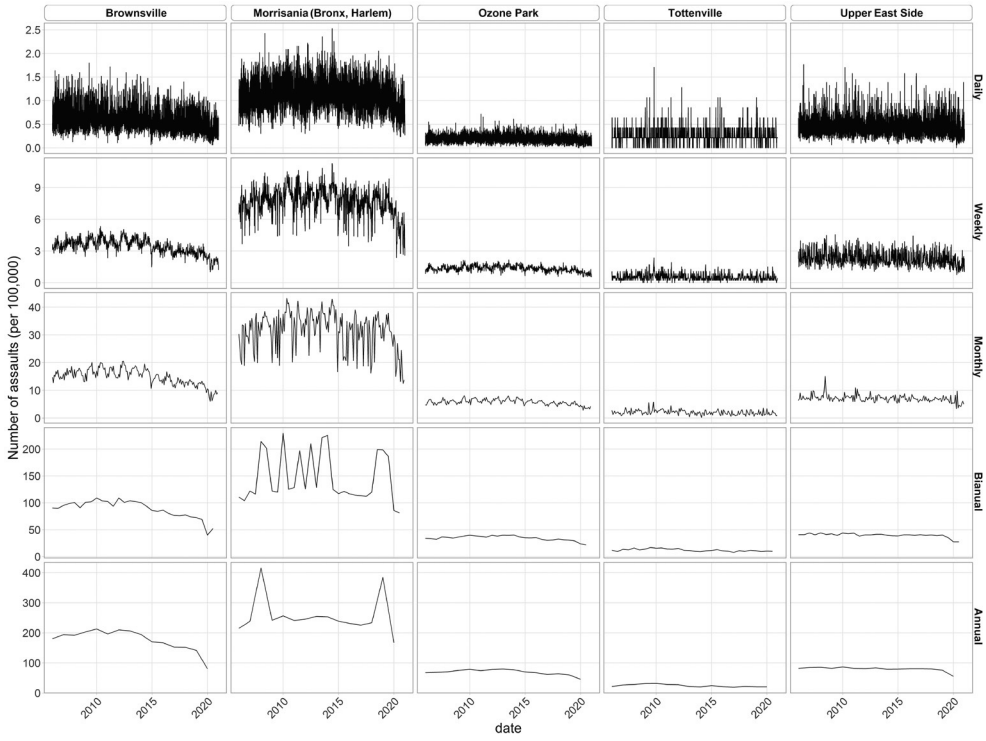


Figure A3.1 Regions of interest in the New York Crime data. Rows indicate different temporal resolutions, namely daily, weekly, monthly, biannually, and annually (from top to bottom) and columns show regions of interest. Each individual panel plots assault rates (summed according to their temporal resolution) against time between January 2006 and December 2020.

Properties of time series data

Time series decomposition

Any time series can be decomposed into three individual parts: trend, season, and random component. The trend describes how the level of a time series changes with time. Season refers to the presence of reoccurring patterns within a calendar year. Identifying the presence of seasonal patterns requires us to be able to partition the data into natural time units such as days, weeks, or months. The random component is what is left of the data after subtracting the trend and seasonal patterns. It is the irregular, non-systematic variation in the time series. Together, these components describe how the time series behaves across the observation period. If the goal is not to explain a specific non-random component (e.g. the trend), it is usually advised to remove trend and season before analysing the time series. In these cases, a time series is usually made stationary prior to further analysis.

A time series can be decomposed with an additive or multiplicative model. The additive model assumes that the raw time series is the sum of its trend, season, and random component. The multiplicative model represents the series as the product of these components. The additive model is suited for time series with constant seasonal variation over time, whereas the multiplicative model is suited for series with increasing seasonal variation. Visually, an additive time series can be identified by its constant frequency and amplitude of seasonal peaks. In a multiplicative model these components vary across time, typically increasing or decreasing.

Changepoints

Across time, the mean level and variance of a time series can suddenly change. A high number of changepoints in mean and variance can make the time series less predictable. Changepoints in an individual's time series may indicate life events, such as job changes or moves. Changepoints which are common to a whole sample may indicate exogenous events, such as switches in political leadership or a financial crisis. We used the packages changepoint and changepoint.np (Haynes et al., 2016; Killick & Eckley, 2014) to estimate changepoints in mean and variance individually and simultaneously. The package offers a large variety of methods to estimate the number and position of changepoints. Our framework includes the modified Bayes information criterion (MBIC) and the Akaike information criterion (AIC) as penalties for selecting between models with different numbers of changepoints (Zhang & Siegmund, 2007). Typically, the MBIC penalizes models with a large number of parameters more strongly, resulting in more parsimonious models with fewer changepoints. Other functionality and penalties from the changepoint package can be easily added to our framework.

Stationarity

A time series whose statistical properties change across time is called non-stationary (Jebb et al., 2015; Young et al., 2020). Observing an increasing trend or changes in variance in the data is therefore a strong indicator of non-stationarity. In this way, the presence of stationarity in itself tells us something about a series' behaviour over time. Many time series models assume that a time series is stationary or can be transformed into a stationary series. Removing systematic variation, such as changes in mean and variance, will often be sufficient to make a time series stationary. A series can be 'detrended' by computing the difference between values in subsequent time periods; this stabilizes the mean. In rare cases, more than one round of differencing is necessary to achieve a constant mean. Taking the logarithm of the time series usually helps to achieve constant variance across time. However, it is advisable to inspect the time series after differencing and to only apply additional measures if necessary (Jebb et al., 2015). The augmented Dickey-Fuller test is often used to assess whether a time series is stationary, where a significant outcome indicates stationarity.

Autocorrelation

Another property of time series is the autocorrelation. Autocorrelation indicates whether the time series is correlated with itself at different lags. A high autocorrelation implies that the current values of a series predict the subsequent values. This is called a lag-1 autocorrelation. Autocorrelation can be also computed for different lags, allowing us

to assess whether current values predict values that are two, three, or four time units away, resulting in lag-2, lag-3 and lag-4 autocorrelations. The higher the autocorrelation, the easier it is to predict future values of a time series based on current values. One problem with the autocorrelation at higher order lags is that the relationship between current values and values at later time periods partially depends on the correlation between current values and intermediate values. To account for this dependency, it is common to additionally compute the partial autocorrelation which removes the influence of intermediate lags.

Autocorrelation and partial autocorrelation can be computed for the raw time series, but also for transformed series such as the stationary time series. Autocorrelation of the stationary time series indicates how predictable the time series is after the systematic changes in statistical properties, such as mean and variance, have been removed.

Framework user manual

The framework is divided into three phases: *Exploration*, *Preprocessing*, and *Extracting statistics*.

Exploration

Our framework can be applied to the raw data directly or to a preprocessed version. To aid the user in making preprocessing choices and in selecting environmental statistics from our framework, we offer the opportunity to first visually explore the data. We offer plots of the raw time series for randomly chosen participants, the decomposed time series, the distribution of augmented Dickey-Fuller test values, the autocorrelation and partial autocorrelation at different lags of the raw and stationary time series, and changepoints in mean and variance (Figure A3.2). For plots related to stationarity, the user can explore different values for the degree of differencing and enable or disable logarithmic transformation of the data.

Preprocessing

During the preprocessing stage the user can remove the trend and/or seasonal components from the time series and apply differencing and/or logarithmic transformations to make the time series stationary. Additionally, we offer the opportunity to split the data at different time periods, resulting in multiple, smaller datasets. This may be useful if the user is interested in separately exploring environmental statistics before and after a specific age or specific time period. After preprocessing it is possible to extract environmental statistics from the resulting data (or datasets) during the next stage of the framework.

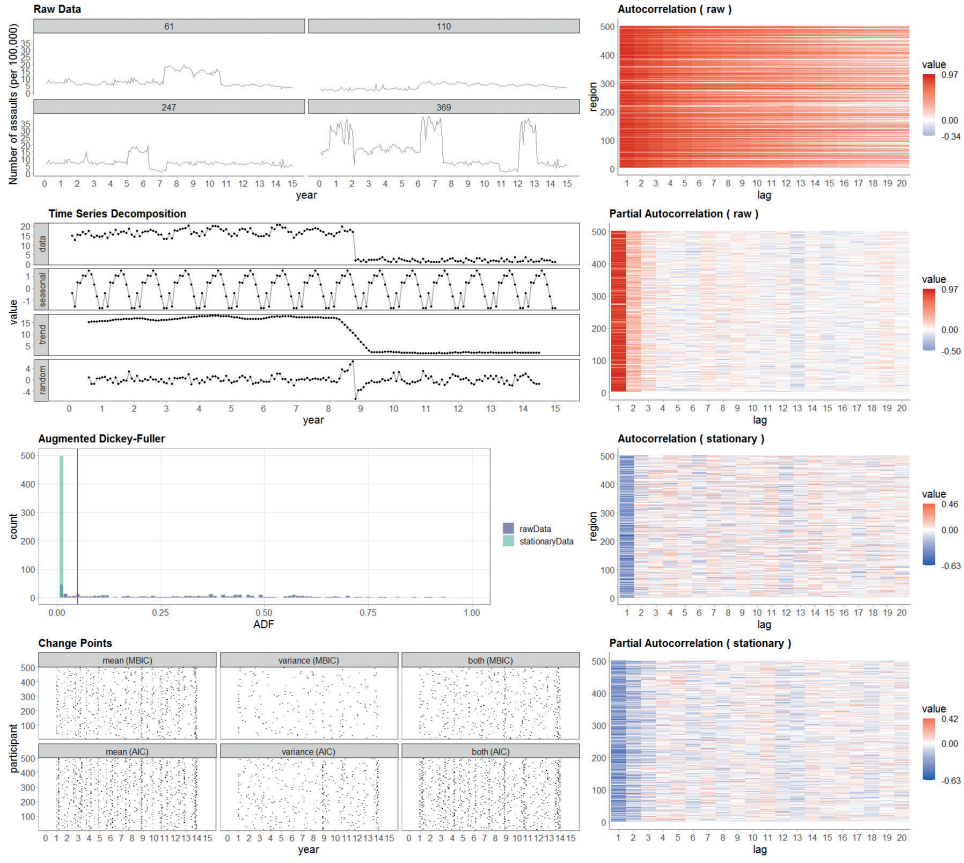


Figure A3.2 Exploratory plots created by our framework. From top to bottom and left to right: our framework plots the raw data for individual participants, the time series decomposition of individual participants, the distribution of augmented dickey-fuller values indicating stationarity, the distribution of changepoints in mean, variance, or both using different computational methods, the distribution of autocorrelation and partial autocorrelation values in the sample of the raw time series and stationary time series.

Extracting statistics

This stage is the heart of our framework. Here, we extract the statistics listed in Table 5.1 (Chapter 5) from the preprocessed dataset. In order to estimate a slope in the data the user can specify a model that is fitted to each participant's time series. In the simplest case this is just a linear model with time as the predictor and the environmental variable as the outcome. The user could also apply more complicated models, such as a nonlinear, polynomial model. As a default the framework also extracts some statistics applied to a participant's stationary time series, as well as the squared deviation from the mean. The latter set of statistics can help to identify systematic changes in variance. In addition to the numeric output, our framework also offers various options for visualizing the extracted statistics.

Analysis

This step is not part of our framework but likely the goal of most users. Once statistics have been extracted and selected, the user may use them for subsequent analysis. If the user attempts to compute inferential statistics for the entire sample using the extracted statistics as predictors, they should ensure a minimum sample size that is appropriate for the fitted model.

Bibliography (Supplementary materials)

Haynes, K., Killick, R., Fearnhead, P., & Eckley, I. (2016). changepoint. np: Methods for nonparametric changepoint detection. *R Package Version 0.0*, *2*.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*(JUN), 1–24. https://doi.org/10.3389/fpsyg.2015.00727

Killick, R., & Eckley, I. A. (2014). changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, *58*(3).

Young, E. S., Frankenhuis, W. E., & Ellis, B. J. (2020). Theory and measurement of environmental unpredictability. *Evolution and Human Behavior*, *February*, 0–1. https://doi.org/10.1016/j.evolhumbehav.2020.08.006

Zhang, N. R., & Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*(1), 22–32. https://doi.org/10.1111/j.1541-0420.2006.00662.x

Summary

Dutch and English academic summaries

Nederlanse samenvatting

Gevoelige perioden zijn perioden (of levensfasen) waarin de ontwikkeling van een organisme meer wordt beïnvloed door ervaringen dan in andere perioden (of fasen). Gevoelige perioden zijn het gevolg van ontwikkelingsplasticiteit, d.w.z. het vermogen van een genotype om afhankelijk van de ervaring verschillende fenotypes te produceren. Ondanks enorme vooruitgang in het begrijpen van de fysiologische mechanismen van gevoelige perioden, weten we weinig over hun evolutie. Wiskundige modellen worden in toenemende mate gebruikt om te onderzoeken onder welke omgevingscondities gevoelige perioden worden begunstigd door natuurlijke selectie. Deze modellen vinden dat gevoelige perioden de neiging hebben om vroeg in het leven te evolueren en dat plasticiteit doorgaans afneemt over de ontogenie heen (d.w.z. de levensfase die relevant is voor de ontwikkeling van een eigenschap), en vaak nul bereikt (**Hoofdstukken 1 en 2**). We weten echter dat gevoelige perioden ook in latere levensfasen kunnen bestaan, zoals puberteit of adolescentie, en dat de resterende plasticiteit laat in de ontwikkeling kan blijven.

Bestaande modellen van gevoelige periode-evolutie bieden weinig inzicht in omstandigheden die verbeterde plasticiteit in latere ontwikkelingsstadia bevorderen. Onderzoekers hebben echter gespeculeerd dat natuurlijke selectie de voorkeur zou kunnen hebben aan gevoelige perioden later in de ontwikkeling, wanneer organismen variatie ervaren in twee hoofdfactoren: de mate waarin ervaringen ('cues') onzekerheid over de omgeving kunnen verminderen ('cue-betrouwbaarheid'), en de omgevingstoestand zelf. De modellen die in dit proefschrift worden gepresenteerd formaliseren deze ideeën. Het eerste model onderzoekt de evolutie van gevoelige perioden wanneer organismen variatie ervaren in de betrouwbaarheid van signalen; het tweede model onderzoekt variatie in de omgevingstoestand over ontogenie. Mijn eerste model laat zien dat, wanneer de betrouwbaarheid van cues over de ontogenie heen toeneemt, gevoelige perioden kunnen evolueren in latere ontwikkelingsstadia (**Hoofdstuk 3**). Wanneer de cue-betrouwbaarheid afneemt over de ontogenie, evolueren gevoelige perioden enkel aan het begin van de ontwikkeling. Over het algemeen lijkt natuurlijke selectie aangepaste niveaus van plasticiteit te hebben om de betrouwbaarheid van cues te volgen. Mijn tweede model laat zien dat, wanneer de cue-betrouwbaarheid constant is over de ontogenie, maar de omgevingstoestand fluctueert, er gevoelige perioden kunnen optreden aan het begin, halverwege en zelfs tegen het einde van de ontogenie (**Hoofdstuk 4**). Deze bevinding contrasteert bevindingen van eerdere modellen van gevoelige periode-evolutie waarin plasticiteit vaak nul bereikt en nooit toeneemt tegen het einde van ontogenie. Ongeacht wanneer plasticiteit tijdens de ontogenie piekt, behouden organismen altijd een resterende plasticiteit laat in de ontogenie wanneer de omgeving fluctueert. Mijn resultaten suggereren dat het onwaarschijnlijk is dat kritieke perioden, waarna de plasticiteit nul wordt, de voorkeur hebben in fluctuerende omgevingen.

Om synergiën tussen modellen en empirische gegevens mogelijk te maken, heb ik ook een raamwerk ontwikkeld voor het bestuderen van omgevingsstatistieken over ontwikkeling heen. Omgevingsstatistieken drukken noties van stabiliteit en verandering van het omgeving uit in ondubbelzinnige en formele termen, d.w.z. als statistische definities.

Beschouw mijn tweede model dat ervan uitgaat dat de omgeving fluctueert tijdens het leven van een organisme. In plaats van alle mogelijke snelheden van omgevingsfluctuaties te onderzoeken, zou het kennen van de waarden van omgevingsstatistieken me kunnen helpen om me te concentreren op snelheden die relevant zijn voor een specifieke soort of eigenschap. Op deze manier zouden mijn resultaten relevanter zijn voor onderzoekers die deze soorten of eigenschappen bestuderen.

Hoewel kennis van de waarden van omgevingsstatistieken nuttig is, is het ook van belang de statistische definities van omgevingsconstructies te weten. Stabiliteit en verandering van de omgeving staan beide centraal in de ontwikkelingswetenschap. Het meeste ontwikkelingsonderzoek veronderstelt, claimt of onderzoekt enkele noties van stabiliteit en verandering in het omgeving. De huidige norm is om noties van stabiliteit en verandering van het omgeving (bijv. variabiliteit, onvoorspelbaarheid, instabiliteit) te beschrijven met behulp van natuurlijke taal, die vaak ambigu is. Deze ambiguïteit verzwakt de match tussen theorie en methoden binnen studies, en leidt tot inconsistenties tussen studies. Ik heb een raamwerk gepresenteerd dat noties van stabiliteit en verandering van het omgeving in ondubbelzinnige en formele termen organiseert (**Hoofdstuk 5**). Het raamwerk is gebaseerd op statistische definities van stabiliteit en verandering in het omgeving die al op grote schaal worden gebruikt in andere disciplines, zoals biologie en ecologie. Om de haalbaarheid aan te tonen, pas ik het raamwerk als een casestudie toe op een dataset van misdaadcijfers in New York City over 15 jaar, met de nadruk op 'onvoorspelbaarheid'. Sommige resultaten generaliseren over statistische definities, en andere zijn afhankelijk van welke statistische definities worden gebruikt. Dit is van belang voor onderzoek in de psychologie dat zich richt op individuele uitkomsten: verschillende definities kunnen leiden tot verschillende conclusies over de impact van onvoorspelbaarheid op belangrijke levensuitkomsten zoals gezondheid, welzijn en psychopathologie. Ten slotte heb ik besproken hoe mijn werk bijdraagt aan de integratie van wetenschappelijke disciplines, en heb ik toekomstige ideeën gepresenteerd voor het modelleren van de evolutie van gevoelige perioden en het uitbreiden van mijn raamwerk van omgevingsstatistieken (**Hoofdstuk 6**).

English summary

Sensitive periods are times (or life stages) in which an organism's development is more affected by experiences than at other times (or stages). Sensitive periods result from developmental plasticity, i.e., the capacity of a genotype to produce different phenotypes depending on experience. Despite immense progress in understanding the physiological mechanisms of sensitive periods, we know little about their evolution. Mathematical models are increasingly used to study under which environmental conditions sensitive periods are favoured by natural selection. These models find that sensitive periods tend to evolve early in life and that plasticity typically declines across ontogeny (i.e., the life stage relevant for the development of a trait), often reaching zero (**Chapters 1 and 2**). However, we know that sensitive periods can also exist at later life stages, such as puberty or adolescence, and that residual plasticity can remain late in development.

Existing models of sensitive period evolution offer little insight into conditions favouring enhanced plasticity at later developmental stages. However, researchers have speculated that natural selection might favour sensitive periods later during development when organisms experience variation in two main factors: the extent to which experiences ('cues') can reduce uncertainty about the environment ('cue reliability'), and the environmental state itself. The models presented in this dissertation formalize these ideas. The first model explores the evolution of sensitive periods when organisms experience variation in the reliability of cues; the second model explores variation in the environmental state across ontogeny. My first model shows that, when the reliability of cues increases across ontogeny, sensitive periods can evolve at later developmental stages (**Chapter 3**). When cue reliability decreases across ontogeny, sensitive periods only evolve at the onset of development. Overall, natural selection appears to have adapted levels of plasticity to track the reliability of cues. My second model shows that, when cue reliability is constant across ontogeny but the environmental state fluctuates, sensitive periods can occur at the onset, midway through, and even towards the end of ontogeny (**Chapter 4**). This finding contrasts findings from previous models of sensitive period evolution in which plasticity often reaches zero, and never increases towards the end of ontogeny. Regardless of when during ontogeny plasticity peaks, organisms always retain residual plasticity late in ontogeny when the environment fluctuates. My results suggest that critical periods, after which plasticity reaches zero, are unlikely to be favored in fluctuating environments.

To facilitate synergies between models and empirical data, I have also developed a framework for studying environmental statistics across development. Environmental statistics express notions of environmental stability and change in unambiguous and formal terms, i.e., as statistical definitions. Consider my second model which assumes that the environment fluctuates within an organism's lifetime. Rather than exploring all possible rates of environmental fluctuations, knowing the values of environmental statistics could help me focus on rates that are relevant for a specific species or trait. In this way, my results would be more relevant to researchers studying these species or traits.

While knowledge of the values of environmental statistics is useful, so is knowing the statistical definitions of environmental constructs. Environmental stability and change are both central to developmental science. Most developmental research assumes, claims, or examines some notions of environmental stability and change. The current norm is to describe notions of environmental stability and change (e.g., variability, unpredictability, instability) using natural language, which tends to be ambiguous. This ambiguity weakens the match between theory and methods within studies, and leads to inconsistencies across studies. I have presented a framework that organizes notions of environmental stability and change in unambiguous and formal terms (**Chapter 5**). The framework draws on statistical definitions of environmental stability and change that are already widely used in other disciplines, such as biology and ecology. To demonstrate feasibility, I apply the framework to a dataset of crime rates in New York City across 15 years, focusing on 'unpredictability' as a case study. Some results generalize across statistical definitions, and others depend on which statistical definitions are used. This matters for research in psychology focusing on individual outcomes: different definitions can lead to different conclusions about the impact of unpredictability on key life outcomes such as health, wellbeing, and psychopathology. Finally, I have discussed how my work contributes towards the integration of scientific disciplines, and presented future ideas for modeling the evolution of sensitive periods and expanding my framework of environmental statistics (**Chapter 6**).

Acknowledgements

This dissertation was 4.5 years and one pandemic in the making. When reflecting on these past few years the phrase 'it takes a village' comes immediately to mind. It did take a village to finish this PhD. Luckily, my village was (and is) filled with wonderful colleagues, friends, and family whom I want to thank.

First of all, I would like to thank my supervisors. Willem, this PhD would have not been possible without you. You shaped it and me as a researcher in countless ways. I am one of your very first PhD students. Looking back, I think it is fair to say that we both grew into and with our roles. Thank you for seeing potential in me, for all the resources and opportunities you provided, for your contagious passion for research, and for always trying to help me grow and develop. Mostly, I thank you for caring about me as a person and for being both a friend and mentor. I am excited to see where the next few years will take us. Karthik, I am so glad you officially joined my supervision team. Your input on our projects always feels like someone is sprinkling glitter all over it (that's a good thing). I look up to you as a scientist and mentor. Also, you are just a really interesting person to hang out with. Thank you for everything. I hope that we will keep working together and that someday you might move to Europe. Toon, I've known you as my promotor, leader of the social development group (program? I am never sure), and director of the institute. You are one of the most hard-working people I have ever met. Despite your role within the institute and how busy you are, I knew I could always knock on your door, and I never felt too intimidated to do so. You have lunches and campus dinners with master students, PhDs, and senior staff alike, never imposing any hierarchy. The BSI is extremely lucky to have you and I was lucky to have you as my promotor. Thank you.

Bill, you are not my supervisor and I am very happy about that. If you were, I wouldn't have been able to get the unfiltered Bill-experience. However, you always did have a mentor-like presence during my PhD. You cared about my well-being and were always looking out for me. I appreciate all work- and non-work-related advice I have ever received from you. Thank you. Anyone who can officially call you their mentor is extremely lucky. Bernd, in a parallel universe you would have stayed on my supervision team and been part of my projects. And, I am sure it would have been awesome. Luckily, I got to keep working with you during my reoccurring gig as a teaching assistant for the statistics classes. You, Bill, and me not working together feels like the end of an era. I will surely miss your creativity when it came to finding ways to make me grade all those assignments. For a Swiss person, you are a great German.

A big thank you goes to the BSI management: Meta, Ralf, and new Ralf a.k.a Sander. You are doing (did in Ralf's case) such an amazing job at keeping things running smoothly. Meta, you are just one of a kind. You make every lunch and BSI event more enjoyable by bringing equal amounts of warmth and sass to the table. You are the definition of 'work hard, play hard'. Thank you all for everything you do. In the same spirit, I would like to thank Marieke who does countless things for our department behind the scenes. To all my

colleagues from social development: thank you. I am really anxious and sad to leave this group. You have been warm and welcoming. I enjoyed our lunches, dinners, events, and writing weeks. On all of these occasions I never felt obligated to spent time with colleagues but rather like I was meeting friends. It doesn't help my sadness that we keep hiring staff and PhDs who are really lovely people. My next department has very big shoes to fill. I would also like to thank all members of any lab groups I attended with special thanks to the people from the DEEP lab. This lab feels like a collection of special individuals handpicked by Willem and I enjoy every bit of it. I hope that we get to have more in-person borrels in the near future.

I also have friends to thank. Jesse, you are one of my very best friends. You are fiercely loyal and really just a big old softy deep down. Some days you're also a big pain in the ass. I often don't know which one it's going to be, but I always know you have my back (even if you're stabbing it a bit). Our time together as officemates in the old building will always have a special place in my heart. Fenneseksi4Eva. Thank you for being such a big part (not a fat joke) of my journey. Achiel, you have demonstrated that genes are not the whole story and that the environment matters. Over the past few years, you have become your own human being and carved out a space that is just Achiel. As much as we all love Jesse, one is enough. Thanks for being you and a great friend. Together you two have always been there for me and supported me when I needed it the most. You're like family (the weird uncles that you keep inviting, although you know they're going to misbehave).

Kevin and Wernard, thank you for all the times we hung out to have drinks together and for the welcome distraction from work. You both express friendship like 5-year-old boys who show affection by bullying other people (at least I hope so) but sometimes you catch me off-guard by being surprisingly sweet and caring. I am grateful for your friendship.

Sarah and Desi, you are as different as two people can be. To me you both are badass, inspiring women that I look up to and care about deeply. I didn't realize how much I needed a female perspective in my life until I became friends with you both. I wouldn't want to miss it ever again. Thank you for the empathy, understanding, and fun times.

Geert, we started out as colleagues but became friends very quickly. You are just a really lovable, frank, and funny human being. I want all the good things for you and am glad that we are in each other's lives. Thanks also to Ruby and Ruben, who (not surprisingly) are just as lovable as you. You and Ruby have some mad baby-making skills because Ruben is such a joy. Niklas, you bring a level of Germanness into my life that I didn't even know I missed. You are definitely crazy and I am a hundred percent there for it. I hope we'll get to see each other more often. Luuk and Marenthe, thank you for destigmatizing couples' nights by making them fun. I am so glad we found people who share our (unhealthy) enthusiasm for niche boardgames and the financial burden that comes with it. Also thank you for just being awesome people and for Valiente. Here, I should also mention Jeroen and Thomas who have defeated me in boardgames countless times. Thank you for keeping my mind sharp. Ethan, in a venn diagram you would live in the intersection of 'friend', 'colleague', and 'mentor'. Thank you for all the different roles you played since I have known you and for all that's yet to come. Also, thanks to all the rest of the people who left and joined our

sanctuary, the Fennemonium, over the years: Silke, Paul, Bella, Ili, Tim, Sterre, Dina, Cata, Peter, Hannah, Yannick, Ivana, Jingu, Jinny, Krisna, Julian and many more. It has been a very special academic bubble with all of you guys and I am looking forward to continuing living in it. Also, more group holidays, please. I would also like to thank the people who have made transitioning to Utrecht easier: thank you, Methee, Ajki, Mathijs, and Edith for making me feel at home.

Stephi und Jana, euch möchte ich für die jahrelange Freundschaft danken und dafür, dass ihr euch meine langweiligen Arbeitsgeschichten antut. Obwohl wir uns selten sehen, fühlt es sich immer direkt vertraut und gut an. Ohne euch wäre die Schulzeit auch furchtbar gewesen. Ihr habt also definitiv einen großen Einfluss auf meinen Werdegang gehabt. Danke für alles. Danke auch an all die Freunde von meinem Bachelor, die dieses Dankwort vielleicht gar nicht zu Gesicht bekommen. Meine ersten Studienjahre waren großartig und das liegt an euch. Danke.

Anton, you have literally made every day that I have known you better. Daily pig-me ups and piggy-back rides got me through paper rejections and a global pandemic. I'm really glad you wore those tight pants when I met you and that they made me ask for your number (shoutout to our cupid Kevin). Thanks for everything you do for me. Thanks also to your family who welcomed me so warmly.

Danke an meine Familie und insebesondere an meine Eltern. Mama, Papa, danke, dass ihr immer für mich da seid und an mich glaubt. Euch stolz zu sehen war immer großer Teil meiner Antriebskraft. Ohne euch wäre ich heute nicht hier und ich hoffe, dass ihr auch stolz auf euch selbst seid.

Lastly, thank you to all you wonderful people with pets out there. Thank you for photos, videos, and in-person cuddle sessions (with the pets that is). I love every single one of them and they always make my day.

About the author

Nicole was born in Nordhorn, Germany, on February 18th, 1992. In 2015, she obtained her Bachelor's degree (B.Sc. with distinction) at the University of Osnabrück (Germany) in cognitive science. As part of her studies she completed an internship in the Nicolelis laboratory at Duke University in Durham, North Carolina (USA). Her work on the decoding of neural signals for brain computer interfaces became part of her Bachelor's thesis. Before starting a Master's degree, Nicole briefly lived in Barcelona in 2015. There, she interned at the Computer Vision Center in the learning and machine perception group. In 2017, she obtained a Master's degree (M.Sc. cum laude) in computing science with specialization in data science from Radboud University, Nijmegen (the Netherlands). During her studies Nicole interned at TNO in The Hague (the Netherlands) where she worked on natural language processing of patient cancer forums. Her time at TNO sparked her interest in research and she decided to pursue a PhD. Nicole worked on her PhD at the Behavioural Science Institute at Radboud University in Nijmegen between 2017 and 2022. In her dissertation she focused on modeling the evolution and development of sensitive periods. This topic allowed her to integrate many interdisciplinary interests: it combines elements from computer science, biology, neuroscience, and psychology. Starting in April 2022, Nicole will be working as a postdoctoral researcher at the University of Utrecht (The Netherlands). She will continue her research on modeling sensitive periods, as well as on quantifying environmental statistics in empirical data.

Publications

PEER-REVIEWED ARTICLES

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2022) Sensitive periods, but not critical periods, evolve in a fluctuating environment: a model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1969), 20212623. https://doi.org/10.1098/rspb.2021.2623

Walasek, N., Frankenhuis, W. E., & Panchanathan, K. (2021). An evolutionary model of sensitive periods when the reliability of cues varies across ontogeny. *Behavioral Ecology*, 33(1), 1–14. https://doi.org/10.1093/beheco/arab113

featured in: Gee, D. G. (2022). When do sensitive periods emerge later in development? *Trends in Cognitive Sciences*, *26*(2), 97–98. https://doi.org/10.1016/j.tics.2021.12.001

Frankenhuis, W. E., & **Walasek, N**. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience, 41*, 100715. https://doi.org/10.1016/j.dcn.2019.100715

Johannes, N., Meier, A., Reinecke, L., Ehlert, S., Setiawan, D. N., **Walasek, N.**, Dienlin, T., Buijzen, M., & Veling, H. (2020). The relationship between online vigilance and affective well-being in everyday life: Combining smartphone logging with experience sampling. *Media Psychology,* 1–25. https://doi.org/10.1080/15213269.2020.1768122

Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., **...** & van der Gugten, R. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical Image Analysis*, *42*, 1-13.

van Oortmerssen, G., Raaijmakers, S., Sappelli, M., Boertjes, E., Verberne, S., **Walasek, N.**, & Kraaij, W. (2017). Analyzing cancer forum discussions with text mining. In: Riano D., Peleg M., Lenz R., Reichert M., Denecke K., Deng Y., Declerck T., Harmelen F. van (red.) *Proceedings of the International Joint Workshop KR4HC - ProHealth 2017 (in conjunction with AIME 2017),* 127-130*.*

MANUSCRIPTS UNDER REVIEW

*Fenneman, J., ***Walasek, N.**, *Griep, Y. Theory construction in organizational behavior: a case for formal modeling. *Group & Organization Management.* *Equal contributions

MANUSCRIPTS IN PREPARATION

Walasek, N., Young, E. S., & Frankenhuis, W. E. (2022) A framework for studying environmental statistics in developmental science. *Child Development.*

Behavioural
Science
Institute

Radboud University