The Aims and Structures of Research Projects That Use Gene Regulatory Information with Evolutionary Genetic Models

by

Steve Elliott

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved January 2017 by the
Graduate Supervisory Committee:

Manfred Laubichler, Co-Chair
Richard Creath, Co-Chair
Brad Armendt
Patrick Forber
Stephen Pratt

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

At the interface of developmental biology and evolutionary biology, the very criteria of scientific knowledge are up for grabs. A central issue is the status of evolutionary genetics models, which some argue cannot coherently be used with complex gene regulatory network (GRN) models to explain the same evolutionary phenomena. Despite those claims, many researchers use evolutionary genetics models jointly with GRN models to study evolutionary phenomena.

How do those researchers deploy those two kinds of models so that they are consistent and compatible with each other? To address that question, this dissertation closely examines, dissects, and compares two recent research projects in which researchers jointly use the two kinds of models. To identify, select, reconstruct, describe, and compare those cases, I use methods from the empirical social sciences, such as digital corpus analysis, content analysis, and structured case analysis.

From those analyses, I infer three primary conclusions about projects of the kind studied. First, they employ an implicit concept of gene that enables the joint use of both kinds of models. Second, they pursue more epistemic aims besides mechanistic explanation of phenomena. Third, they don't work to create and export broad synthesized theories. Rather, they focus on phenomena too complex to be understood by a common general theory, they distinguish parts of the phenomena, and they apply models from different theories to the different parts. For such projects, seemingly incompatible models are synthesized largely through mediated representations of complex phenomena.

The dissertation closes by proposing how developmental evolution, a field traditionally focused on macroevolution, might fruitfully expand its research agenda to include projects that study microevolution.

DEDICATION

To the clear and *distinct* memories of Linda Lou Elliott (1946–2005) and Walter Wendell Elliott, Jr. (1941–2008).

ACKNOWLEDGMENTS

I gratefully acknowledge the School of Life Sciences for a Fall 2016 Dissertation Completion Fellowship, which enabled me to finish this project. Thank you to the National Institute for Mathematical and Biological Synthesis in Knoxville, Tennessee, and to Stevan J. Arnold and Joe Felsenstein, who accepted me to their Evolutionary Quantitative Genetics Tutorial at the Institute during the summer of 2014. Their instruction helped me understand the mathematics discussed in this dissertation. Thank you to various organizations that financially supported this research and its dissemination: the Center for Biology and Society, the School of Life Science, ASU's Graduate and Professional Student Association, and the National Science Foundation.

Many good people made this dissertation possible. Thanks are due:

To those who shepherded me through ASU's bureaucracy with unwavering goodwill: Rob Booher, Andrea Cotrell, Yvonne Delgado, Scott McAdams, Rick Olson, Wendi Simonson, Judy Swartz, Linda Zarroli, and especially Jessica Ranney.

To those who took extra efforts to ensure I developed into a researcher: Karin Ellison, Greg Fitch, Jürgen Gadau, Andrew Hamilton, Ted Humphrey, Jürgen Liebig, Joan McGregor, Martin Wojciechowski, and especially Michael White.

To the wonderful group of students and researchers who made the Center for Biology and Society a receptive and interesting place: Ken Aiello, Jovanna Anzaldua, Caroline Appleton, the Abbouds, Liz Barnes, Zane Bartlett, Jenny Brian, Tito Carvalho, Matt Chew, Cecilia Chou, Jennifer Craer, Julia Damerow, Mary Drago, Kelle Dhein, Kevin Gleason, Lijing Jiang, Nathan Johnson, Grace Kim, Katherine Liu, Claudia Nunez-

Eddy, Deryc Painter, Erick Peirson, Christopher Rojas, Christian Ross, David Steffes, Nevada Wagoner, Johnny Winston, Grant Yamashita, and Yawen Zou.

And to those colleagues who each (perhaps unwittingly) now safeguard a piece of my heart with them wherever they go: Carlo Altamirano-Allende, Wes Anderson, Badingquiying, Marci Baranski, Sean Cohmer, Nathan Crowe, Cera Fisher, Michael Gifford, J.J. LaTourelle, Kate MacCord, Paige Madison, Catherine May, Karla Moeller, Ernest Nkansah-Dwamena, Erica O'Neil, Valerie Racine, Chanapa Tantibanchachai, Federica Turriziana Colonna, Mark Ulett, Karen Wellner, and Justin Wolter. Guido Caniglia has two pieces.

To Jane Maienschein, who rescued me from myself, repotted me, and entrusted me with the Embryo Project.

To those who had special impact on this dissertation, either by teaching me computational methods or by commenting on early drafts of chapters: Erick, Julia, Deryc, Ken, Wes, Valerie, Yawen, Kate, and Guido.

Most importantly, to my committee members. They believed in me when I didn't, they drove me to excel, and they patiently waited as I concocted the mongrel document before you. Richard Creath, Manfred Laubichler, Brad Armendt, and Stephen Pratt have shaped my intellect more than any others. Patrick Forber supported this project even after I got him lost in the Superstition Mountains. If this dissertation does nothing else, I hope only that it earns me their respect.

To my kin: The Fahrenheit and the Yaaaar!

Finally, to Kelley Karnes. The five years of two-more-years are over.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This dissertation is about the problem of knowledge at the interface of developmental biology and evolutionary biology. In this chapter, I review the notion of problem of knowledge, then I show how it is relevant to the interface between developmental biology and evolutionary biology. Next I list the questions that drive this dissertation, and I conclude with a list of questions about meta-scientific methodology, questions that must be addressed as preliminaries to the driving questions. Throughout, I preview the later chapters of this dissertation.

1.1- Problem of Knowledge

The phrase "problem of knowledge" refers to a set of questions that we ask about knowledge itself, especially scientific knowledge in its specialized domains, fields, or disciplines.[1] The questions are related to each other, but exhibit no hierarchy of importance in relation to each other. For a given discipline or field, the set of questions traditionally includes at least:

What questions do those in the discipline ask?

How do they address those questions?

By what criteria do they evaluate the quality of responses to questions?

[1] See the introduction to (Cassirer 1950) for a succinct introduction to the idea of the problem of knowledge, as well as a summary of the transition, at the turn of the 20th century, from the application of general philosophical systems to specific domains of science, to the piecemeal study of different domains to elucidate their unique theories of knowledge.

What counts as knowledge in this discipline?

What form does that knowledge have?

What are the tasks or functions of such knowledge?

What methods do people employ to generate that knowledge and re-deploy it?

Like others who address those and related questions, I assume that disciplines and subdisciplines of science needn't be led or evaluated by some most general system or theory of knowledge. Rather, in asking those questions, we assume that researchers in different fields construct often distinct and specific theories of knowledge to help them understand the abundance of phenomena in the universe.[2]

At their best, studies about the problem of knowledge serve several ends. They help us describe and learn about the epistemological workings of disciplines. They help us compare such workings across disciplines, enabling us to get a synoptic view of the sciences, though not necessarily a unified one. These studies also help us describe the history of knowledge and explain its evolution. In some cases, these studies can help practitioners in their fields to develop those fields.

In this dissertation, I describe some aspects of an epistemology now being developed at the interface of the fields of evolutionary biology and developmental biology. I also use my description for a further task: to indicate the prospects for the further development of that field.

[2] This position is consistent with, but doesn't entail, claims that the world is metaphysically un-unifiable (Dupré 1995). I am agnostic about such claims.

1.2- Evolution and Development

For at least the last 35 years, and with deeper historical roots, researchers have been trying to combine developmental biology with evolutionary biology (Laubichler and Maienschein 2007). This interface is nebulous. Some argue that novel theories are being developed at this interface. Others argue that a new theory of knowledge about evolution is being developed. Others argue that the interface is a nebula only in appearance, and that little of theoretical or epistemological interest is happening there.

For this dissertation, I focus on one region of that nebula, the interface between evolutionary genetics and developmental genetics. I focus on this region for several reasons. There are good arguments to think that this region is a likely area to develop a novel theory of knowledge for evolutionary phenomena. Furthermore, few have studied this region of the nebula, especially since about 2005. Finally, any insights about the aspects of a nascent theory of knowledge have the potential to influence the course of theorizing and research about evolution. I discuss each of those points below, and I use them to motivate the questions that drive this dissertation.

In 2003, Scott Gilbert argued that, in syntheses of evolutionary biology and developmental biology, the most important integrations were between models of evolutionary genetics (population genetics and quantitative genetics) and models of developmental genetics.

> But more importantly, not all parts of developmental biology and not all parts of evolutionary biology are involved in these new unions. What is happening, I believe, is a series of interactions occurring between **population genetic** models of evolution and **developmental genetic** models of evolution. Both of these models emphasize genes. But in one case, evolution depends upon the frequency of gene variants within a population. In the other case, evolution depends on variations of gene expression between populations (Gilbert 2003, 348; Gilbert's emphasis).

Later commentators, however, focused less and less on evolutionary genetic models as tools to be integrated with developmental genetic models.

Much of that shift in focus was due to Ron Amundson. Amundson argued that the two kinds of models are constructed and deployed in distinct investigative programs (Amundson 2005). For the adaptationist program, biologists use evolutionary genetic models to show how selection causes the evolution of species and of population-level phenomena (like gene frequencies). For the structuralist program, biologists use developmental mechanisms to show how developmental types like the quadruped limb or the avian feather evolved. Amundson noted that that models used to explain the evolution of population phenomena couldn't be used to explain the evolution of developmental types. The deep problem was that the logic of population concepts differed too much from the logic of developmental type concepts. He concluded that no evolutionary theory could synthesize the two programs.

Amundson's conclusions had a large impact. First, after Amundson most commentators increasingly ignored evolutionary genetics as it related to developmental biology. Instead, they discussed how researchers could use developmental mechanisms to study a wide range of macroevolutionary phenomena related to developmental types. They focused on issues of morphology, novelty, modularity, homology, and evolvability, all related to developmental types. Insofar as anyone discussed the role of evolutionary genetics, they mostly (re)asserted its importance without addressing Amundson's arguments (Hoekstra and Coyne 2007; Lynch 2007; Wray et al. 2014).

Second, though Amundson discussed models of developmental mechanisms generally, his conclusion also applied to gene regulatory network models (GRNs). GRN

models are a subclass of developmental genetic models, which are a subclass of developmental mechanism models. So if no developmental model could be integrated with any evolutionary genetic model, then more specifically, neither could any GRN model. Few commentators have discussed relations between GRN models and evolutionary genetics models.[3]

But while Amundson argued that the adaptationist and structuralist programs were incompatible, he tempered his conclusion.

> History has a marvelous way of making philosophical and methodological difficulties disappear (poof!) in the face of scientific success…If both evo-devo [representing the structuralist program and developmental mechanisms –SE] and population genetics continue to be successful, a way will somehow be found to see them as consistent (Amundson 2005, 257).

This dissertation begins with the conditional at the end of Amundson's passage. It's been ten years since Amundson published his book, and the structuralist program and the adaptationist programs continue to be successful, at least by metrics applicable to both. Each continues to draw students, develop research projects, secure funding, publish thousands of papers, address research questions according to the best criteria of their programs, and to export theories, data, and results throughout their disciplines.

How, then, do practicing biologists see the two programs as consistent? In the decade since Amundson's discussion, no one has systematically addressed this question. How might it even be addressed? Amundson's arguments show that the issues between the two programs are partly epistemological. As a result, the few people who have

[3] But see (Wagner 2000; Laubichler 2010; and Davidson 2011) for some such discussion, most of which echoes Amundson's more general conclusion.

addressed it have used purely philosophical methods, exploring the conceptual relations between the most abstract formulations of the programs.[4]

I pursue a different strategy here. I study projects in which researchers directly face Amundson's challenge of incompatible models. In such a project, a research team uses both evolutionary genetic models and developmental mechanism models, more specifically GRN models, to understand a phenomenon. Such projects provide perhaps the best source of data to see how researchers make those two kinds of models consistent with each other.

I also work to infer aspects of a common epistemology between those kinds of projects. Amundson's conditional indicates that any rapprochement between evolutionary genetics and developmental genetics will require a theory of knowledge that differs from the ones employed in their separate domains. I can't hope to detail all, or even many, aspects of a developing theory of knowledge common across those fields. But I can address a few.

To study the theory of knowledge shared among these projects, I focus on one question from the problem of knowledge. What is the task of knowledge created in these projects? I treat knowledge partly as the theories, models, and other empirically supported representations that researcher create, use, and export from their research projects. I treat the task of such representations as their epistemic functions. Such functions include to represent phenomena, predict it, or explain it. In Chapter 3, I develop this account of epistemic functions/aims. Given those explications, one question pursued

[4] On such study is (Craig 2015), who nicely elucidates the programs and defends Amundson's general conclusion.

in this dissertation is: For projects that use evolutionary genetic models and GRN models, what are the epistemic aims of those projects?

I focus on that question because most everyone whose publishes about the interface of developmental biology and evolutionary biology agrees it is important. Richard Burian, one of the earliest commentators on the nebula, stressed that integrations of programs and of models should be evaluated in how well they achieve their objectives (Burian 1986). Most who have proposed integrative fields, from evo-devo to the extended evolutionary synthesis to devo-evo and beyond, have fought about the most important objectives to pursue.

Besides the objectives that commentators think ought to be pursued, there are those actually pursued. Amundson convincingly argued that evolutionary genetic models can't be used to understand the evolution of developmental types. From that he concluded that the evolutionary genetic approach and the structuralist approach were incompatible. This dissertation shows that his conclusion was too broad. Some practicing biologists use GRN models, important tools in the structuralist approach, to understand the evolution of populations, phenomena from the evolutionary genetic approach. The two approaches are compatible, just not yet for the kinds of phenomena Amundson considered most worth explaining.

1.3- Driving Questions

A few questions drive this project. The overarching questions stems from Amundson's conditional:

1. How do biologists see evolutionary genetics as consistent with developmental genetics?

To address it, I closely study and compare two cases of research projects in which the teams who conduct the projects use evolutionary genetic models with GRN models to understand the evolution of complex phenomena. The first project was conducted by Norman Johnson, Adam Porter, and Alexander Tulchinky at the University of Massachusetts, Amherst, from 1998 to 2014. They studied simulated populations. The second project was conducted by Greg Wray, David Garfield, Daniel Runcie, and a few more colleagues at Duke University in Durham, North Carolina, between 2008 and 2013. They studied wild-caught purple sea urchins, which they bred in their laboratory.

Two further questions drive my analyses of those projects.

2. What models of evolutionary genetics do those projects use? What GRN models do they use?
3. What are the epistemic aims of the projects that use those models?

In Chapters 2 and 3, I develop methods to address those questions. In Chapter 4 I frame those questions against several prominent debates in philosophy of science about models, mechanisms, and explanation. I also preview the structure of the next two chapters. In Chapters 5 and 6, I provide detailed case descriptions to answer those questions.

With those descriptions, I have tools with which to systematically compare the cases to address, in Chapter 7, the final two driving questions.

4. How do the researcher teams relate the GRN models to the evolutionary genetic models in such a way that they are consistent with each other?
5. What are the prospects for similar projects, and for the field of developmental evolution more broadly?

For brief preview, my answers to the above questions are as follows. For question 2, the teams construct custom evolutionary genetic models that enable them to treat gene expression, rather than adult traits or behaviors, as the relevant phenotype. For question 3, the projects largely aim to describe the phenomena they study, while the Johnson team also aims to predict its phenomena, and the Wray team also aims to discover new phenomena. For question 4, the teams use a shared and implicit gene concept to link their GRN models to their evolutionary genetic models. Furthermore, the models link to each other, or are integrated, only with respect to complex phenomena implicit in the teams' research systems. The models are not integrated at the level of abstract theory.

For question 5, there are reasons to be optimistic, and there are reasons to be pessimistic, about the projects and their kin. But it's still too early to tell their long-term fate. As for the field of developmental evolution, I conclude that its prospects are most bright in comparison to evo-devo and to traditionally evolutionary theory insofar is its proponents conceptualize it as a field in which research use models of developmental mechanisms, regardless of the kind of model, to understand, regardless of the kind of understanding, evolutionary phenomena, regardless of the taxonomic index of such phenomena (macro or micro evolution).

The chapters that follow detail and justify those answers.

1.4- Meta-scientific Methodology

In addition to the above driving questions, issues about methodology also color this dissertation. Scientific research projects exist in space and time, and as such should be described with empirical methods. Much of this dissertation provides a strategy for studying research projects with methods commonly used in the social sciences, but rarely (if ever) applied to the problem of knowledge.

This dissertation is part of the tradition in which people use historical analysis, either of the history of ideas or of case studies of scientific research, to address the problem of knowledge for specific scientific disciplines. I conceptualize theories, models, research projects, methods, etc. as historical objects or processes, just as biologists similarly conceptualize organisms, populations, genotypes, phenotypes, developmental types, etc.[5] As historical objects, we can collect evidence about when, where, how, and by whom those objects were created, developed, redeployed, etc.

But I also go beyond the historiographic methods deployed in that tradition. While many in the 20th century ably used those methods to study the theories of knowledge in scientific disciplines, other noted flaws, sometimes trenchant, in those methods, and of the research that resulted from them.

To forestall some of those critiques, I deploy methods from the empirical social sciences, especially those aided by computers. While some have begun to apply those methods to study science; especially citation networks, research topics (Evans and Foster 2011), and collaborations (Leahey 2016; Wuchty et al. 2007); we're just now at the edge

[5] I'm most influenced by (Hull 1986) on this front, though I remain agnostic about his account of evolutionary epistemology.

of using those methods to describe theories of knowledge in scientific disciplines, that is, to address the problem of knowledge.

While the questions listed in the previous section drive the overall dissertation, the ones below influence the strategy by which I address the driving questions. My proposals aren't the ideal or only possible ones. They are just the ones I could manage. Ultimately, this dissertation is a step, albeit a small and cautious one, towards more robust empirical studies and characterizations of the problems of knowledge pursued in scientific disciplines. In terms of methodology, I'd be happy to find it soon replaced by stronger methods.

How do you identify and isolate a domain, area, or developing field of science? Theorists, historians, and philosophers often talk about scientific fields, but often provide only a fuzzy idea about the scope of those fields. That shortcoming is potentially acute for this dissertation, which talks about the nebular region between the fields of evolutionary biology and developmental biology, and focuses still further on the interface of evolutionary genetics and developmental genetics. This nascent field, if it can be called that, lacks unique textbooks, manuals, and organizing as reference points to help readers see it as a developing region of science. To show that it exists, and to indicate its scope, I represent the field as a set of research papers that share common themes. In Chapter 2, I show how to use computer-assisted bibliographic analysis to construct the population.

How do you conceptualize scientific research so that you can isolate cases for comparison? Many theorists and philosophers address this issue by focusing on theories and models abstracted from their real-world construction and use. On the other hand, many historians and sociologists focus on those products in their social contexts, focusing on large social units like paradigms, traditions, or programmes, or on smaller units like laboratories. The first approach alone can't help me address my driving question, and the socio-epistemic units in the second approach are too broad to provide much help when talking about nascent fields. In Chapter 2, I propose a new socio-epistemic unit that I call research projects. Such projects share a general local epistemology that I call rationales. For this dissertation, I use those concepts to help historicize actual scientific theories and models. As those who study science, we can compare research projects to each other, and we can compare rationales to each other.

How do you select cases for study and structure their descriptions without cherry-picking or biasing your cases? Those who study the epistemology of science often appeal to case studies to support or discredit claims about that epistemology. But many argue that the use of such cases exhibit bias by the person using them, either in how she describes or represents the cases, in how she selects the cases to begin with, or in the scope of her inferences from those cases. In this dissertation, I aim to forestall or at least to lessen those worries as they relate to my use of case studies. To do so, I construct an explicit population of actual research projects from the population of research papers. That population of projects provides the scope for inferencea I make from my cross-case analyses. Furthermore, I study and compare more than one case, a practice that enables

me to compare cases and suggest generalizations. Next, and to limit bias in case selection, I follow a protocol for partitioning the population of projects, and for selecting cases from the partitions such that they are independent of each other, and are relevant to the driving questions of this dissertation. I do all those tasks in Chapter 2. Finally, to limit bias in case descriptions, I use a common structure for those descriptions, which I introduce in Chapter 4. Finally, I approach the driving questions with no preconceived ideas about what their answers should be, nor with a particular axe to grind.

How do you explicate the notion of "the task of knowledge"? Among philosophers of science, there's a growing movement to study not just the structures of theories, models, and concepts, but also their functions (Brigandt 2012; Woodward 20014; Woody 2015; Lloyd 2016). Few have developed systems of such functions. In the context of research projects, I argue that theories and models function to address questions, ameliorate problems, meet epistemic aims, and satisfy values. Each of those general functions is a task of knowledge. I address the first two in (Elliott 2016). In Chapter 3 of this dissertation, I provide an account of the third. I propose a concept of epistemic aims/functions, and a taxonomy of such aims/functions. I also operationalize it for use with content analysis methods.

How do you collect data about the aspects of research projects and of rationales? For those who study science, one primary source of information is the scholarly articles published by scientists. Such articles aren't perfect sources of information, and they are systematically constructed to obscure details about the history of the research described

in them. But they are useful. The articles provide some information about the epistemic aims of research projects and the epistemic functions of theories, models, etc. They also carry much implicit information, and I argue some of that information is about epistemic aims/functions. Those who study science often just read the articles, summarize their models or results, and move on to other tasks. That method won't work to reveal the implicit information of the articles. To uncover implicit information from those articles, and to tie my inferences about epistemic aims/functions to bits of text from the articles, I use formal content analysis (Krippendorff 2013). In Chapter 3, I develop the framework for using content analysis on scientific articles, and I tabulate the results of those analyses in the case descriptions of Chapters 5 and 6. Content analysis methods provide a tool by which others can check the quality of my claims about the kinds of epistemic aims pursued by the two research teams I study.

How do you ensure the reliability of your methods, and the reproducibility of your results? Even though I use methods like corpus analysis, content analysis, and cross case analysis, the mere use of those methods doesn't ensure the strength of the inferences I draw from them. At the very least, I should also show that the methods are reliable, that if I use them repeatedly, they return the same data. Similarly, I should show that my inferences are replicable, that other can study the same phenomena and infer the same conclusions. To ensure that my corpus analysis methods are reliable and yield replicable results, I operationalize my concept of research projects so that it applies to the population of research papers, I create and use protocols to infer the presence of research projects, and I use computer programs to conduct nearly every step. Most of those topics

are discussed in Chapter 2. To ensure that my content analysis methods are reliable and return replicable results, I operationalize my taxonomy of epistemic aims/functions, use an explicit content analysis design, create and use explicit data collection protocols, and use computer programs to collect the data. Most of those topics are discussed in Chapter 3. To ensure that my case analysis methods are reliable and return replicable results, I create and use a protocol to construct case descriptions, and I construct the case descriptions not as narratives, but as reports in relation to the dissertation's driving questions. Most of those topics are discussed in Chapter 4. Finally, I include all of the protocols used for this dissertation in the Appendices.

CHAPTER 2

HOW TO CONCEPTUALIZE AND STUDY SCIENTIFIC RESEARCH PROJECTS

2.1- Introduction

This chapter proposes a concept of research project, an account of project rationales, and it provides protocols to delimit populations of research projects and to select instances from those populations for case study analyses. It ends with an example, drawn from research projects in evolutionary biology, of how to use those protocols. Those tools don't exhaust the tools needed to study research projects. But the tools described here enable those who study scientific projects with cases to specify both populations of projects related by a common feature, and the projects drawn from those populations as cases for focused study.[1]

Researchers have proposed many kinds of socio-epistemic units with which to study the products of science. Such units include paradigms (Kuhn 1962), research programmes (Lakatos 1970), fields (Darden and Maull 1977), research traditions (Laudan 1977), laboratories (Latour and Woolgar 1978, Latour 1987), scientific practices (Kitcher 1984), styles of thought or research (Fleck 1935; Hacking 1992; Harwood 1993; Crombie 1994), disciplines (Lenoir 1997; Becher and Trowler 2001), movements (Frickel and Gross 2005), and repertoires (Leonelli and Ankeny 2015), to list just a few.

Such units have important uses. Researchers proposed the above kinds of units to help them study and explain science as scientists conduct it, and not merely as the sum of abstract theories, hypotheses, or other lexical phenomena. To that end, researchers have

[1] Thanks to Jane Maienschein for helping me to see the importance of this topic.

used the units above to reveal much about actual science. And while some of the units capture similar phenomena, they also foreground different aspects of those phenomena, yielding surprisingly little theoretical competition between the units themselves.

But the units indicated above also face serious issues. Those units are rarely specified with concepts that aren't ambiguous or vague. Their authors rarely show how to precisely isolate an instance of one kind of unit from another instance of that kind, nor do they provide protocols to study those instances, nor do they explain how to relate the above units to each other. As a result, not only do studies that employ such concepts face issues of construct and external validity and replicability, they also provide other researchers tools that are sketchy at best.

In this chapter, I propose a concept of research project. That concept enables research into science as it is actually conducted, just as the above units do. But the concept of research project is comparatively precise and operationalizable, so that we can test the validity and reproducibility of studies that employ it. I don't argue that the concept replaces or is superior to the concepts for the units listed above. On the contrary, I think that we can use the concept of research project to make precise the concepts listed above, though I don't pursue that project here.

I also provide a protocol for studying research projects. It enables me to construct reliably a population of research projects, to distinguish among research projects, and to check that a supposed research project is in fact a research project. I also indicate how to select individual projects from the population of projects for specific case studies.

Strategy for the Rest of the Chapter

To accomplish the above goals, I follow the following strategy in the rest of this chapter. In the next section, I contextualize this paper in debates about how to use case studies to investigate science. I distill two problems that confront those who wish use case studies to investigate science, and I use them to motivate the construction of the tools provided in later sections. There are five primary tools: a concept of research project, a map of epistemic relations for research projects, an operationalization of the concept, and two protocols to delimit projects. I also provide an example of how to use the concept and protocols in relation to the driving questions of this dissertation.

Section three details a concept of research project, a map of epistemic relations that applies to those projects, and it provides one way to operationalize that concept so as to establish the existence of such projects, and populations of similar projects, from collections of research articles.

Section four describes a general protocol to collect that evidence, and it includes tests for reliability of methods, external validity of results, and construct validity of the operationalized concept. Section four also and it provides a strategy for selecting a small sample of projects for case study analysis. Section four concludes with an example of how to use the protocols to identify a population of similar research projects. The example is about research projects at the interface of evolutionary biology and developmental biology. This chapter concludes in section five, in which I review some limitations of the tools.

2.2- A Context of Case Studies

Background

The tools I describe in later sections can help ameliorate issues about how philosophers, historians, and social scientists use case study analyses to investigate science. That said, the tools aren't specific to case study analyses, and they could be used profitably in other research frameworks. But case studies provide a particularly useful context in which to see the potency of the tools.

Philosophers, historians, and social scientists of science use case studies ubiquitously. If we open a journal from any of the common science study disciplines, we see many articles that claim to employ case study methods, or to describe cases of some more general phenomena. Furthermore, in their instructions to authors, the editors of those journals often indicate how much case study content they seek in submitted articles.[2]

A quick analysis reveals the ubiquity of case studies in those journals. I did a quick analysis of case studies in the Web of Science database of scholarly articles. I focused on the ten-year period between and including the years 2005 and 2014, and I collected all of the articles from the following journals: *British Journal for the History of Science, British Journal for the Philosophy of Science, Isis, Perspectives on Science, Philosophy of Science, Social Studies of Science*, and *Studies in History and Philosophy of Science*. Excluding all but research articles and conference proceedings, there were 2,372 articles in the set. Of those, 439 articles, or 18.5% of the total, explicitly invoked or used the terminology of cases in their titles or abstracts. Of those, 95 articles, or 4.0% of

[2] For a good example, look to this official blog post by Steven French, editor of the *British Journal for the Philosophy of Science*. http://thebjps.typepad.com/my-blog/2015/01/deskrejectionfrench.html

the total, explicitly discussed or mentioned in their titles or abstracts the use of case study methods.[3]

Those who study science use case studies for many functions. Among other functions, researchers use cases as heuristics or examples by which to elucidate some abstract claim about science, they use them to provide some evidential support for or against some claim about science, and they use them as exemplars with which to engineer new concepts and to show others how to deploy those concepts (Currie 2015). In the majority of science studies articles, researchers construct and deploy cases for at least one of those myriad functions.

But those who study science, especially philosophers, have developed several discussions about case studies as a style of science. In one such discussion, they study how researchers use case studies in general (Forrester 1996; Ruzzene 2011; Morgan 2012; Morgan 2014). In a second discussion, they study how other researchers use case studies in specific fields, such as psychoanalysis (Grunbaum 1988), applied ecology (Schrader-Frechette and Mccoy 1994), political science (Crasnow 2011; Crasnow 2012), developmental biology (Ankeny 2012), or medicine (Ankeny 2014).

In the third discussion, researchers debate about the appropriate role of case studies as evidence for or against more general claims about science. Some are pessimistic that case studies can provide such evidence (Brooke 1981; Meehl 1992; Faust and Meehl 1992; Pitt 2001; Faust and Meehl 2002), while others are optimistic that case

[3] If I'd included more journals, including the dozens of specialty journals, we'd have seen not only an increase in the absolute number of articles that discuss or invoke cases and case studies, but I suspect the percentages would be higher as well.

studies can provide such evidence, at least in some situations (Donovan, Laudan, and Laudan 1988; Burian 2001; Chang 2012; Currie 2015; Kinzel 2015).

I situate this chapter in the third discussion.[4] I agree with those who say that we can use case studies to provide evidence that confirms or disconfirms somewhat general claims about science, but I acknowledge that the pessimists have raised important issues. If we want to produce better case studies, we need to continue to address those issues.

Problems and Success Conditions

The above literature notes many problems with case studies as they're commonly used in studies of science. Here, I focus on two such issues:

1) Researchers often bias their case studies to support their positions.
2) Case studies provide poor empirical grounds from which to generalize about science.

The first issue is often called the problem of cherry-picking, according to which researchers don't randomly select cases for study, but they instead preselect cases that will confirm their pet theories. The second issue is that even a well constructed case provides, at best, a single data-point with which to confirm or disconfirm a general claim about science, so case studies can only disconfirm maximally general claims about science, and they otherwise provide little evidence for or against other claims. As such,

[4] While I situate this chapter in the third discussion, the overall dissertation provides an instance of what Currie calls the ahistorical project of using cases as instruction manuals by which to show others how to use newly engineered concepts (Currie 2015). For this dissertation, the newly engineered concepts are those of research projects and of epistemic goals. By showing how to investigate research projects and epistemic goals in my three cases, I attempt to show others how to redeploy my concepts in other situations. For more on instructional uses of case studies, see especially (Burian 2001) and (Chang 2012).

we cannot use them to decide between competing claims about science, especially philosophical claims like the realism or anti-realism about science.

Many, especially since 2000, have addressed those issues in fruitful ways. For instance, Kinzel notes that case studies are no different than any other empirical style, and that they are prone to issues of, or related to, theory ladenness (Kinzel 2015). She indicates that such theory ladenness often gives the appearance of cherry-picked cases, but that such appearances can be wrong. I agree, but Kinzel's argument doesn't rule out that many case studies have been, are, or could be, cherry-picked. And I don't think her argument will convince the pessimists who raise issue (1), most of whom were well aware of the empirical character of case studies and of issues of theory ladenness. I take Kinzel's analysis a step further and ask: from the set of cases that might have been cherry-picked, how can we distinguish those cases that weren't cherry-picked?

As for the second issue, Burian stressed that case studies provide evidence for or against claims about science that have a limited and specific scope, especially for what I call local rationales, but others have called local epistemologies (Burian 2001). More recently, Currie argued that case studies provide evidence for claims more general than the ones Burian countenanced, provided that those employing the case studies established a patchy-unity among the set of cases covered by the general claim (Currie 2015). I agree with Burian and Currie, but again I think that their arguments won't convince the pessimists. We face two further questions: how do we precisely specify scope and delimit local epistemologies, and how do we establish patchy-unity?

To answer those questions, I pursue a different strategy. I don't think that arguments based upon reflection alone will answer the above questions or satisfy case

study pessimists. Nor will such arguments alone yield tools to strengthen our methods and strategies of using case studies to investigate science or to provide evidence for or against claim about science. In addition to such arguments, I propose that we construct, make explicit, evaluate, share, and improve strategies and methods for case study research. While arguing for the evidential merits of case studies is helpful, what is more fruitful, and ultimately more convincing, will be research designed to explicitly address and mitigate the issues raised by pessimists.

Later, I develop a set of tools that help ameliorate, but not completely solve, the two issues listed above. The tools include a concept of research project and two protocols for investigating them. The two issues above become success conditions for the total suite of tools provided. As success conditions, I rephrase the above issues as:

1. Show how to delimit a population of cases, provide criteria to select cases—justified according to the research question, and select cases.
2. Make explicit how to evaluate inferences from evidence with tests of reliability, construct validity, and validity.

Preliminaries

Before I describe the tools, I highlight some issues about case studies that I don't address in depth in this chapter, but that nonetheless color the rest of my discussion.

First, researchers use the terms 'case' and 'case study' with many different meanings, and because those terms are imprecise, authors often talk past each other when they talk about cases and case studies. Some take cases to be idealized instances of general phenomena, while others take 'case' as synonymous with 'case report'. Some

take case studies to be about only single cases, while others take case studies to be about multiple cases. We need a conceptual framework with which to talk about the many flavors of research involving cases, a project for future research. For this paper, I treat a case as an instance of some phenomena of specified generality, a case report as a description of that instance, and a case study as being about more than one case and involving multiple case reports.

A second and related issue is about case reports. Almost all of the authors cited so far assume that case reports have the structure of historical narratives or stories. While many reports do have that structure, and while that structure can serve the ends of a given case study project, it isn't the only valuable structure (Eisenhardt 1991). This assumption is especially prevalent among philosophers and historians, who have a long running discussion about the appropriate relations between historical and philosophical studies of science. For instance, insofar as case studies aim to give explanations of phenomena, and insofar as historical narratives provide explanations, we can fruitfully use historical narratives of cases to explain phenomena. But while I agree that all data used to construct case reports is historical, I hold that not all reports must provide historical narratives (Creath 2010; Eisenhardt 1991).

A third and corollary issue is about the aims of case study analyses and their relations to evidence. Some, especially from the social sciences more broadly, hold that we only use case studies to test explicitly causal claims (George and Bennett 2005; Gerring 2006). I disagree. While we can use case studies to test causal claims, we can also use them as evidence for or against descriptive claims. In this dissertation, and following the suggestions of (Burian 2001), I use case studies to describe the local

rationales of research teams, and to compare those rationales across two research teams. I do so to establish trends, not to causally explain them (Gerring 2012). More generally, I hold that we can use case studies to further any number of epistemic goals, from discovering phenomena or describing them, to explaining or predicting them.

Finally, there is an underlying theme to the issues raised by those who are pessimists about the use of case studies as tools to provide support for or against claims about science. The issue is that, insofar as case study research is a style of research, those who claim to use the style often fail to employ the standards or tools of that style. The pessimists cited above raise few issues about the uses and limitations of case studies that haven't been raised by case study theorists in other disciplines.

In the social sciences more broadly, researchers have struggled for decades with issues of how to delimit populations of cases and individual cases, how to select cases, how to theorize with cases, and how much evidence cases provide. While those who study science could contribute much to those methodological debates, they could also learn much from the results of those debates. Social scientists have developed many explicit strategies and tools by which to design, conduct, and use case studies to investigate socio-epistemic units and organizations (Eisenhardt 1989; Eisenhardt 1991; King, Keohane, and Verba 1994; Yin 2003; Gerring 2004; George and Bennett 2005; Gerring 2006; Flyvbjerg 2007; Seawright and Gerring 2008). Insofar as we study socio-epistemic units and organizations in science, we can borrow and improve upon many of those strategies and tools.

In short, if we want to use case studies to provide evidence for or against claims about science, the best way to quiet the pessimists is to design our research, collect data,

make our inferences explicit, and evaluate our methods and results according to standard cannons such as reliability, reproducibility, and the various flavors of validity. The tools below enable researchers to design and evaluate the initial stages of case study research, but not the later stages in which they compare cases and infer causes, trends, predictions, etc.

2.3- A Concept of Research Project

In this section, I develop a concept of research project. I begin with a rough idea of the concept, then I explicate that rough idea into a categorical concept, I articulate that a related map of epistemic relations, and finally I provide one possible operationalization of that concept. The concept will help us begin to answer the question: That case is a case of what exactly? The answer will partly be: That case is a case of a research project.

The Rough Idea

Compared to many of the socio-epistemic units of science listed in the introduction, real world researchers more often use the phrase "research project" to refer to their own work. They litter their papers and websites with that phrase as a way to bound their work and to relate it to the work of others. When they apply for grants, the grants are for projects. They use the phrase at conferences to introduce themselves and their work in their presentations or when they meet people. They also train their students to identify, pursue, and evaluate their own research projects. In short, researchers ubiquitously use the concept of research project, but it has received little if any attention,

even from those who study science and who might profitably use the concept to help ground their own studies.[5]

From that ubiquitous use, I highlight some implicit aspects of research projects. A research project has at least some spatio-temporal aspects and some abstract epistemological aspects. I discuss each kind of aspect in turn.

Spatio-temporal Aspects

Like many of the units listed in the introduction, research projects exist in space and time. For that reason, we can collect evidence about them and we can test claims about them. Research projects have beginnings and ends, though those markers may be fuzzy. Research projects can have lulls in their temporal persistence, with periods of activity and output interpolated by periods of little activity or stasis. Like laboratories, research projects exist in space, which is generally the space of the researchers pursuing the project, and over time the location of the project can change. For instance, a particle physicist might have a project elucidating the quantum electrodynamic background for neutrino emissions, which she has conducted at the same university for several years. If she leaves for a higher paying job in a government agency in a different city, she might take her project with her.

Research projects are conducted by people, either individuals or teams. If the project persists for a good length of time, the people who pursue the project can change, just as the members of a band can change over time. For instance, a primary researcher

[5] There are some heuristic cookbooks that tutor new graduate students on how to conduct research projects in specific fields. But those cookbooks are often collections of best-practices, not sustained developments of the concept of research project.

who has a lab in a university might have a project that persists for about five years. In that five years, the primary researcher might have several different post docs, graduate students, and undergraduate students cycle through the project. Yet through those changes, the project persists.

Compared to the kinds of units listed in the introduction, research projects refer to the smallest socio-epistemic units. From the list in the introduction, laboratories (Latour and Woolgar 1978) and repertoires (Leonelli and Ankeny 2015) are the smallest units. Research projects are smaller than laboratories, which are also centered on people and research teams. But in a laboratory, the researchers can, and often do, pursue many research projects. And an individual researcher, who collaborates with no one, also often pursues several projects. Research projects are also smaller than repertoires, which by definition take successful research projects as exemplars from which to develop a toolkit of practices and strategies.

The bigger a project gets, with more people and more tools and materials, the more it can grade into units like repertoires and programmes. Items labeled as projects might fruitfully be described as projects and as programmes. The Human Genome Project is an example.

In a research project, a research team studies phenomena, which themselves exist in space and time.[6]

Finally, a team creates many kinds of outputs as part of a research project. Those outputs can be data, theories, models, computer programs, grant proposals, grant reports,

[6] In this chapter, I assume that research projects are empirical endeavors. I think that my terminology could be loosened to create a concept that also refers to projects in purely analytical fields, but I don't pursue that task here.

public outreach pieces, protocols, articles in journals and books, books, etc. A team receives feedback about those products at least in terms of funding, citations, awards, and by its ability to grow its project into repertoires, programmes, and paradigms. When we study research projects, those outputs become sources of evidence by which to infer information and test claims about those projects and their kin.

Abstract Aspects

Like many of the units listed in the introduction, research projects have abstract epistemological aspects. We can study the logical relations among those aspects. In a given project, the research team uses theories, hypotheses, models, and similar items collected under the label of scientific products. The team also collects data and infers new scientific products from those data. In doing those tasks, the team invokes and contributes to a special vocabulary and to a suite of methods.

Researchers focus their research projects on specific issues within a broader topic. While the phenomena studied by a project exist in space and time, the topic that describes and contextualizes those phenomena is abstract. Topics are aspects of fields, and they are often about general phenomena. For instance, in the field of evolutionary biology, one topic might be the evolution of pupfish, and a specific project might study the evolution of pupfish in a half dozen ponds near the border between California and Nevada.

In addition to focusing a topic, a team pursues a project to address questions, achieve epistemic aims, and to ameliorate problems. Fields or paradigms or programmes all have collections of questions, aims, and problems. But for a research project, the team specifies from those broad collections the particular problems or set of problems they

pursue. The same is true of particular epistemic goals and questions. Ultimately, the problems, questions, and goals pursued in a research project are much more specific compared to the sets of problems, questions, and goals that partly define units like paradigms, fields, etc.

Furthermore, research projects exhibit a network of epistemic relations. For a given research project, the specific goals, questions, and problems pursued logically connect with data and scientific products to form a project rationale, or a logic of inquiry, according to which researchers can evaluate the quality of a research project. Often for fields, paradigms, etc., those logical connections cannot be delimited because the sets of problems, questions, and aims are so large. Or, if connections are made, they are too broad to be of much use. Not so in research projects, which regularly face evaluation of their quality in grant proposals, article submissions, institutional reviews, and the marketplace of ideas. Therefore, research teams often take pains to explicate the logical connections among their problem statements, questions, epistemic goals, scientific products, and data.

A Concept of Research Project

From the discussion in the previous section, I distil a concept of research project. The concept enables us to operationalize it and thus to collect evidence about research projects "in the wild". But the concept provided is a first attempt, and as it is an empirical concept, we might later revise it in light of new evidence. Of the many kinds of concepts, the one provided is a categorical concept, which enables us to class phenomena as research projects or not.

Research Project

Some x is a research project only if:

1. A research team of at least one person conducts x
2. x has a beginning and persists for a period of time (maybe to a termination)
3. one set of outputs of x includes research publications, grant proposals, grant reports, institutional reports, public outreach pieces, all of which at least some of the team members create
4. another set of outputs of x includes data, computer programs, and scientific products (hypotheses, models, theories, etc.), all of which at least some members of the team create
5. as part of x, the team pursues a specific and relatively precise set of problems, questions, and epistemic goals
6. as part of x, the team studies a specific and relatively precise set of phenomena
7. x is partly constituted by those actions of the team members in which they pursue the problems, questions, and epistemic goals from (5); study the phenomena from (6), and produce the outputs in (3) and (4).
8. x has an internal logic that fits together the pieces from (4–7) in a rationale and is reported in the items in (3).

I note some features of the concept. First, it provides a set of necessary, but not jointly sufficient, conditions for an item to be a research project. That feature indicates

not only that might we revise the individual conditions in light of experience, but that we might also profitably add to them. For instance, some of the conditions describe outputs, but none describe inputs. I think that research projects have inputs, but I can't yet think of a way to discuss them without biasing the condition towards projects conducted in laboratories. The concept needs conditions about inputs, a project I leave for the future. Similarly, projects might incorporate the a-epistemic aims of team members. Such aims could be for esteem, money, career advancement, etc. The concept above is silent on those issues, and on topics about the complex social reinforcements between scientists, thus limiting its ability to more fully describe projects.

Second, condition (2) mentions the beginnings of research project, but it hedges on terminations. I hedge because while many projects have demonstrably ended, while others are active and ongoing. Someone studying a research project will have to determine how she wants to bound later temporal parts of the project.

Third, the outputs listed in condition (3) often describe many of the aspects of research projects. They thus provide primary sources of evidence for research projects. However, they don't provide the only sources, and those studying a given project can collect evidence about those projects from other sources, such as by interviews of the team members or by observations of the team in action.

Fourth, I treat the outputs listed in condition (4) as exportable tools. Not only does a research team create and employ those tools to address the project's problems, questions, and epistemic aims, but other teams can get and redeploy those tools for their own research projects. As they are tools created by teams with histories, those tools also have histories, and we can study those histories.

Fifth, condition (5) distinguishes research problems from questions, and both from epistemic goals. That condition assumes that we can conceptualize those things somewhat precisely and as distinct from each other, an assumption that some may disagree with. See the Chapter 3 and (Elliott 2016) for reasons to adopt that assumption.

Sixth, conditions (5) and (6) invoke a phrase about specificity and relative precision. Such specificity and precision of research projects is relative to laboratories, repertoires, programmes, and other socio-epistemic units listed in the introduction.

Finally, condition (8) mentions a rationale for a research project. for those who work to describe a given research project, they must describe that rationale as best as they can. To help in that task, I provide the following map of epistemic relations.

Map of Epistemic Relations

Burian and Currie left us with two questions: how do we delimit local epistemologies, and how do we establish patchy-unity? To begin to address those questions, I relate the concept of research project to a map of epistemic relations within research projects. Here, I posit it hypothetically, but I don't show how I came to construct it or how it relates to similar maps. I leave that task for a future project.

The Map

The map in Figure 2.1 is at first glance overwhelming. There are nine nodes and twenty-six edges between them. I've already discussed most of the nodes, most of which are part of the explicit concept of research project. From that concept, Research Teams are part of condition (1), Products and Data are part of conditions (4) and (7), Problem(s)

and Question(s) and Epistemic Aims are parts of conditions (5) and (7), and Data Collection Methods and Inference from Data Methods are reified parts of condition (7). The only new node is that of Values, which we might include as part of condition (5), but which I keep separate.[7]

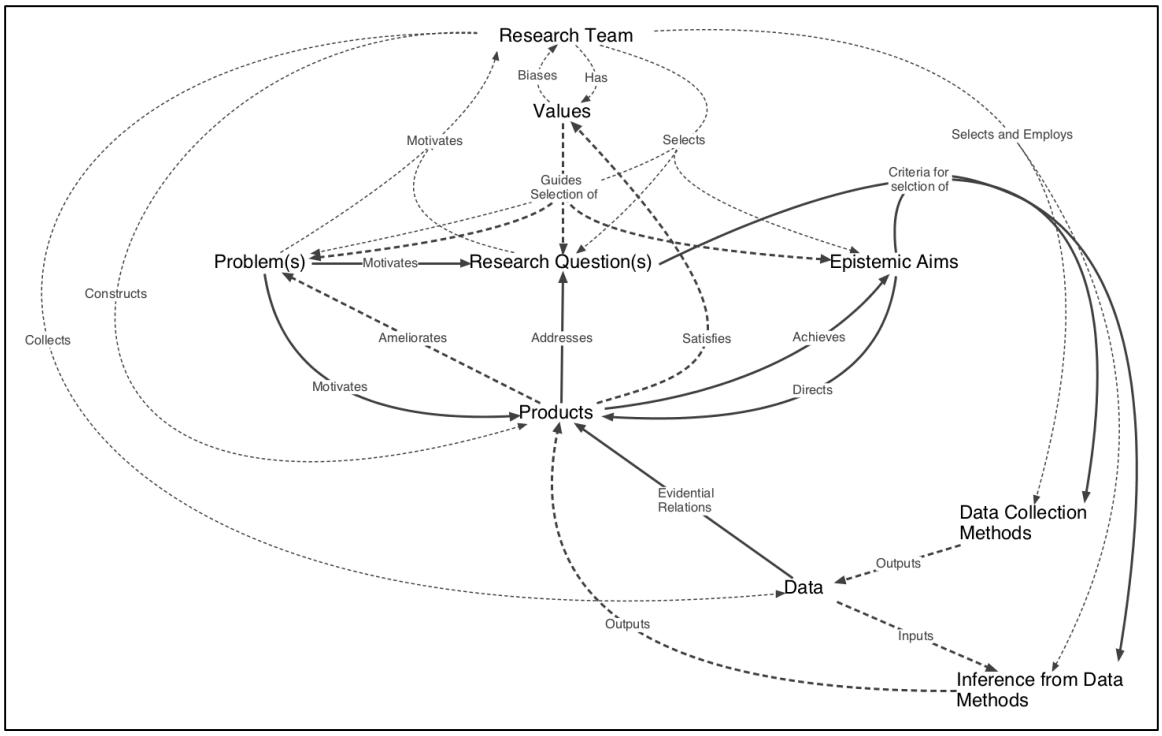

Fig. 2.1. Map of Epistemic Relations for Research Projects. The model represents nine nodes and twenty-six edges, most of which comprise parts of the concept of research project. The lightly dashed lines represent causal or empirical relations between nodes. Solid lines represent analytical or logical relations between nodes. Heavily dashed lines represent relations that can be empirical or logical, depending on the aims and questions of the research *about* the research project.

There are three kinds of edges, those with lightly dashed lines, as between Research Team and Data; those with solid lines, as between Data and Products; and those with heavily dashed lines, as between Products and Problems. The eleven dashed lines

[7] Many philosophers of science conflate epistemic aims with values. I think they are importantly different, so I distinguish them. But I've yet to develop a condition for the concept of research project that adequately captures my thoughts about values. I leave the development of that condition to later work.

represent connections between nodes that we most easily study with empirical means. They are the causal relations, about which we can collect empirical evidence when we investigate research projects as social scientists or as historians. Furthermore, the lightly dashed lines emanating from the Research Team node represent the actions described in condition (7) of the above concept. When we empirically investigate a research project, we often must first describe and specify the nodes and the causal relations between the nodes.

The seven solid lines represent connections between nodes that we most easily study with analytical means. They are logical relations, which we analyze and tinker with when we investigate research projects as theoreticians or as philosophers. Together with their nodes, they represent the internal logic of the project discussed in condition (8). The eight heavily dashed lines represent relations that we might study with empirical or with analytical means. Our choice of means depends on the questions, aims, etc. that guide our own projects when we investigate the projects of others. Social scientists and philosophers might both study how a team's products satisfy the team's values, and if the social scientists and the philosophers differ in their conclusions, its likely because their own research questions stress different aspects of relations of satisfying.

Burian and Currie

Burian proposed that we could use case studies to study the local epistemologies of research groups, which we then compare to other research groups, and from those comparisons we abstract ever more general epistemologies (Burian 2001). Burian has pursued this task in his own studies. But aside from his and similar examples, we have

few tools with which to construct local epistemologies such that we can systematically compare those epistemologies across groups.

I argue that the map above provides a theory of local epistemologies, which I call rationales, for research projects. A rationale is an instance of the general map that is specific to a given research project. When we isolate a research project, we bound something for which we can make systematic comparisons to similarly bounded things. When we describe a project, we specify the nodes and edges between the nodes that are represented in the general model. The specification of those nodes and edges is almost always an empirical task.

The map also provides a pattern that enables not only systematic description of single research projects, but also systematic comparisons across research projects. It enables us to precisely show that, for instance, distinct research teams develop different products to achieve the same epistemic aims, or that they develop the same products to achieve different aims, etc. That precision, I propose, enables those who study science not only to better understand science and its evolution, but also to better identify conceptual, empirical, and social roadblocks in the evolution of research traditions. Furthermore, when we compare projects and reveal similarities between them in their nodes or edges, we establish the patchy unity at which Currie and others have gestured.

Notes on the Model

I present the model as a descriptive tool. When we study research projects, we can use it to guide us as we study the local epistemologies of those projects. But the model is more than a heuristic. Though I pose it here only hypothetically, it is testable, and given

data, we may revise it. Furthermore, the model enables certain kinds of counterfactual studies into research projects. For a given project, were the specific contents of a node or edge to change, the contents of the nodes they influence are likely to change as well. That counterfactual aspect enables causal testing of the dynamics of research projects. It also enables researchers to specify the meanings of concepts according to the roles those concepts play within a project and across compared projects.[8]

An Operationalization of the Concept of Research Project

We might operationalize the concept specified above in many fruitful ways. For the process of operationalization, we start with a somewhat precise concept like the one detailed above. Then we specify ways by which to collect evidence about the objects to which that concept refers according to the conditions of that concept, and we collect those ways in a list. I treat such a list as the operationalized concept. We can develop different operationalizations of the same concept based at least on the conditions of the concept that we stress, on the kind of evidence we have available, and on the questions we ask of the objects referred to by the concept. Thus, the relations between a starting concept and its operationalizations is one-many, and no single operationalization provides the true or best definition of the starting concept itself. While below I provide one operationalization of the concept of research project, others could fruitfully develop other operationalizations.

[8] Thanks to Rick Creath for encouraging me to think about counterfactual comparisons of research projects as potential tools for conceptual role semantics of scientific terms. On this topic, I'm especially influenced by Sellars (1953) and now by Creath (1994).

The operationalization that I provide in this chapter has the following bases. I ask of research projects: how can we identify a research project and isolate it from other research projects that are similar to it either because they share a general topic or because they are pursued by the same people? To help address that question, I focus on the first three conditions from the concept of research project. Finally, I focus especially on research publications as the outputs of projects that provide evidence for the existence and spatio-temporal bounds of research projects.

I represent the operationalization of the concept of research project in Table 2.1. For each condition of the operationalized concept, the table specifies a name, the relevant conditions from the concept of research project, an informal characterization of the condition, and a more formal hypothetical characterization of the condition.

To use the table, I start with a collection of research publications. Some of those publications may cluster together as outputs from distinct research projects, but we don't yet know which will do so. To help infer the existence of research projects, it will help if the papers in the collection aren't totally random, but are about a common but general topic. For the remainder of this chapter, I treat topics as collections of publications that share themes. In the protocols in the next section, I show how to generate such topics.

Given a topic, I use the conditions in Table 2.1, especially condition ORP-4, to generate networks of papers. Each network indicates the existence of a research project. I use condition ORP-3 to identify the members of the team that conducted the project. And I use condition ORP-5 to bound the project in time. We need use no other conditions to infer networks that represent research projects. But the networks are at best first approximations, and many of the networks will falsely indicate that networks of papers

represent distinct research projects. Often, those networks will be too inclusive, representing several research projects as one.

Table 2.2 is a little different, and the conditions in it help us overcome the issue described in the previous paragraph. We can interpret the conditions in Table 2.2 in at least two ways. In the first, the conditions in Table 2.2 just add to those in Table 2.1, and they further specify the conditions in the concept of research project. In the second, the conditions in Table 2.2 provide means to check the validity of our inferences from the conditions in Table 2.1. That is, we use the conditions in Table 2.2 to check that the networks inferred via the conditions in Table 2.1 in fact represent distinct research projects. I prefer the second interpretation of Table 2.2, as Table 2.1 provides minimal conditions to infer the existence of projects, but nothing is wrong with the first interpretation.

Each table represents its conditions with a natural language or warrant form and with a hypothetical form. I use the natural warrant form to understand the gist of the condition, and I use the hypothetical form to make explicit our inferences from data. The hypothetical forms are each fallible, as the concept of research project is an empirical concept.

TABLE 2.1

OPERATIONALIZATION OF *RESEARCH PROJECT*

Operational Condition	Relevant condition(s) from the Concept of Research Project	Natural Warrant Form	Hypothetical Form	Notes
ORP-1	3	The papers published by a team that relate to a general topic provide evidence that the team had a research project related to that topic.	If a team has a research project related to a topic, then if we check the team's outputs, we're likely to find papers about that topic.	This condition makes explicit an otherwise implicit assumption not specified as part of the concept of research project.
ORP-2	3	The more papers a team publishes from its research project, the more information it provides about that project.	If team B provides more information about its project than does team C about its project, then if we check the outputs of the two teams, we're likely to find more papers by B than by C.	This condition extends ORP-1.
ORP-3	1	The authors of a paper indicate the members of a research team.	If a team includes people, and if it published papers, then if when we check the authors of those papers, the list of authors will indicate the team members.	This condition enables us to identify the members of a research team by a nearly universal convention about authorship.

TABLE 2.1—CONTINUED

OPERATIONALIZATION OF *RESEARCH PROJECT*

Operational Condition	Relevant Condition(s) from the Concept of Research Project	Natural Warrant Form	Hypothetical Form	Notes
ORP-4	1,3	A common author of two papers related to a topic indicates, but doesn't guarantee, that the papers are products of a single project.	For any two papers related to a topic, if the papers are products of single project, then if we compare their authors, the papers will share at least one author.	This condition is extremely fallible, and inferences from it should be validated against further conditions specified below. This condition enables us to join papers in networks of common authors.
ORP-5	2	Publication dates of papers from a common research project indicate the temporal period of that project.	If a research project persists for a period of time, then if we check the publication dates of those papers, those dates will indicate the span of that period.	This condition yields approximate dates, as any project must exist before it can produce its first publication. Inferences from this condition should be revised according to evidence from sources other than the research publication of that project.

TABLE 2.2

VALIDITY CONDITIONS FOR *OPERATIONALIZED RESEARCH PROJECT*

Operational Condition	**Relevant Condition(s) from the Concept of Research Project and from the Operationalized Concept of Research Project (ORP)**	**Natural Warrant Form**	**Hypothetical Form**	**Notes**
V-ORP-6	1,3, ORP-4	The more authors two papers from the same topic share, the more likely the two papers are from the same research project.	If two papers from a topic are outputs of the same project, then if we compare them to two random papers from that topic, the first two papers will share more authors than will the second two papers.	
V-ORP-7	1,3, ORP-4	Shared first or last authors across two papers indicates, but doesn't guarantee, that the two papers are from the same project.	If two papers are from the same project, than if we check the first and last authors of the papers, those authors will be the same across the papers.	This condition enables us to distinguish projects within a network of co-authored papers and is based standard authorship conventions among scientists
V-ORP-8	1,3, ORP-4	In a network of papers connected by common authors, two papers that share no authors are likely, but not necessarily, from different projects	If two papers in a network of co-authored papers are from distinct projects, then if we check the authors of those papers, there will be no shared authors.	This condition increases in fallibility the longer a project persists and evolves.

TABLE 2.2—CONTINUED

VALIDITY CONDITIONS FOR *OPERATIONALIZED RESEARCH PROJECT*

Operational Condition	**Relevant Condition(s) from the Concept of Research Project and from the Operationalized Concept of Research Project (ORP)**	**Natural Warrant Form**	**Hypothetical Form**	**Notes**
V-ORP-9	2, ORP-5	In a network of papers connected by common authors, the closer in time two papers are published, the more likely they are from the same project.	If two papers about a topic that share an author are from the same project, then if we compare their dates of publication to the dates of publication for any two random papers in the topic, the dates of the first two will be closer to each other than will the dates of the latter two.	This condition enables us to distinguish projects within a network of co-authored papers.
V-ORP-10	4,5	The more two papers discuss the same instances of research problems, questions, epistemic aims, data, methods, and scientific products, the more likely they are from the same project.	If two papers about a topic are from the same project, then if we compare them to two random papers from the same topic, the first two papers will discuss more similar instances of problems, questions, aims, methods, products, phenomena, etc., than will the latter two papers.	

2.4- Protocol to Infer Research Projects

While the previous sections describe tools to conceptualize research projects, this section provides simple protocols to isolate them from one another, to collect them in populations, and to select projects from those populations for case study analyses. Protocols to ensure the reliability and replicability of our studies, features that I listed as desiderata on the total suite of tools provided in this chapter.

In this section, I provide a general protocol to infer populations of distinct projects. In Appendix A, I provide a more specific protocol for the same task that employs a suite of digital tools. Third I provide a general strategy for selecting small samples of cases. Fourth, I use the specific protocol to infer a population of research projects at the interface of evolutionary genetics and developmental genetics, and I use the general strategy to select two cases from it.

General Protocol

To use the general protocol below, I must begin with a topic, or a collection of research papers about a common theme. Anyone can use this protocol when they aim to infer a set of possible research projects from that collection. So the protocol presupposes condition ORP-1. The protocol also presupposes that, if we use it, we have a rationale for our own projects that includes our research questions, aims, phenomena of study, etc. The general protocol is:

1. Use informal methods to list a small collection of research projects (4–8) related to the theme of interest. This is a set of known cases.
2. Form a topic, or a collection of papers about a common theme.

3. For any two papers, compare the authors, and place the papers that share an author in a common pile or folder (ORP-4).
4. From the papers in each pile or folder, infer the team members (ORP-3) and the temporal period of the project (ORP-5).
5. List each putative research project, its members, and its time period. Each pile or folder represents a putative research project. The collection of piles or folders represents a population of research projects.
6. Check for false negatives in the population of cases. Use the set of known cases from step (1). Compare the population to that set, and if the population lacks many of those cases, begin again at step (1). Use a new method to collect papers for step (2).
7. Check for false positives. Focus on putative projects with many more papers or members, or with much longer temporal periods, than other putative projects in the population. Those putative projects may represent several research projects. To distinguish those projects from each other, use inference methods (V-ORP-6 through V-ORP-9). Furthermore, check that authors with the same name are in fact the same person.
8. Check for false positives. If step six doesn't resolve large putative projects into distinct smaller projects, use V-ORP-10 on the abstracts of the papers.
9. Check for false positives. For each project, search for other papers by the same authors in roughly the same time period about the same theme. Additional papers further support the existence of that project.

10. Check for false positives. For each project, search for other kinds of documents that could be outputs of the project. Such documents include grant applications or reports, white papers or institutional documents, outreach pieces, etc. Additional documents further support the existence of that project.
11. Validate the results. For each project, if possible, contact the members of the team. Ask them if, given the concept of research project, they would classify the papers listed as outputs of the same project. Assent provides further evidence for the existence of the project. Dissent provides evidence against the existence of the project, and perhaps a need to modify the set of papers, the relevant team members, or with enough dissent, the concept of research project.

Notes on the General Protocol

First, the protocol itself isn't surprising. Many who study science implicitly use procedures similar in many respects to the one described above. But insofar as they do, they don't publish those procedures so that others can evaluate their work or redeploy those procedures for novel phenomena. Furthermore, they don't tie their procedures to an explicit concept of research project, as I do in steps (3) and (4). The protocol above makes my procedure explicit, and it enables us to evaluate the reliability of that procedure. It also enables us to check the replicability of results that my procedure yields. As the protocol explicitly ties to an empirical concept, we can evaluate the construct validity of the concept within the procedure. As such, the protocol helps me meet many of the desiderata of strong case study methods.

Second, the protocol makes explicit its steps to check the validity of a representation of an inferred population of individual projects. Steps (1) and (6–11) stipulate validity tests by which we can ensure that the protocol yields not just reliable results, but also true or empirically well confirmed results. Steps (6–11) represent just a few of the possible means by which we might test the validity of those results. Furthermore, they represent further avenues of research. For example, as stated above, steps (1) and (6) represent a rough metric for evaluating the quality of an overall population of research projects. To be more useful, that metric should be made precise.

Finally, step (2) provides something of an anchor for the overall protocol. Insofar as those who study science collect a small number of papers, or focus on the papers of already established research groups, the populations that they infer will be biased, as will studies of the cases in those populations, thus inviting legitimate concerns of cherry-picking. If the protocol above is to yield a good representation of a population of research projects inferred from many research articles, then step (2) must be further specified.

A More Specific Protocol

In Appendix A, I provide a more specific protocol. I use it to overcome the limitations of the general protocol, to address the above issue facing step (2), and to further specify steps (2–5). The protocol employs a suite of digital tools. Those tools include the Web of Science database of scientific papers, software for analyzing those papers, specifically the Tethne program and research system developed by Erick Peirson and his colleagues (Peirson et al. 2015), and software for representing networks of those papers, specifically Cytoscape (Kilcoyne et al. 2009; http://www.cytoscape.org/).

The output of the protocol is an image of multiple networks. Each node represents a paper, each edge represents a relation of having an identical author between two papers, and each network putatively represents a research project.

As each network represents a putative research project, the collection of such networks represents the putative population of research projects related to a topic. The putative population is an approximation to the actual population of projects related to a topic. The above protocols bias the putative population in at least two ways. First, they reconstruct only those projects that had research papers as outputs. Second, they reconstruct only those projects that published papers in journals indexed and sorted by the academic database used, Web of Science in this case. Furthermore, much depends on the sets and structures of keywords used to generate a corpus in the databases, and those who use the protocols will have to tinker with those sets and structures before they get usable results.

Furthermore, there is a trade-off between the specificity of a theme and the size of the population of research projects. The more specific the theme, as represented by the structure of the intersections and unions of the search results of many keywords, the smaller will be the population. The less specific the theme, the larger the population. If the population is small, we face issues of how to select projects for individual case study without biasing our results. I address this issue in the next section. Also, if the population is large, many of the networks will represent clusters of many research projects, and a case study researcher will spend much time distinguishing distinct projects.

Finally, if we conduct detailed case studies on a research project or on projects identified by the above protocols, the results of those studies generalize to the population

from which they were selected. Thus, the concept of research project and the protocols enable us to specify an answer to the question raised in the introduction. That case is a case of what, exactly? If we use the concept and the protocols, we can answer: That case x is a case of research projects RP related to topic T. Each individual answer must specify a single case x, a population of cases RP, and T.

Selecting Small Samples

An issue remains. I have developed a concept of research project and protocols for finding populations of research projects. My aim is to use those as tools by which to address the problem of (potential) cherry-picking of cases by those who study science. If we wish to do a case study of science, if we focus on research projects and use the protocols above, and if we develop sufficiently large populations and samples of research projects, then we should be able to select individual projects from the population for focused study without cherry-picking. But we still face the issue about the representativeness of the cases.

But if a sample of cases is small in proportion to the population, even if the sample is randomly selected, the sample will exhibit bias (Seawright and Gerring 2008). Even with well-delimited populations, almost all case studies have small samples. Thus, even if we don't cherry-pick our examples, we will still face the problem of which cherry picking is just one source: imprecise samples and therefore imprecise results.[9] The problem is one of practice, not of principle. For a given case study, those conducting the

[9] I suspect, contra (Kinzel 2015), that this issue, more than issues of theory-ladenness, yields the appearance of cherry-picked cases when researchers didn't in fact cherry-pick their cases.

study *could* intensively study enough cases to have a large enough sample to yield unbiased results.

We can see the weight of the practical problem, however, when we examine the size of a population. For small populations of less than 100 individuals, as populations of research projects often are, precise samples would require the researcher to select at least half of the cases to infer conclusions about the population with high statistical significance. For larger populations, the proportion of a sample in relation to the overall population lessens, but the absolute number of the sample will be larger than those for small populations. For instance, a population of around 1,000 would require a sample of roughly 278 for conclusions to be statistically significant according to standard norms of high significance.

So those who wish to use case studies to investigate research projects face a dilemma. Even when they precisely delimit a population of cases, if they randomly sample a few cases for focused study, their results will still be imprecise. But if they randomly sample enough cases to avoid bias, they will assume an especially onerous task.[10] If we accept the dilemma, the most straightforward path is to accept the second horn and to grind out the onerous task. But I think we can find techniques to lessen the force of the first horn, and thus still fruitfully study small samples of cases.

So how do we select a small sample of cases in such a way as to minimize the imprecision of the sample? Random sampling isn't sufficient. Seawright and Gerring (2008) address this question by specifying seven selection techniques that are specific to

[10] I find this dilemma to be more problematic than the one (Pitt 2001) proposed. Among the case study pessimists, Faust and Meehl (Meehl 1992; Faust and Meehl 1992; Faust and Meehl 2002) came closest to appreciating and postulating this dilemma.

the kinds of questions asked by those doing the case study research. For one such technique, Seawright and Gerring suggest that researchers partition a population according to some categories or variables, and that they select cases from those partitions. Within each partition, researchers might select cases for any number of reasons and techniques, such as randomly or because the selected cases exemplify some feature that makes them typical to the other cases in their partitions. While Seawright and Gerring focus on relatively large populations and on causal investigations, I argue that their techniques also work for smaller populations and for non-causal investigations.

When we sample instances from a population at random, we do so to maximize the chances that the selected instances are independent of each other. When we partition a population, we do the same thing. Those who study research projects might use any of a variety of categories to partition their population and to ensure the independence of the cases across partitions. They might partition projects into distinct groups according to the institutions at which the projects were conducted, or to the time periods when they were conducted, or to the academic lineage of those who conducted the projects, etc.

I propose the following strategy to select a small number of cases from a population of research projects. First, delimit the proposed population and validate the proposal by checking it for false positives and false negatives. Second, from the question that drives the research into the population of cases, identify a category by which to partition the population. The category functions as a criterion of independence among the cases. Apply that category to each case and partition the population. Select cases from each partition. The default selection technique is random, but look to (Seawright and

Gerring 2008) for other legitimate selection techniques. I illustrate this general strategy below.

Example from the Interface between Developmental Genetics and Evolutionary Genetics

As part of a broader project, I ask: What are projects in which researchers use models from evolutionary genetics with models from developmental genetics? I use the protocol developed in this chapter to answer that question. The question seeks a description of phenomena, not a causal explanation or a prediction on phenomena. The protocol above yields a description that addresses the question. I follow the ten steps to show how to produce that description.

Step 1

Using an informal literature search, I developed a set of five research projects that I could use to later check for false negatives in an overall population of research projects I would infer. I limited the search to include the years 2000 to 2014. The set is:

1. True and Haag on developmental systems drift. (True and Haag 2001; Haag 2007; Haag and True 2007; Haag 2014).
2. Klingenberg on morphometrics and quantitative trait loci in mice. (Klingenberg et al. 2004; Klingenberg 2008; Klingenberg 2010).
3. Hansen and Wagner on the evolution of genetic architecture. (Hansen and Wagner 2001a; Hansen and Wagner 2001b; Hermisson et al. 2003; Carter et al. 2005; Hansen 2006; Hansen et al. 2006).

4. Stewart and Plotkin in the evolution of binding site motifs (Stewart et al. 2012a; Stewart et al. 2012b; Stewart et al. 2013a; Stewart et al. 2013b).
5. Plomin and his team on genome wide association studies linking QTLs and psychological traits in humans. (Plomin and Kovas 2005; Butcher et al. 2008; Plomin and Davis 2009; Plomin et al. 2009).

Step 2

To form a topic, I used the Web of Science (WoS) database of scientific publications. I developed a structure of keywords related to my overall question, which seeks projects that use models from evolutionary genetics and developmental genetics. I represent the structure in Table 2.3.

TABLE 2.3

WEB OF SCIENCE SEARCH STRUCTURE

Evolutionary genetics	**Developmental genetics**
"evolutionary geneti*"	"developmental geneti*"
"population geneti*"	"molecular geneti*"
"quantitative geneti*"	"gene regulat*"
	"gene network"
	"genetic pathway"
	"cis regulat*"

Each cell represents search terms used in a single search. Terms are in quotation marks to ensure that search results return complete phrases. Stars (*) enable WoS to return results with similar suffixes. For instance, "gene regulat*" returns results for both "gene regulation" and for "gene regulatory". Each column represents the union of the results returned for the single searches in its constituent cells, and the table as a whole represents the intersection of the results from the columns.

For each cell, I used the terms in the cell to do a unique search on WoS. Then, for each column, I combined the search results from the cells into single databases. Finally, I took the intersection of the two columnar databases to create a final database, which had 632 scientific papers and was the topic of the project.

Steps 3–5

To group papers in the topic by shared author, I used Tethne, its author_coupling() function, and default program settings. I exported the results as a GraphML file. I opened the file in Cytoscape, and using a Prefuse force-directed layout, I represented the network of papers as Figure 2.2. Each distinct network represents a putative research project. Each node represents a paper, and each edge represents a relation of same author between papers.

The image represents 367 isolated papers, 48 networks with only two papers, and 32 networks with three or greater papers. Cytoscape enabled me to output the networks, their authors, and dates, and thus to infer the project team members and the temporal periods of the projects.

Step 6

To check for false negatives within the overall population, I searched the representation for the authors and projects listed in Step 1. The representation included at least some of the papers for all of the projects, except for the Hansen and Wagner project, which wasn't represented at all. I concluded that the representation captured the vast

majority of projects relevant to my questions, but that it could be missing as many as 20 percent of them.[11]

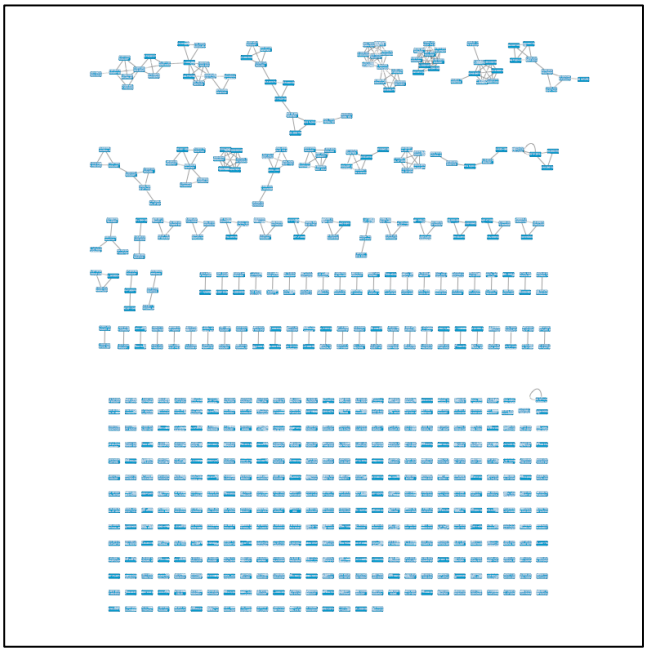

Fig. 2.2. Population of Putative Research Projects. The figure above shows networks that represent putative research projects. Each node represents a paper, and each edge represents a relation of shared authorship between the nodes. Each isolated network represents a putative research project. Further work must be done to delineate actual research projects.

Steps 7 and 8

To delineate projects in large networks, I applied conditions V-ORP-6 through V-ORP-9 to the networks represented in Figure 2.2.

[11] I said earlier that we need to build stronger and more quantitative metrics and procedures of external validity for the results of the protocol. Once we have those metrics and procedures, if the false negative number proves to be too high, we need to return to and revise the keyword list and begin the process again.

Steps 10 and 11

I applied these steps only to the two cases I selected for further study, as per the discussion below.

Case Selection Steps

From the population of projects represented above, I wanted to select a sample of cases for further in-depth study. But I faced issues of selecting small samples.

To select cases, I began with the putative projects represented in Figure 2.2. I used ORP-2 to eliminate all of the isolated networks of single papers as unlikely to provide much information about sustained research projects. Next, I partitioned the remaining projects intro three groups based on the kinds of populations studied in the projects, as described in the abstracts of the papers. Those projects in the first group studied purely theoretic organisms as part of modeling processes. Those projects in the second group studied organisms in laboratory populations. Those projects in the third group studied organisms in wild populations. Those three kinds of populations, theoretic or lab or wild, represent independent phenomena, and they require identifiably distinct research strategies and assumptions (Winther et al. 2015). To simplify this dissertation, I combined the partitions for lab populations and wild populations into a single partition.

From each partition, I selected a case at random, for a total of two cases.

1. Theoretic Population: Norman Johnson, Adam Porter, and Alex Tulchinsky's studies of hybrid incompatibility due to mutations in gene regulatory architecture.

(Johnson and Porter 2001; Porter and Johnson 2002; Tulchinksy et al. 2014a; Tulchincsky et al 2014b).

2. Wild/ Lab population: Greg Wray's group studying evolving *cis* regions in wild sea urchins. (Balhoff and Wray 2005; Garfield et al 2012; Runcie et al. 2012).

Next, I confirmed with further documentation that each of the three cases listed above in fact were the outputs of single research projects. For instance, with case 1, I found further papers that were outputs of the same project (Johnson and Porter 2000; Johnson and Porter 2007; Johnson 2007), and personal communication with the authors confirmed the existence of a distinct project (Johnson 2015; Porter 2015).

Finally, with well-articulated topics, populations of research projects, and individual cases, I could in-depth case studies, described in later chapters. Given those topics, populations, and cases, I could compare the cases to each other and specify the generalizability of my results. I could also use the map of epistemic relations in Figure 2.1 to describe the parts of each project.

2.5- Conclusion

Limitations of the Tools

Each of the tools has some limitations. The concept of research project is a categorical concept, and it faces all of the problems that such concepts face in empirical domains (Cassirer 1923). As research into research projects develops, it may prove useful

to replace or refine the categorical concept with a concept of a different logical form, especially for tasks other than description of research projects.

For example, we can't use the categorical concept to definitively distinguish all research projects from all other research projects. Tough cases remain, and often those tough cases are prevalent in populations or networks of research projects. Furthermore, we can't use the concept, without adding to it, to predict the development of new research projects.

The protocols have their limitations, as well. The protocols work best when studying research projects that focus on relatively narrow topics in recent history. As we focus on broader topics, the protocols yield increasingly larger populations of projects, connected in increasingly more complex networks. Those who study broad topics in science will have to work harder to distinguish research projects than those who study narrow topics. But even if they focus on narrow topics, the older the time period studied, the less relevant become scientific journal articles and the less helpful become digital journal databases. Given the temporal scope of articles in JSTOR, WoS, and Scopus, those databases are of no help for research into scientific projects that existed only before 1920 or so. Thus, the protocols are limited in their robustness.

But given those limitations, the tools remain useful. While together they don't perfectly pick out research projects or populations of them, they still increase the precision, transparency, and replicability of our attempts to do so. A critic may not like how someone distinguished research projects form each other, but if the person who distinguished them used the tools above, the critic will be able to pinpoint exactly where

she disagrees with the procedure, and she'll have a framework in which to suggest revisions.

Questions Raised and Addressed

Throughout this chapter, I raised several questions. I return to them now and review them in light of the success conditions I set. The conditions are:

1. Show how to delimit a population of cases, provide criteria to select cases—justified according to the research question, and select cases.
2. Make explicit how to evaluate inferences from evidence with tests of reliability, construct validity, and validity.

One question is: how can we identify a research project and isolate it from other research projects that are similar to it either because they share a general topic or because they are pursued by the same people? The concept and operationalization of research project, and the protocols enable us to do so.

Together, they meet the two success conditions. The concept and operationalization enable us to identify and isolate research projects, and they enable us to employ the protocols to delimit a population of cases. The concept of research project enables tests of the construct validity of various operationalizations of it, while the V-ORP conditions of the operationalization enable validity checks of isolated and putative research projects. The protocols provide replicable and reliable procedures for inferring the existence of research projects from corpora, and for checking the validity of populations of putative research projects.

A second question is: how can we distinguish those cases that weren't cherry-picked? I argue that cherry-picking is just one manifestation of a larger problem of selecting cases that aren't independent from each other. If we want to ensure that a researcher selects independent cases, we need to check that he provides a population of projects and a partition of that population. Next, we need to check his criteria for partitioning that population, and we need to ensure that those criteria are theoretically grounded.

A related question is: how do we select a small sample of cases in such a way as to minimize the bias of the sample? We use theoretically grounded criteria to partition the population. Then we specify our technique for drawing from the partitions. Finally, we strive for multi-case studies.

A fourth question is: how do we precisely specify scope and delimit local epistemologies, and how do we establish patchy-unity? I argue that we use the map of epistemic relations to describe the local epistemology of a research project. We compare the instantiated models across projects to establish the similarities and differences in those projects. Similarities reveal patchy unity.

The question that motivated this chapter is: That case is a case of what, exactly? Given the tools provided above, a general answer is: That case x is a case of research projects RP related to topic T. Each individual answer specifies a single case x, a population of cases RP, and T. For the example of evolutionary developmental biology provided in section 4, the answer is: Johnson and Porter's project to study hybrid incompatibility due to mutations in genetic architecture in simulated populations is a case

of the 80 or so projects in which researchers use models from evolutionary genetics with models from developmental genetics.

The tools provided in this chapter provide only some of the tools needed to study research projects within case study designs. But they enable such projects to begin from a good foundation. For studying the particular aspects of particular research projects, and for systematically comparing those aspects across research projects to address investigative questions, further methods are needed.[12]

[12] I argue for three main steps in case study research of projects. 1: Identifying a population of projects and selecting some for study; 2: Collecting data and inferring conclusions about particular research projects; 3: Comparing those data and inference across cases. In the next chapter, I develop a protocol for an instance of the second step.

CHAPTER 3

HOW TO CONCEPTUALIZE AND STUDY EPISTEMIC GOALS

3.1- Introduction

This chapter proposes a concept of epistemic goals or aims such that they are aspects of scientific research projects, and it describes a general procedure to study them with content analysis methods. The chapter ends with some of the documents required to use content analysis in this dissertation. The tools provided in the chapter enable those who study science to address some perennial issues that they face. First, the concept provides an explicit but fallible framework with which researchers can study the aims of sciences, a historically vague topic. Second, the content analysis procedure provides an instance of a general procedure that enables researchers to systematically collect data from, and make inferences about, scientific texts. That procedure provides one route by which those who study science can improve the reliability and replicability of their results.

Those who study science, especially philosophers of science, have a recurring interest in the aims of science, sometimes called the ends or epistemic aims of science. Recent discussions often argue for one preferred or primary aim of science (Kitcher 2001; Potochnik 2015), for the need to study such aims (Hardcastle 1999; Kitcher 2004), for their social aspects (Fallis 2007), or for their relations to concepts or theories (Brigandt 2010; Brigandt 2012; Love 2013). Furthermore, some have noted that there are multiple aims, and that philosophers too often focus on explanation at the expense of other aims (Douglas 2009 and references therein).

Those discussions evolved from discussion from the late 1970s to the early 2000s in which researchers often collapsed concepts of epistemic aims with those of epistemic virtues or cognitive values (Kuhn 1977; Laudan 1983; Laudan 1987; Giere 1988; Laudan 1990a; Laudan 1990b; Kitcher 1993; Rooney 1992; Longino 1996; Goldman 2002). While some researchers distinguish goals from values, acknowledging that the two are conceptually related, sometimes they are still lumped together (Douglas 2013). Older still are canonical discussions about the aims of science (Duhem 1914; Hempel 1952; Popper 1957).

But given those discussions, we still lack explicit tools with which to study epistemic aims as actual aspects of scientific practice. We lack a reasonably precise concept of epistemic aims with which to distinguish those aims from epistemic values, and the concept of epistemic aim remains implicit. We also lack a tentative and revisable list of the different species of epistemic aims, so we tend to focus only on explanation and prediction. And we lack methods with which to empirically study the epistemic aims of actual scientific research projects, or with which to settle conflicting hypotheses about the aims pursued in those projects.

This chapter aims to ameliorate, but not to completely remedy, those issues. I propose a general concept of epistemic aims as it applies to scientific research projects, and I provide a taxonomy of different kinds of epistemic aims. Furthermore, I outline a procedure for studying the aims of research projects, and collecting data about those aims, using content analysis methods.

I organize the rest of this chapter as follows. In the next section, I review my account of research projects and my map of epistemic relations for those projects, in

which values and epistemic aims are central but distinct aspects. In section three I briefly present a rough idea of epistemic aims, then I propose a focused concept of epistemic aims, and I provide a taxonomy of kinds of aims. Section 4 provides a brief review or introduction to content analysis methods, and it shows how to adapt them to study epistemic goals. Section 5 provides operational definitions, specific to content analysis methods, for some of the species of aims, and it provides a protocol for data collection. Section 6 concludes this chapter and looks ahead to further chapters.

But first some caveats. First, this chapter doesn't provide a final theory of all the aims that have ever been or will be pursued in science. It provides an incomplete list of aims that researchers have pursued as part of research projects, for at least in a short period of time that includes the 1950s to the 2010s. Furthermore, the tools provided are first attempts, and will require revision in light of new evidence.

Second, I treat the term 'aims' as roughly synonymous with the terms 'goals', and 'ends', and to some extent, 'functions'. When I focus on people or research projects, I talk of aims, goals, and ends. When I focus on models, theories, or other scientific products, I talk of functions and of their epistemic functional relations to phenomena. I acknowledge that in other contexts there are further and subtle differences in meaning for those terms. For my purposes in this chapter, such further subtleties are unimportant.

3.2- The Context of Research Projects

I conceptualize epistemic aims as components of scientific research projects. Epistemic aims function centrally in the rationales of those projects. When researchers

design research projects, they do so partly by specifying the aims they wish to accomplish. Thus, we must appeal to a research team's epistemic aims to describe or explain the rationale of its projects. Furthermore, when we evaluate a team's rationales for proposed or for completed projects, we appeal partly to the team's proposed or avowed aims and to the achievability of those aims from the team's proposed or used research procedures.

For a given project, the specific aims pursued have several functions. First, they provide one set of criteria by which the team selects its methods to collect data. Second, they provide a standard that directs the team as they construct their scientific products or deploy them in relation to data. Finally, they provide one of several benchmarks by which the team and those outside of the team can evaluate the effectiveness of the products and the success of the project.

Over the career of a project, the team may revise or change its epistemic aims in light of many factors, including unexpected data, thus altering the project's rationale and import. Hence, in the career of a project, a team most strongly commits itself to specific epistemic aims, or any other component for that matter, when it rationally reconstructs the project and publishes it in the format of research presentations, posters, and papers.

While we evaluate projects as epistemic units according to the aims the teams ultimately set for those projects, we needn't evaluate scientific products or data according to those same aims. Those who produce scientific products or data can select poor goals by which to evaluate their products, sometimes due to lack of imagination but likely more often due to theoretical or social constraints. Products or data from projects may be, and

often are, repurposed much later by other teams in other contexts with other rationales. In that way, the value of data and products outlives the value of research projects.

3.3- A Concept and Taxonomy of Epistemic Aims

This section proposes some conceptual tools. While we can empirically describe the rationales of proposed and completed research projects, we can also evaluate those projects according to at least the problems they ameliorate, the questions they address, the values they satisfy, and the epistemic aims they meet. If we want to empirically describe those aspects, we need tools and concepts by which to identify them. This chapter focuses on epistemic aims, and below I describe a rough idea of epistemic aims, then I propose a concept of such aims, and finally I propose a taxonomy of aims.

The Rough Idea

Epistemic aims are by definition a species of a more general concept of aims. Before I propose my concept of epistemic aims, I first discuss some aspects of a rough idea of aims. That discussion will provide some focus and common ground, and from it I propose my concept of epistemic aims in the next subsection.[1] In this and the next few subsections, I use the term 'goals' more often than the term 'aims', though I treat them as synonyms.

[1] To be clear, the rough idea doesn't provide a justification or an exact target for the concept later proposed. Instead, discussing the rough idea is more like collecting the clay from which to fashion a pot. Such a discussion, though now somewhat unfashionable, used to be called clarifying the explicandum (Carnap 1950).

When we think of goals, we often think of the goals that individuals set for themselves or the goals that social groups set for the group. Psychologists study the first kind, and many kinds of social scientists study the second, including researchers who study organizations, city planning, design, etc. Psychologists have done more than others to propose concepts of goals, to operationalize them, and the to test the construct validity of those concepts.

For individuals' goals, useful concepts are trickier to specify than might seem. Most dictionary definitions interdefine goals with aims and ends, giving little help. I start by thinking of goals as things that people set or specify and work to achieve or avoid (Elliott and Fryer 2008). In that sense, goals partly motivate people to act, and they regulate those actions so that people fallibly strive to meet those goals.

We also use goals to evaluate the actions and projects of others, especially organizations. Some agent can set a goal but do nothing to achieve it, such as when the US government explicitly aimed to irrigate the western US for mostly family farms, but did nothing to ensure that irrigated water in fact reached such farms. Furthermore, someone can, with perfect strategy and execution, achieve a goal. But if we deem the goal wrong, perhaps such as Kit Carson's achieved goal of military subjugation of the Navajos, we criticize the endeavor.

Researchers in many different fields sometimes disagree about how to further define goals. Are goals states of affairs or are they representations of those states of affairs? Those who argue that goals are states of affairs stress that for an agent to achieve a goal, it must achieve a change in the world and not just create a representation.

Those who argue that goals are representations stress that, in causal processes, future states of affairs have no causal power on prior states, whereas prior representations of future states of affairs can have causal influence on future states of affairs. So if we want to appeal to goals to causally explain events, we should conceptualize them as representations. Furthermore, only the later account can capture our natural discourse in statements like "I have a goal to get a job" or "The electric company aims to stifle solar energy".

I tend to agree with the latter camp about the usefulness of treating goals as representations, but the former camp makes an important contribution about the importance of states of affairs.[2] I next propose a somewhat clear concept of goals. But rather than fitting all of the desiderata of such concepts into a single concept, as many in the previous two camps try, I propose a cluster of concepts to meet those desiderata.

Goal State:

a possible or actual state of affairs that

1. some agent could represent
2. some agent could want to happen (or whatever other cognitive attitude)
3. without action, will not become or persist in being actual.

Goal Representation:

1. a representation of a goal state

[2] Both camps make useful points. But there may be a fact of the matter between them about which concept more closely captures natural language concepts, a fact that matters to researchers when they design surveys to ask people about the goals they or their social groups pursue.

Goal:

a possible or an actual state of affairs (goal state) that some agent

1. actually has or represents
2. wants to happen or to be actual (or whatever other cognitive attitude)
3. acts to bring about or to cause to happen
    a. and so acts because of (1) and (2)
    b. those action directively correlate with goal state

The above concept of goal is strongly influenced by Raimo Tuomela's account of joint goals (Tuomela 1990). But Tuomela didn't propose concepts of goal states or of goal representations. Neither have psychologists or other social scientists. As a result, previous conceptual toolkits for talking about goals has seemed limited, engendering disputes like the one described above. Such disputes are likely verbal.

Some notes about the above concepts. First, while we tend to think of goals as aiming at the future, we needn't. Given a point in time, goal states are often future states of affairs. But they needn't be, as goals may already have been achieved, in which case the goal states would be prior states of affairs. Furthermore, goal states may be achieved in repeated or regular fashions, in which cases goal states could be past and present states of affairs.

Second, the second conditions of the concepts of Goal State and of Goal describes the possible or actual wanting of agents. Wanting is a cognitive attitude, but it may prove to be the wrong or an infelicitous attitude to include in the definitions. A different

cognitive attitude, or a group of them, may prove more useful, but I leave the task of arguing for such attitudes for a later project.

Third, the concepts of Goal States and of Goals both refer to agents, which can be many kinds of things. Such agents can include individuals, organizations, businesses, and maybe some kinds of computers. But for automatons, many kinds of organisms, and many kinds of devices, which are not canonically agents, the above concept won't apply. For those kinds of things, the above concepts must be modified, especially the conditions about representations and about cognitive attitudes.[3]

A Concept of Epistemic Goals

For epistemic goals, researchers generally pursue one of two strategies to define or account for them. For the first strategy, they argue for one ultimate epistemic goal or another, such as truth (Goldman 2002), significant truth (Kitcher 1993; Kitcher 2001), or understanding (Potochnik 2015), and they argue that all other epistemic goals resolve in or derive from the ultimate goal. I disagree with those who employ the first strategy, but the disagreement may be verbal. I take truth to be a component of all science, but I don't classify it as a goal. Next, I think significant truths and findings are important, but I think significance derives from questions, as does Kitcher, and ultimately from problems. As such, significance isn't a feature of epistemic goals *per se*. Finally, I argue that there are many ways to understand phenomena, so the appeal to understanding simply passes the

[3] I claim, with (Sommerhoff 1974), that for those kinds of things, representations can be replaced by the more general concept of information. A more difficult problem, as indicated by robust literatures on organismic goals and functions, is how to replace cognitive attitudes.

buck, and often prejudices projects or products that yield causal understanding over other types, such as descriptive or predictive understanding.

For the second strategy, some argue that such ultimate goals don't exist (Hardcastle 1999), nor do general concepts of scientific or epistemic aims, and that the best we can do is to study specific goals for specific teams. On this second strategy, researchers list one or two kinds of epistemic aims, generally explanation or prediction, but they provide no comprehensive list, accounts, or unifying concept. I agree that the there are no universally true concepts of epistemic goals, and that we should study specific research teams and their specific goals. But I don't think we can do so without explicit and somewhat general concepts of epistemic goals, though we must accept that those concepts evolve over time and vary across social groups.

For my proposed concepts, epistemic goals comprise a subclass of goals *simpliciter*. The rule to specify and identify epistemic goals is a further specification of the rules to specify and identify goals. The concepts below each have a list of necessary but not sufficient conditions for those concepts. As research into those concepts progresses, those conditions could change, or we could add to them.

Epistemic Goal State:

a possible or actual state of affairs that

1. some agent could represent
2. some agent could want to happen (or whatever other cognitive attitude)
3. without action, will not become or persist in being actual

4. without a scientific product with explicit relations to evidence, will not become or persist in being actual

Epistemic Goal Representation:

1. a representation of an epistemic goal state

Epistemic Goal for a Research Project:

a possible or an actual state of affairs (goal state) that some agent/team

1. actually has or (incompletely) represents
    a. but if some individuals in the team don't have or represent it, they have or represent subgoals
2. wants to happen or to be actual (or whatever other cognitive attitude)
3. acts to bring about or to cause to happen
    a. and so acts because of (1) and (2)
    b. especially by constructing or using scientific products
    c. such that those actions directively correlate with the epistemic goal state
4. provides a criterion by which an agent in a research project partly selects methods used in the project
5. to be realized or achieved, requires at least one scientific product with explicit relations to evidence
6. provides a standard by which partly to evaluate

a. the scientific products used in the project such that they function to achieve the epistemic goal state
b. the actions of the agent in the project such that they directively correlate with the epistemic goal state and helped yield or employ the scientific products
c. the quality of a research project

Such goals are epistemic in that they relate scientific products to evidence. If we talk of beliefs, an epistemic goal is one component of a justified belief such that the component meets the above conditions. Insofar as the preponderance of the evidence supports the scientific product used to generate that belief, the belief is true. Insofar as there are different kinds of epistemic goals, there are different kinds of knowledge, and there are different kinds of ways to understand the world. For instance, if we can use evidence to support a product that describes some phenomenon, and if we can use other evidence to support a product that explains that same phenomenon, then description and explanation are equally legitimate, but perhaps not equally useful, ways of understanding the phenomenon.

If we limit our talk to publically scrutable statements, an epistemic goal is a statement that meets the above conditions. Insofar as the preponderance of the evidence supports the scientific products described in those conditions, the product is true, or (perhaps) is similar to the world in some respect, and we can use it to achieve the epistemic goal. While the product may be empirically true (or similar, etc.), it may not

meet the epistemic goal set for it as part of a research project. A further claim is that the product in fact achieves or doesn't achieve the epistemic goal.

For my purposes here, I treat epistemic goals as statements or as parts of statements. The above concepts enable us to talk about, or construct statements about, goals in general. Those concepts enable us to pick out *species* of goals, but not *instances*. Types but not tokens.

To pick out instances of epistemic goals, we need a bit more conceptual machinery. We need a list of types of epistemic goals. Again for my purposes here, I treat types of epistemic goals as scrutable statements. Such statements have at least one of two forms. The first form highlights goals as actions and the second highlights goals as states of affairs. In the first form I treat goals as verbs, in the second form I treat goals as nominalizations. The difference is one, for instance, between *explains* and *is an explanation*.

Type of Epistemic Goal (Action)

To X is a type of epistemic goal only if

1. it's possible for an agent to use scientific products to X some phenomenon or another scientific product
2. to X some phenomenon or scientific product as part of a research project, an agent must identify a state of affairs that meets the conditions listed in the concept of Epistemic Goal for a Research Project
3. to X is to realize or achieve the epistemic goal state if one meets further conditions specified as part of an explicit concept 'to X'

Type of Epistemic Goal (Nominalization)

An X is a type of epistemic goal only if

1. it's possible for an agent to use scientific products in an X of some phenomenon or other scientific product
2. in an X of some phenomenon or scientific product, and as part of a research project, an agent must identify a state of affairs that meets the conditions listed in the concept of Epistemic Goal for a Research project
3. an X realizes or achieves the epistemic goal state if it meets further conditions specified as part of an explicit account of 'an X'

Whether or not we choose to talk about types of epistemic goals as action or as states of affairs, the difference is of little consequence. I use the two versions interchangeably depending on the grammar I need to construct my sentences.

Now for an example. Imagine that we observe a pinecone falling from a tree, and we wish to understand that phenomenon. There are many such ways we could understand it. We could describe or represent the phenomena for further study, we could develop tools to predict when other pinecones will fall in the future, we could develop tools to explain how they fall, etc.

Select the task of explaining how the pinecone fell, and let's assume we want a causal explanation. That information fits into the concepts of Epistemic Goal and Type of Epistemic Goals as follows. The relevant possible state of affairs is the state of having explained the pinecone's fall. Insofar as I represent that goal, I meet condition 1 of

Epistemic Goal, and insofar as I want to explain that phenomenon, I meet condition 2. If those conditions motivate us to create or employ scientific product to achieve the goal state, I meet condition 3.

To meet conditions 4 through 6, I need first to employ our concept of Type of Epistemic Goal. I need a detailed account of what logical form or representational pattern causal explanations must have. Select Salmon's causal mechanical account of causal explanation (Salmon 1984). Given that account, I can meet the first two conditions the concept Type of Epistemic Goal, and Salmon's account provides the further conditions, required by condition 3 of Type, to detail what form or pattern causal explanation must have in his account.

Given that account, I can select the methods needed to collect data that will help me use or employ scientific products to achieve that pattern. Thus I meet condition 4. Insofar as I collect data, relate those data to the scientific product, and evaluate those relations according to standards of evidence, I partially meet condition 5. Insofar as the scientific products meet the form or pattern of Salmon's account, they meet the rest of condition 5 and realize the epistemic goal of putatively explaining the pinecone's fall. Finally, given the representation of what I want, and given Salmon's account, I can evaluate whether or not a putative explanation in fact satisfies his account, whether or not my actions in pursuing the project helped me build or employ products that would satisfy his account, or whether or not I was right to even pursue a causal explanation of the pinecone's fall.

Before I list some species of epistemic goals, I note a potentially confusing issue. My proposed concepts above focus on goal states as things pursued, achieved, and

maintained by agents. But we often use the words or phrases of types of goals in other ways. For instance, we often say that a theory or a model, not an agent, explains or predicts phenomena. In this way, we abstract away from the particulars of research projects and describe supposedly eternal features of models. That move is largely rhetorical, as it lends an extra air of objectivity to scientific products by removing agents from the picture. That said, I think that we can legitimately talk about scientific products as explaining, predicting, describing, etc.

When we shift from talking about agents to the tools they create, we should shift from talking about goals to talking about functions. Consider the following statements.

1. The Sears team explains how limbs develop in mice.
2. The representations of causal regularities of morphogenesis explain how limbs develop in mice.

The two statements are similar in grammar, with just the subjects of the statements differing. Yet 'explain' picks out two different concepts in the two sentences. In the first, it picks out an epistemic goal of a research team. In the second it picks out an epistemic function of a scientific product.[4]

[4] To explicate many commonplace statements, then, I might need to develop tinker with the concepts above to specify an account of epistemic functions of scientific products. I don't pursue that end here, but I flag it for a future project. For my purposes here, I note that functions don't exist without relevant systems, or for tools, to relevant agents. When we replace talk of epistemic goals of research projects with epistemic functions of scientific products, we change the relevant social system that gives meaning to the products and their ends. The relevant social system is no longer a local research team, but instead it's a larger community of researchers, perhaps better characterized as a movement, tradition, discipline, or paradigm.

Types of Epistemic Goals

One issue that motivates this chapter is that those who study science too often focus on the epistemic aims of prediction and especially of explanation. To ameliorate that issue, I provide a list of different types or species of epistemic aims.

Philosophers of science, I argue, have spent much of the last one hundred years providing accounts of specific types of epistemic goals. In the most general sense, I treat an account of a type of an epistemic goal as provided the pattern, form, or schemata that some representation must meet or satisfy to count as an instance of a putatively realized goal. Most often, philosophers provide sets of conditions or logical schemata for their accounts of types of epistemic goals, but not always.

While philosophers often provide competing accounts of explanation, they also have addressed representation, prediction, control, and a several other aims. The list provided below is in some way a taxonomy of much of that work. I undoubtedly missed some species of epistemic aims and some accounts. The list is a first attempt and will require revision after further research into specific epistemic aims. But it provides a starting point.

Another issue that motivated this chapter is that those who study science often conflate epistemic aims with epistemic values, sometimes called cognitive virtues of theories. The standard lists of such virtues include simplicity, conservatism, economy, refutability, modesty, generality, accuracy, fruitfulness for discovery, and explanatory power (Quine and Ullman 1970; Kuhn 1977). Critics argued that those virtues weren't purely cognitive and had political content (Rooney 1993; Longino 1996). Those critics suggested further values of novelty, applicability to human needs and dignity, and

diffusion of power. Others suggested that all cognitive virtues derived from the virtue of tending to produce true beliefs (Kitcher 1993; Goldman 2002). The set of virtues has expanded so much that Heather Douglas proposed a taxonomy of virtues based on when we use them in the investigative process (Douglas 2013).

I argue that some things often classed as cognitive virtues would more helpfully be conceptualized as epistemic aims. But even with the concepts detailed above, I provide no perfect tool to always distinguish epistemic goals from cognitive values. That issue partly stems from a lack of clear concepts, as opposed to lists, of cognitive virtues.

Roughly, I argue that we use values, cognitive or otherwise, to help us select among competing problems, questions, and epistemic aims (Brigandt 2015), and to help us select between scientific products that equally ameliorate the same problems, address the same questions, and achieve the same epistemic aims. Within a rationale of research projects, virtues and aims have different functions.

Consider a scientific product that meets a value, such as conservatism, but that doesn't achieve an epistemic aim, such as predicting phenomena. In such cases, the product has no use as a tool, and that it meets a certain value provides us with no reason for selecting it from the toolshed. Now consider the reverse case, a product that fails the virtue of conservatism but predicts phenomena. Its function (achieving an aim) provides a reason for selecting it from the toolshed to begin with. But if we find next to it in the toolshed a product that predicts phenomena and meets conservatism, we have a reason for choosing the latter product over the former.

Cognitive virtues and values provide criteria for selecting between products that achieve the same epistemic aims, not for choosing scientific products *simpliciter*. We

evaluate products partly on if and how well they achieve epistemic aims. Only after that should we evaluate products on other advantages they provide. That's not to say that values have a derivative importance. A product might well achieve an epistemic aim, but if the aim is abhorrent, such as predicting the effects of non-medicated syphilis in uninformed black patients, we can criticize the aim or the project or the team that pursued it.

From the common lists of cognitive virtues, I treat explanation, discovery, prediction, control, and unification as epistemic aims for which we might employ scientific products, not as cognitive values. But my selection is fallible. Some research teams might use concepts like the ones detailed above to conceive other virtues, such as simplicity, as epistemic aims for a project. In such cases, teams would have to specify the pattern or schemata for their account of simplicity, which may prove difficult.

The two tables below provide two lists of epistemic aims. Table 3.1 lists epistemic aims that relate scientific products to phenomena. Table 3.2 lists aims that relate scientific products to other scientific products. There is a column of type of epistemic goals, each unique type underlined. Within the types, there are subtypes, which are more specific accounts, generally provided by theoreticians or philosophers of sciences.

TABLE 3.1

TAXONOMY OF EPISTEMIC AIMS FOCUSED ON PHENOMENA

Family	Type (Subtype)	Rough Description	Select Sources
Phenomena-focused	Description/ Representation of phenomena	Products represent or describe phenomena without necessarily explaining or predicting them	Duhem 1906; Bogen and Woodward 1988; Woodward 1989; Giere 2004; Van Fraassen 2008; Gerring 2012
	Prediction of phenomena	Products forecast future phenomena, yet they needn't also explain that phenomena.	Duhem 1914; Hesse 1966; Salmon 1981; Rescher 1998; Sprites, Glymour, and Scheines 2001
	DN and IS Models	Predictions as arguments with logical derivations from descriptions of laws and conditions to descriptions of phenomena	Nagel 1961; Hempel 1965
	Discovery of phenomena	Products help researchers discover novel phenomena or causes, yet other products may explain those phenomena or use the causes to explain phenomena	Hanson 1958; Sprites, Glymour, and Scheines 2001; Cartwright 2007; Pearl 2009; Caniglia 2010;
	Explanation of phenomena	Products explain phenomena, without necessarily predicting similar future phenomena	Popper 1934/1959; Hesse 1966; Salmon 1989; Humphreys 1989
	DN and IS Models	Products are arguments with logical derivations from descriptions of laws and conditions to descriptions of phenomena	Hempel and Oppenheim 1948; Nagel 1961; Hempel 1965
	Other Statistical Models	Products aren't necessarily arguments, but involve statistics	Salmon 1970
	Erotetic	Products answer questions about phenomena	Bromberger 1966; Van Fraassen 1980
	Causal Mechanical Causal Interventionist	Products display causal structure Products display causal relations that are invariant under manipulation	Salmon 1984; Craver 2006; Craver 2007 Woodward 2003; Pearl 2009; Woodward 2014
	Unifying of phenomena	Products unify many phenomena	Friedman 1974; Kitcher 1989
	Control of phenomena	Products help researchers control or constrain phenomena without necessarily being able to predict or explain it	Popper 1934/1959;

The table represents one family of epistemic aims for scientific products such that those aims focus on phenomena. Within that family, the table lists five types: description or representation of phenomena, predictions of phenomena, discovery of phenomena, explanation of phenomena, and control of phenomena. For some of those types, the table list subtypes. For all types and subtypes, the table provides a brief description of the aim and a few sources by which to learn more about the aim.

TABLE 3.2

TAXONOMY OF EPISTEMIC AIMS FOCUSED ON PRODUCTS

Family	Type (Subtype)	Rough Description	Select Sources
Product/Theory- Focused	Discovery of new products	One pattern is inference to the best explanation	Duhem 1914; Peirce 1908; Hanson 1958; Lipton 2004
	Explanation of products	Products explain other products	Nagel 1961; Salmon 1984
	DN and IS Models	Products are arguments with logical derivations from theories and laws to theories or laws.	Hempel and Oppenheim 1948; Nagel 1961; Hempel 1965
	Amalgamation of products	One pattern is theory reduction.	Duhem 1914; Nagel 1961
	DN and IS Models	Products are arguments with logical derivations from theories and laws to theories or laws.	Hempel and Oppenheim 1948; Nagel 1961; Hempel 1965
	Bayesian	Products that unify other products gain extra evidential support	Myrvold 2003
	Interfield theories	Products incorporate descriptions or laws from more than one field	Darden and Maull 1977; Maull 1977
	Integration	Products incorporate descriptions or laws from more than one field without unifying those fields	Mitchell 2003; Mitchell and Dietrich 2006

The table represents one family of epistemic aims for scientific products such that those aims focus on other scientific products. Within that family, the table lists four types: discovery of new products, explanation of products, unification of products, and integrations of products. For some of those types, the table lists subtypes. For all types and subtypes, the table provides a brief description of the aim and a few sources by which to learn more about the aim.

3.4- Content Analysis Methods to Study Epistemic Goals

Research projects are empirical objects. We can collect evidence about them and about their components, and we can collect evidence about the epistemic goals pursued in a project. There are many potential sources of evidence about epistemic goals within the framework of research projects. But most boil down to extracting information from texts that describe those projects and their constituent aims. This section briefly describes the method of content analysis, which enables researchers to extract that information so that it's verifiable and replicable, and via methods that are repeatable. Later in this dissertation, I use content analysis to extract information from research articles about the epistemic goals of research projects.

Introduction to Content Analysis

Researchers use content analysis for roughly two aims. In the first, they use it to describe manifest content of texts. In those situations, content analysis provides the methods and data with which to test claims roughly like: That text clearly mean/says/states XYZ. Often, such methods are used to read large amounts of texts, numbering in the thousands, and with the aid of computers.

For the second aim, researchers use content analysis to infer implicit content or meanings of texts. Many content analysis methods originated as tools for the second aim, especially in World War II when Allied researchers developed the techniques to infer information about their opposing armies from the latter's propaganda.

Content analysis methods are widely used in the social sciences. Social scientists generally agree that the distinction between manifest and non-manifest or implicit content

is one of degree, not of kind. Increasingly, they argue that the theoretical machinery needed to infer implicit content is also required to ground studies of manifest content.

There are many manuals for content analysis, but the standard one is Klaus Krippendorff's (Krippendorff 2013). His manual covers everything from how to conceptualize and design studies that involve content analysis, to collecting documents, coding them, collecting data, making inferences, evaluating those inferences, and answering research questions.[5] He summarizes all of that information into several diagrams. The diagram reproduced below summarizes the overall process of content analysis by ordering some of its subprocesses.

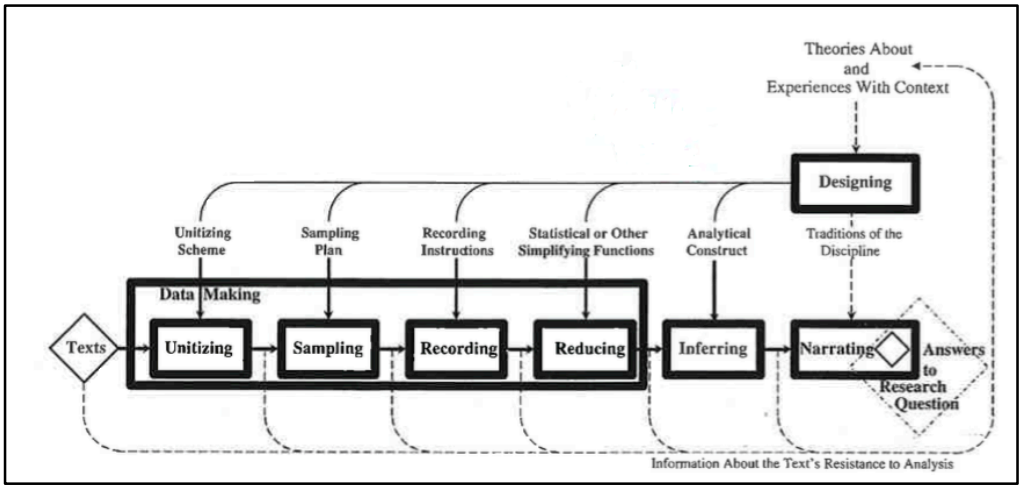

Fig. 3.1. Standard Model of Content Analysis. (Reprinted with Permission from (Kriippendorff 2013, 86)).

[5] Like 'case study analysis', the phrase 'content analysis' has several senses, meaning sometimes the general family of methods, sometimes a specific technique, other times the results of a technique. I hope that the sense appropriate to a sentence will be somewhat clear from the sentence's context.

The diagram represents eight subprocesses, each represented by heavy lined boxes. It also represents eight sets of documents needed, represented in unboxed text near the boxes. The process starts by designing the research with the aid of theories about content and some research questions. Given a set of texts, a content analyst unitizes them into chunks, say paragraphs or sentences. Based on a sampling plan, she samples some of those units, and based on some coding or recording instructions, she records data from them. Given simplifying tools, she reduces or cleans the data. Those processes together comprise the main data-collection subprocess. Once she has completed them, and given an analytic construct of the particular phenomenon that interests the researcher, she infers the content of the texts, which she uses to answer the research questions in a report that accords with the standards of her particular discipline. At any point in the process, a researcher may learn something that convinces her of a problem in her design, and she may halt the process, revise her design, and begin again.

Documents of a Content Analysis

Krippendorff's diagram nicely captures the process of content analysis. But in my use of content analysis methods, I found that Krippendorff's diagram has two problems, both of which are easily remedied. First, the diagram pays little attention to two key processes in content analysis. The first process is evaluating the reliability of the data collecting methods, and the second is evaluating the reliability and validity of the inferences made from those data with the aid of the analytical construct. Both processes are central components of any content analysis, and they mustn't be overlooked. Readers should mentally add those stages to the above diagram.

The second problem is about the documents needed for content analysis. I take the term ‘document’ in a liberal sense, referring to any self-contained set of printed information. The diagram gives too little attention to the breadth of documents needed to complete a content analysis, or to the order in which a researcher must create them before she can move from one stage of the overall process to the next. To remedy that issue, I provide the following diagram.

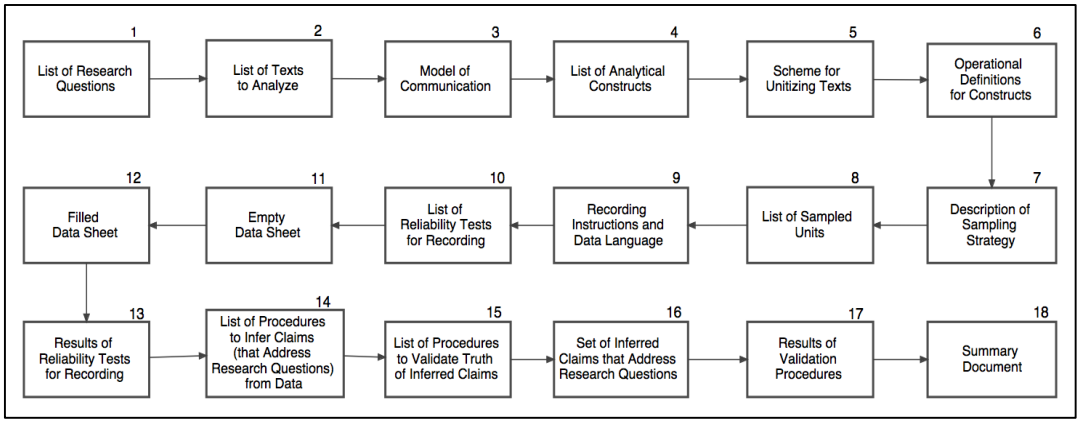

Fig. 3.2. The Documents of Content Analysis.

The above diagram shows 18 documents. A researcher who pursues a content analysis should create at least those 18 documents. And for those evaluating, reviewing, or learning from the analysis, the above 18 documents provide public information that can be criticized, revised, borrowed, etc. Though I order the documents linearly, that order shouldn’t imply a perfect stepwise procedure to content analysis. At any point in the process of creating those documents, a researcher may return to drafts of earlier documents, revise them, and begin the analysis anew.

I won't review the general features of the above 18 documents. Rather, I refer readers new to content analysis to Krippendorff's manual and to the expansive literature on content analysis methods cited in his book. In the rest of this chapter, I provide instances of the above documents specific to this dissertation. In this chapter, I provide instances of documents 1–10, and 14. I collect data for two projects, and I summarize the data from Document 12 and in tables in the next two chapters, which also provide Document 18 for each project. The remaining documents I collect in protocols in Appendices C and D at the end of this dissertation.

3.5- Content Analysis for Epistemic Goals

In the previous chapter, I selected two cases for further study. Each case represents a research project in which the teams in some way use models of evolutionary genetics with models of, or information about, gene regulation. Each of the sub sections below provide the information by which I designed and completed a content analysis on the research publications of those three projects.

1 List of Research Questions

While the overall dissertation pursues several questions, I use content analysis to address only the following one.

What are the epistemic aims of the two selected projects that jointly use models of gene regulation with models of evolutionary genetics?

I use content analysis to provide evidence for the specific kinds of aims pursued in each project. While I could use it to identify many other features of the projects, such as the models employed, I do not. Those features aren't in doubt, at least for my project, so I needn't provide systematic evidence and reconstruction. What are in question, however, are the epistemic aims of those projects, so I use an explicit methods to generate systematic evidence for or against claims about the pursuit of some species of epistemic aims in those projects.

2 List of Texts to Analyze

I use content analysis on a specific kind of text. Each project has several kinds of outputs, one of which includes their published research reports. Those reports include research articles, review articles, and dissertations. For each project, I focus on the set of published research reports. I don't here list all of the articles selected for each project. I instead list them in the following two chapters, in the relevant subsections specific to the individual projects.

I ignore other kinds of outputs from the projects. I didn't analyze grant applications or reports, interviews, or research notebooks. Those kinds of documents surely have much information about the epistemic aims that the teams pursued in their projects. But I leave them for additional studies, and for future tests of the validity of my results.

3 Model of Communication

Each content analysis assumes a model of communication. Content analysts conceive of content as not always perfectly manifest in texts, but instead as something that holds between the creators of texts and the consumers of those texts (Krippendorff 2013). If the creators of a text remain constant, then the contents of the texts will vary with different kinds of consumers. Content analysis involves inferences from textual data to content. To ground those inferences, it presupposes a model of communication that specifies the creators, the kinds of consumers, and the behaviors of those consumers that the texts may influence.

Even though many have studied communication in science, few have developed well-articulated models of communication for research reports. Many content analyses of such reports forego a communication model, and thus they limit or undermine the strength and clarity of their inferences. To better ground their inferences, and ultimately their causal explanations of science and its progress, those who study science have much to gain by further specifying their communication models.

Of those who have developed models, they often build on the sender-receiver model of Shannon's model of information flow (Shannon 1948). Communication theorists often describe that model as too simplistic to capture the complexity of communication between people, but when they propose alternative models, they generally build on Shannon's model and add arrows for feedback from destination to source. I similarly take Shannon's model as basic.

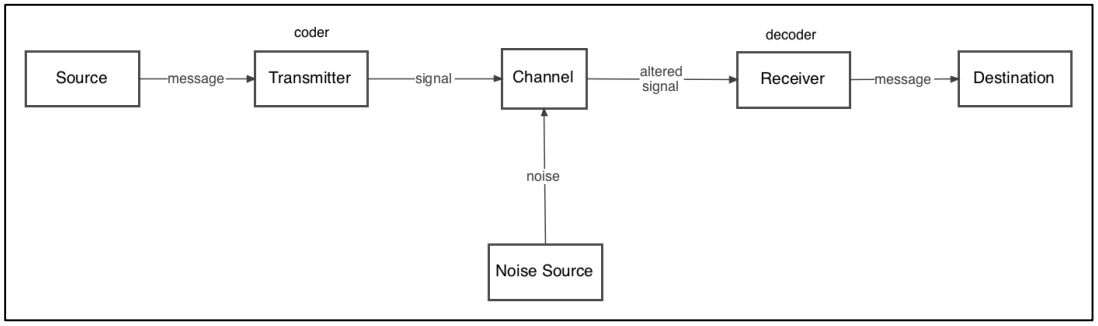

Fig. 3.3. Shannon's Model of Information Flow. (Redrawn from (Shannon 1948)).

In the early 1970s, Garvey and Griffith developed the now most commonly cited model of communication among scientists, which traces a normal process of development for scientific research, from informal discussions to presentations at meetings to published papers (Garvey and Griffith 1971). The model built on their earlier work, in which they'd argued that to investigate communication among scientists, researchers should conceptualize that communication as a social phenomenon that produces scientific products and information, among other things, which feed back into the overall system of communication (Garvey and Griffith 1967). Garvey and Griffith's model remains standard today, modified only slightly by the advent and implementation of contemporary technology, such as personal computers, digital databases, the internet, and email (Brown 2010).

I focus on only the final stages of Garvey and Griffith's model, in which researchers publish articles of their own research or consume the articles of others. That stage lacks detail. If we focus on those stages, we can loosely model the interactions as a Shannon processes. For one such process, the source is the set of article authors, the

destination is the set of intended readers, and the article is the channel, or perhaps a packet of information in the channel.

But I need to add a bit more to the model. Critics of Shannon's model argue that the model is too simplistic for human communication, as it incorporates little about feedback and because it abstracts away from a key component of human communication: the purposes for which people engage in it.

Both points are correct, but they don't undermine Shannon's model. Shannon's model captures information flow in its most general and abstract form. If we conceptualize human communication as a kind or species of information flow, then we should de-abstract the model a bit so that it more fully captures features that we wish to study. For instance, when theorists add arrows to the model to represent feedback from destination to sender, which they often do, they have de-abstracted the basic model to capture a more specific activity. I do the same for the purposes of scientific communication.

To succeed, communication among people must satisfy at least some purposes from at least two sets. The first set includes the purposes of those who send the information. The second set includes the purposes of those who receive the information and return feedback to the sender. The more purposes satisfied, the more successful the communication.

For a research article, we can focus those purposes. For an article to succeed as a bit of communication, it should satisfy at least some purposes of the authors or at least some purposes of the readers. The more purposes satisfied, the more successful the article

functions in the process of communication. Those who study communication between scientists have noted those two sets of purposes.

Researchers have studied the purposes of scientific authors. Garvey and Griffith argued that authors of scientific articles aimed to change the behaviors of those who read those articles (Griffith and Garvey 1967; also Bazerman 1981). According to them, authors want their readers to recognize a record of priority and complete information and to cite the articles in review articles or textbooks. Gusfield stressed that authors aim to convince their readers about the truth of the claims in their articles (Gusfeld 1976). He also argued that authors also interpret their data, especially in social sciences, in such a way as to subtly influence the attitudes that readers form about the objects studied (Gusfield 1976). While some have studied the potential aims of those who author research reports, much more research is needed (Hyland and Salager-Meyer 2008).

Researchers have also studied the purposes of those who read articles. Many have noted that scientists read the reports of others to discover relevant background knowledge and literature for their own research (Garvey and Griffith 1967; Bazerman 1981). Tenopir and Volentin surveyed biologists to find the purposes for which biologists read academic literature (Tenopir and Volentin 2012). They reported that biologists read articles to inspire new thinking in their research, to improve their own results, or to change their research focuses. Biologists report that, of the articles they read, they read almost 80% for research purposes. But Tenopir and Volentin found further purposes as well, including 7% for teaching and 11% for current awareness. Other sciences reported similar numbers.

From the above information about purposes, I sketch a tentative communication model for communication between scientists via published research articles.[6]

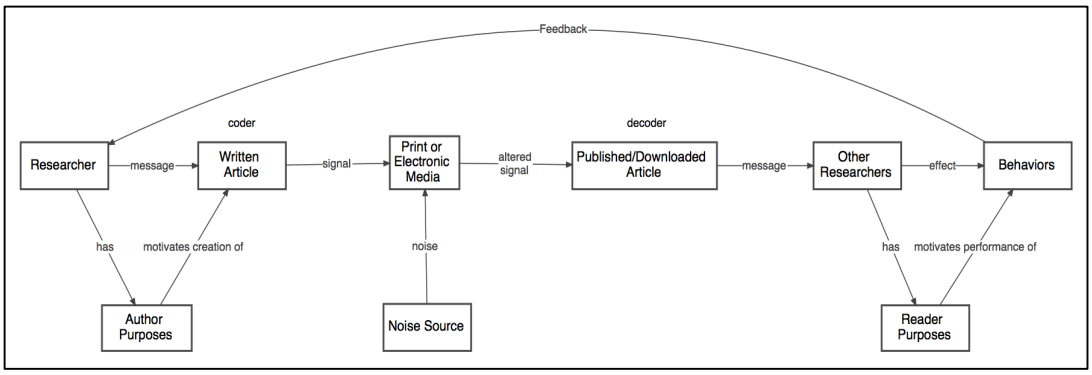

Fig. 3.4. Basic Model of Communication via Scientific Publication. The model represents key nodes in the process of information transfer among scientists when they publish scientific articles. While it specifies the basic nodes of a Shannon model, it also adds three new nodes: Author Purposes, Reader Purposes, and (Reader) Behaviors. It also adds a channel of feedback from the readers (behaviors) to the original researcher.

Figure 3.4 specifies the nodes in Shannon's model, and it adds three new nodes: Author Purposes, Reader Purposes, and (Reader) Behaviors. Those nodes help explicate the social aspect of the communication process, the goal-directedness of it, and that when a research team communicates about their research, they ultimately aim to change the behaviors of their readers. Figure 3.5 adds more detail to Figure 3.4, and it lists species or kinds of purposes or behaviors.

[6] I stress two caveats. Perhaps no human communication is as purpose-driven as is that among scientists via their published reports. Thus, the model specified below may little represent the process of communication via published documents in non-scientific communications. Furthermore, even among scientists, little communication between them is as purpose-driven as is that via their published reports. Thus, the model specified below may little represent other kinds of communication among scientists.

The above model has many uses. I use it in this dissertation to ground descriptions of the phenomena I study, the epistemic goals pursued in research projects. But researchers often use Shannon's model, and derivatives of it, to explain and predict phenomena. To use the above model to explain some processes of communication among scientists, researchers must show that the additional nodes are isolable and manipulable. I doubt that their isolability will be difficult to show. More interesting, and more fruitful, will be their manipulability. But that is a project for another day.[7]

For this dissertation, I use the Figures 3.4 and 3.5 to ground my content analysis of the articles published as the outputs of the research projects I study. Given the above models, a research team partly writes and publishes research articles to convince others to use those products or to believe the results of those products. There are many strategies to convince readers. A common one is to use the articles to display the products and to show how they helped the research team achieve its epistemic aims. Sometimes authors are explicit about their epistemic aims, but often those aims are implicit. Content analysis helps reveal those aims, be they explicit or implicit.[8]

[7] To show that they are manipulable, we must show, for example, that altering a researcher's purposes for communicating will alter how she constructs her article, the medium in which she publishes it, the researchers who read it and change their behaviors (or not) because of it, and the feedback the original researcher receives also changes.

[8] Given the above models, there are many interesting projects that those who study science could pursue. One that seems especially important is to examine the rhetorical structure, or content model (Introduction, Methods, Results, Discussion, Conclusion), of research articles to see whether or not that structure helps or frustrates the sale of scientific products to readers. I leave that project for another day.

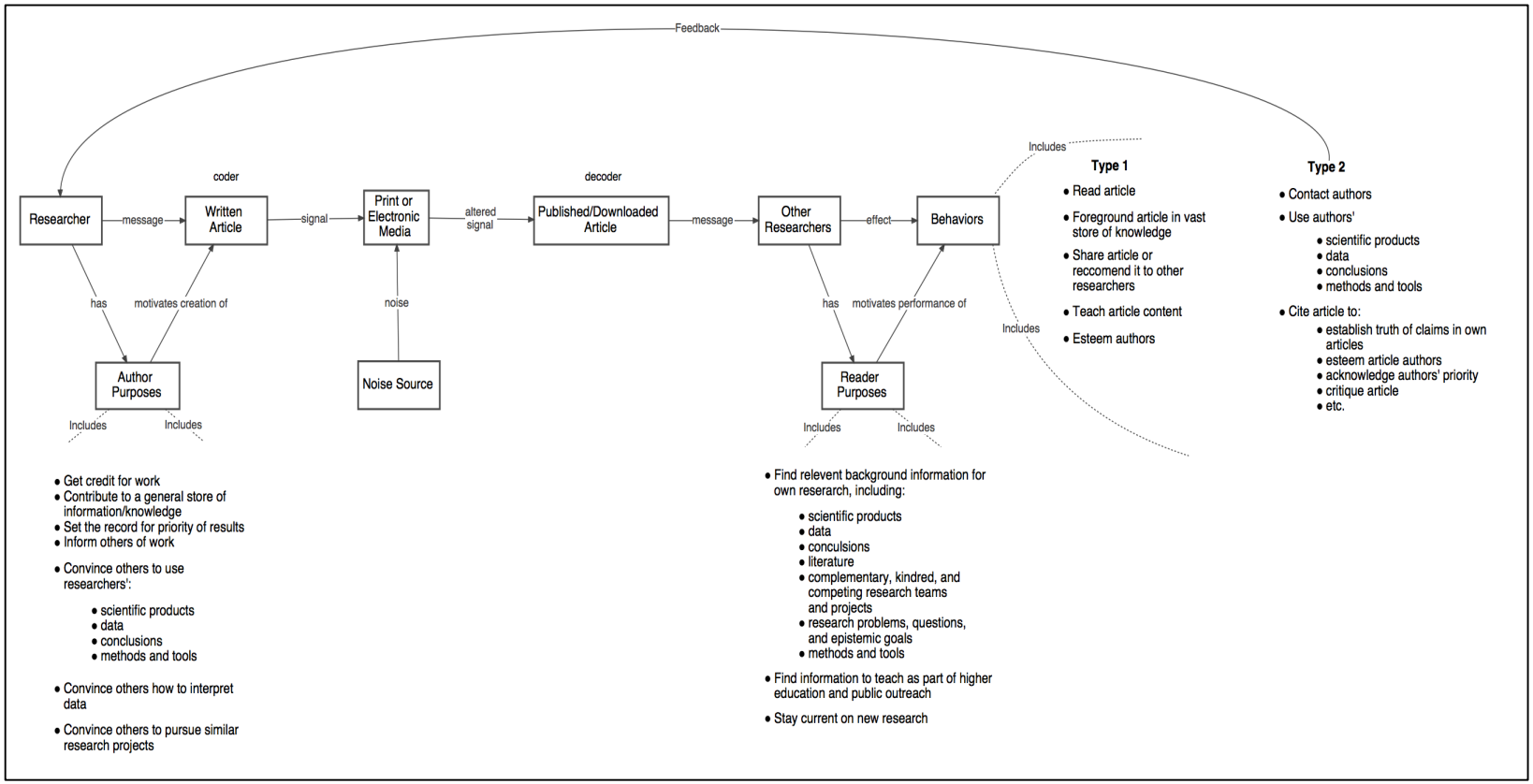

Fig. 3.5. Fuller Model of Communication via Scientific Publication. This figure lists some of the kinds of purposes and behaviors that may matter in communication among researchers via published article.

4 List of Analytical Constructs

This dissertation relies on two analytical constructs. The first is that of research project, reviewed in section 2 of this chapter and more fully discussed in Chapter 2. The second is that of epistemic goal, detailed in section 3 of this chapter. The first construct enables me to conceptualize and bound the cases I study. The second enables me to focus on a specific phenomenon of interest in those cases.

Especially important is the taxonomy of epistemic goals listed in Tables 3.1 and 3.2 above. Those tables provide the specific goal concepts that I operationalize below.[9] While I also study the models deployed in the research projects that I study, those units of analysis are less difficult to identify that are research projects or especially epistemic goals. As such, I don't treat models or arguments as analytical constructs, which means I don't provide explicit definitions or operationalizations of those notions. The choice is one of convenience, and a more in-depth study could treat those units of analysis as constructs.

5 Scheme for Unitizing Texts

To unitize the texts studied in this dissertation, I followed the following scheme.

1. Turn each text into a txt document.
2. Each paragraph is a unit.
3. The abstract is a unit.
4. The article title is the first unit.
5. Each header is a unit.

[9] Thanks to Wes Anderson for stressing this requirement to me.

6. Each numbered indent (e.g. a model assumption) is a unit.
7. Each graphic/table/figure caption is a unit.
8. Each sidebar counts as a unit, regardless of the number of paragraphs within it.
9. All units are numbered with whole numbers, starting with 1.
10. Numbering proceeds top to bottom, left to right.
11. Numbering proceeds in the order of sections demarcated by section headings.
12. All items of text, as described in steps 2–5, are numbered first, and without break, from the beginning of section to the end, taking the sections in order, including appendices.
13. All sidebars and figure captions are numbered next. The are the final paragraphs in the sections in which they are first mentioned in the text.
14. Supplementary methods sections or text documents should be included and unitized.
15. The sources/ bibliography *are not* units.

6 Operational Definitions for Analytical Construct

Here, I operationalize the concept of epistemic goals so that I can use content analysis methods to collect data about them. Specifically, I collect data about which goals are pursued in the three research projects I study. I based the operationalizations below on the definition of epistemic goals and the taxonomy of kinds of epistemic goals.

I provide four operationalizations in Table 3.3. All of them refer to codebooks, which are added to this dissertation in Appendix D. Each codebook lists a set of words

TABLE 3.3

OPERATIONALIZATION FOR CONCEPTS OF EPISTEMIC AIMS

AND TYPES OF EPISTEMIC AIMS

Operational Condition	Relevant Concepts and Conditions	Natural Warrant Form	Hypothetical Form	Notes
OK-1	None	Paragraphs that have instances of the terms and phrases from [Know] indicate the that the authors discuss knowledge in the paragraph.	If a paragraph describes a possible state of knowledge, then if we check the paragraph, we'll find instances in it of the terms and phrases in the [Know] list.	We can use OK-1 to see if authors use knowledge terms only when they aim to explain phenomena, or if they also use it for other types of goals.
OEG-1	Epistemic Goals, Condition 1	Paragraphs that have instances of the terms and phrases from [Goal] indicate the that the authors had or pursued a goal described in the paragraph.	If a paragraph describes a goal of project, then if we check the paragraph, we'll find instances in it of the terms and phrases in the [Goal] list.	We can use OEG-1 to identify all kind of goals, not just epistemic goals.
OEG-5	Epistemic Goals, Condition 5	Paragraphs that describe a goal and a product indicate that the product is relevant to the achievement or realization of the goal.	If a paragraph describes an epistemic goal of a project, then if we check it we'll find references to the product AND instances of the terms or phrases from at least one of [Know], [Goal], [Amalgam], [Control], [Describe], [Discover], [Explain] and [Cause], or [Predict].	I provide no codebook of terms for scientific products, and I leave to coders the evaluation of whether or not a product is mentioned in a paragraph.
OTEG-3	Type of Epistemic Goals, Conditions 1–3.	Paragraphs that have instances of the terms and phrases from the following lists indicate the that the authors had or pursued a specific kind of goal described in the paragraph. The lists are: [Amalgam], [Control], [Describe], [Discover], [Explain] and [Cause], and [Predict].	If a paragraph describes a specific type of epistemic goal of a project, then if we check the paragraph, we'll find in it instances of the terms or phrases from the lists: [Amalgam], [Control], [Describe], [Discover], [Explain] and [Cause], and [Predict].	

that connote a shared meaning. For instance, the codebook for the concept of general goal includes terms like ‘aim’, ‘end’, and ‘goal’, among many others.

The first operationalization is OK-1. It is an operationalization of a general concept of knowledge, but not of a more specific concept of epistemic goals. The next operationalization is OEG-1, which operationalizes the first condition of the concept of Epistemic Goal for a Research Project. The third operationalization is OEG-5, which operationalizes the fifth condition of the concept of Epistemic Goal for a Research Project. I don’t operationalize conditions 2,3,4, or 6 of the concept of Epistemic Goal for a Research Project. Those conditions provide routes by which, when operationalized, we might use them to provide (in)validating evidence for conclusions based on data drawn from the first two conditions. The final operationalization below is OTEG-3, which operationalizes condition three of the concept Type of Epistemic Goal.

I use those four operationalizations to collect data about the existence or pursuit of specific kinds of epistemic goals in the contexts of research projects.

7 Description of Sampling Strategy

I don’t use a sampling strategy for several reasons. First, my corpus is small, and the data I require can be collected quickly and efficiently. Second, I’m unsure that a larger corpus, which might require a sampling of lexical units, would help me address the descriptive question that guides this research. The data I seek are often so few and subtle that they’d likely be missed in large scale statistical analyses. Finally, larger corpora are good for identifying large trends, but for the fine grained analysis I pursue, it’s better to

bound phenomena in scientific research projects and then to study the documents in those contexts.

8 List of Sampled Units

The list of sampled units includes all paragraphs for all of the documents I study. On the raw data sheet, each paragraph is numbered and identified as a column. Each sheet in a workbook represents a different scientific publication, and each workbook represents a different research project.

9 Recording Instructions and Data Language

To record data, I employ the following instructions. I require no special data language.

1. Select a text and unitize it into paragraphs according to the scheme described above.
2. For the first paragraph of the text, and for the first list of words from the codebook, look for the first word from the list in the paragraph.
3. If you see an instance of the word:
    a. on the spreadsheet, find the row for the operational definition tied to the list of words on the codebook.
    b. find the column for the numbered paragraph.
    c. at the intersection of the two, list the word in the topmost empty cell.
    d. at the intersection of the two, list the containing sentence for the word in the next empty cell.

4. If you don't see an instance of the word, move on to the next word listed on the codebook and repeat the above steps.
5. Repeat the above steps for all of the words in the first list of words from the codebook.
6. If you don't see any words from the first list in the first paragraph, leave blank the intersection cells of the paragraph and the operational definition.
7. Repeat the above steps for all of the words in all of the lists in the codebook.
8. Repeat the above steps for all of the paragraphs of the text.
9. Repeat the above steps for all of the texts studied in the research project, with data for each text stored on an independent sheet in a spreadsheet workbook.
10. Review the raw data recorded. For each word and sentence recorded:
    a. Read the relevant paragraph, sentence, and word.
    b. Determine if the word is used to describe epistemic goals, or if it is used for some other function.
    c. If it is used for epistemic goals, mark 'yes' at the bottommost empty cell at the confluence of the numbered paragraph and operational definition.
    d. If not, mark a 'no'.
11. For each text, make a refined data sheet.
    a. Delete all of the data for the cells marked 'no'.

The last two steps are steps to clean the data and to separate noise from data. Coders should store both raw data files and scrubbed data files.

10 List of Reliability Tests/ Procedures

There are several ways in which researchers interpret the reliability of data. I follow Krippendorff, who argues that reproducibility is perhaps the most important interpretation. As such, I must develop and describe methods such that others can reproduce my data. That task is relatively easy, as I analyze all paragraphs in texts, and as I require no statistical analyses. For my corpus, if others follow the unitizing scheme and the recording instructions described above, they should be able to *exactly* recreate my raw data. I also use the detailed protocol in Appendix C.

To further indicate the replicability of my data, I also use computational tools. Given my corpus of unitized texts, I employ the instructions above within the software package WordSmith Tools (Scott 2012). That software enables researchers to quickly find words in texts, their containing sentences, and their paragraph numbers. The software automates steps 2 through 8 in the recording instructions. I use the software to ensure that I get the same raw data every time.

One issue remains. In the recording instructions, the final two steps indicate how a researcher should scrub the raw data to remove the noise from the information. To ensure that such steps aren't idiosyncratic, content analysts often use multiple coders to scrub the data, and they rate the inter-coder reliability. For this dissertation, only one researcher scrubs the data, and there are no inter-coder reliability scores.

The reliance on only one coder is a flaw, however minor, of the project. It is necessary for reasons of expediency. It is mitigated due to the perfect replicability of the rest of the data collection methods, and as the raw and scrubbed data are both published

as part of the project. Future work should aim to evaluate inter-coder reliability of the scrubbed data.

14 List of Procedures to Infer Answers to Questions

I use a set off inference procedures to infer answers to my research question from the collected data. The set of procedures is listed below. Other researchers may wish to categorize some of the procedures as further operational definitions of the concepts of Epistemic Goal for a Research project or of Type of Epistemic Goal. Furthermore, some may wish to use some of the inference procedures as tools to guide the scrubbing of data. I see nothing wrong with either tactic, but for reasons of clarity and brevity I don't pursue them here. The procedures are:

1. If a paragraph has at least one instance of terms or phrases from [Amalgam], then that paragraph indicates that one goal of the project was amalgamation of phenomena or theories.
2. If a paragraph has at least one instance of terms or phrases from [Control], then that paragraph indicates that one goal of the project was control of phenomena.
3. If a paragraph has at least one instance of terms or phrases from [Describe], then that paragraph indicates that one goal of the project was description or representation of phenomena.

4. If a paragraph has at least one instance of terms or phrases from [Discover], then that paragraph indicates that one goal of the project was discovery of new phenomena or theories.
5. If a paragraph has at least one instance of terms or phrases from [Explain], then that paragraph indicates that one goal of the project was explanation of phenomena or theories.
6. If a paragraph has at least one instance of terms or phrases from [Predict], then that paragraph indicates that one goal of the project was prediction of phenomena.
7. If a paper has more instances of terms or phrases from [Amalgam] than for any other type of epistemic goals, then the paper indicates that amalgamation of phenomena or theories was the primary goal of the project for the period of research described by the paper. The same holds for [Control], [Describe], [Discover], [Explain], and [Predict].
8. From the relative frequency in a paper of terms in [Amalgam], [Control], [Describe], [Discover], [Explain], and [Predict], we can infer the relative rank or importance of those kinds of epistemic goals of a project for the period of research described by the paper.
9. If the same paragraph has at least one instance of terms or phrases from [Know] AND from [Goal], then the paragraph indicates that knowledge was a goal of the project.

15 List of Procedures to Validate Truth of Inferred Claims

There are at least two primary kinds of validity that my project must satisfy. The first is construct validity. To show that my project meets construct validity, I must show that I use explicit concepts for my analytical constructs of research projects and epistemic goals. I must also show that I operationalized them in such a way so as to collect evidence about them. Most importantly, I must show that the data I collected provides evidence about specific cases of the explicit concepts, especially of epistemic goals, and not about closely related features.

Much of this chapter enables me to establish construct validity. I provide explicit concepts of research project, of epistemic goal, and of type of epistemic goal. Next, I operationalize the concepts of epistemic goal and type of epistemic goal so as to collect evidence about them from scientific articles. All that remains is to establish that the data collected provides evidence for the existence of epistemic goals, and not some similar feature, within research projects. The most likely similar feature is that of epistemic value or virtue.

To show that the feature of epistemic value differs from that of epistemic goal, I must provide concepts of both and show how they differ in contexts of research projects. I haven't done so. Therefore, others may justly question the construct validity of my results.

I mitigate the impact of such questions in at least two ways. First, I note that the project of explicating concept of epistemic goals falls within a larger project of explicating a model of local epistemologies of research projects, as depicted in Figure 2.1. Though I don't explicate the concept of values for research projects as part of this

dissertation, I must eventually do so. Second, I urge critics to look to my operationalizations and codebooks for epistemic goals. Critics should provide their concepts of epistemic values and show that my tools provide more evidence about those values than they do about my proposed concepts of epistemic aims.

There is a second form of validity that my project should meet. I label it corroborative validity. I must show that my results stand up to potentially disconfirming evidence. A common route is to analyze one set of documents, in this case scientific articles, and then analyze a somewhat independent set of documents, perhaps grant applications, and then compare the results. Such corroboration remains a future project.

3.6- Conclusion

This chapter provides a series of tools with which to study the epistemic aims of research projects. It provides a general concept of epistemic aims and a taxonomy of specific kinds of epistemic aims. It also provides many of the tools needed to collect evidence about epistemic aims with content analysis methods.

This chapter also describes 18 documents, which are themselves tools, needed for content analyses. Of special importance, this chapter details a novel model of scientific communication via research articles, and it provides operationalizations for the concepts of Epistemic Goals of Research Projects and Type of Epistemic Goals. For the design of content analysis, this chapter provides a scheme for unitizing texts studied, data-recording instructions, procedures for inferring descriptions of goals from data, and

reliability and validity procedures. Appendices C and D provide detailed protocols and codebooks for conducting the content analyses.

The next two chapters provide case descriptions, and they describe the results of the content analysis described above.

CHAPTER 4

CONTEXT FOR CASE DESCRIPTIONS

4.1- Introduction

Each of the next two chapters provides a description of a research project in which a team somehow uses gene regulatory models with evolutionary genetic models. This chapter introduces the structure of those descriptions, and it addresses some of the assumptions implicit in them. I discuss those topics here to make the descriptions shorter, unburdened by interpretative footnotes, and easier to read.

This brief chapter has several primary sections. After this introduction, I describe the aims and basic ontology presumed by the descriptions, and I preview the five sections of each description. In section 4.3 I discuss the general methods used to create the descriptions, and in the next section I describe some issues with collecting and refining the content analysis data. In 4.5, I discuss some of the limitations of the descriptions. I conclude with a section about three concepts I presuppose in my case descriptions: models, mechanisms, and research questions. Those topics are the subjects of robust debates, and many may wish to see how I approach those topics in my descriptions. The final section indicates how I do so.

4.2- Descriptions

Aims

In the two chapters, I aim only to describe the two cases. I don't aim to explain, predict, control or discover new phenomena. As such, with the descriptions, I aim primarily to answer the questions:

1. What models of evolutionary genetics do researchers use with gene regulatory network (GRN) models?
2. When researchers use those two kinds of models in tandem, what epistemic aims do they pursue?

To answer those questions, I focus on two cases for extended study. Each case is about a research project. Case study, however, isn't the only way to address those questions, and others may use different methods.

In Chapter 7 after the descriptions, I aim to do more than just escribe the cases. In that chapter, I comparing them to each other. I separate the steps of description and comparison for purposes of replicability. If people wish to replicate the methods and results of this dissertation, they should focus on one of the two steps at a time, because each step pursues different aims and employs different methods and logics to achieve those aims.

Ontology

The descriptions rely on an ontology developed in earlier chapters. That ontology includes at least research topics, research projects, and research teams. For this

dissertation, the research topic is the use by scientists of gene regulatory models with evolutionary genetics models. In Chapter 2 I inferred a set of research projects about that topic, and I represented those projects as networks of research papers connected by similar authors. Projects tend to span multiple years, and have multiple research publications. If I focus on a single publication, and the rationale of that publication, I call the object of that focus a *study*. In a sense, a research project decomposes into studies. For each project, its research team is the set of people who conducted the research project.

Sections

Each of the two case descriptions has five sections. The first section describes the research team, locating it in time and space, and each of its members. The second section describes the research project. It lists all the documents treated as outputs of the project, and which provide the basis for most of the information and textual analysis of the project. For another description of the same project, we might exclude some of those outputs or include others, so the list provides an explicit basis.

The second section also describes the type of population studied by the team, either simulated, laboratory, or wild (Winther et al. 2015). I use those three types of populations to partition into three parts the set of research projects in which researches use evolutionary genetic models with GRN models. However, for purposes of this dissertation, I partition the set into two subsets: projects about simulated populations; and populations about laboratory or wild populations.

The second section also describes many aspects of the project's rationale. It describes the primary phenomenon studied, the problems and questions pursued in the project, and the methods the team used.

The third section describes the models used in the project. A team may use many models, and many kinds of models in a research project. I aim to describe all of them, and to focus on evolutionary genetic models and GRNs. I classify some models as mechanistic. I further discuss my assumptions about models and mechanisms in a later section of this chapter.

The fourth section describes information related to the epistemic aims of the project. I provide two kinds of information about the aims of a project. The first is a collection of whole sentences that I judged as explicitly relevant to the aims of the project. The second is a set of tables that represent count-data for terms related to different epistemic aims. That latter information is the result of the content analysis described in previous chapters. I discuss it a bit further in a later section of this chapter.

The final section of each description summarizes the main conclusions from each study of the project.

4.3- General Methods

To develop the case descriptions, I used the following methods. First, I identified a population of cases and selected two for study according to the procedures described in Chapter 2 of this dissertation. Second, I drafted the case descriptions without the content

analyses. Third, I conducted the content analyses and summarized their results with tabulated information in the case descriptions.

The second and third steps decompose into many sub-steps. To ensure consistency in procedure across cases and that other could replicate my descriptions, I wrote, revised, and followed protocols for writing the case descriptions and for doing the content analyses. The protocols are Appendices B, C, and D to this dissertation.

For most sections of the case descriptions, I used methods familiar to historians and philosophers of science. For a given research project, and with a specific question in mind, I carefully read the documents that I had classed as outputs of the project. For instance, when I asked what were the problems pursued by the team, I read the documents and annotated the sections of text that were relevant to the question. From those annotations, I summarized my answer to the question, and in some cases I provided lists of sentences as evidence from the analyzed texts.

For other sections, I used methods that historians and philosopher may find unfamiliar, but that social scientists will find familiar. Content analysis is one such method, though I apply it only to find the epistemic aims pursued in the projects. Another is my use of protocols for constructing case descriptions. Such protocols enabled me to mechanize much of the data collection and description processes. Social scientists will recognize my use of protocols as an attempt to ensure the reliability of my methods and the reproducibility of the case descriptions. Some historians and philosophers struggle to, or don't care to, meet those criteria for objective results.

Similarly, historians and philosophers, but not social scientists, may find the structure of my case descriptions unfamiliar. Philosophers and historians often construct

case descriptions to tell historical narratives. Such narratives serve important functions (Beatty 2016). But not all products of historical research need be narratives, and my case descriptions are not narratives. While my descriptions don't tell stories, they enable comparisons across cases, which I pursue in Chapter 7 of this dissertation (Eisenhardt 1989; Eisenhardt 1991).

4.4- Content Analysis

A Note about Focus

I take content analysis to be a better method by which to interpret texts according to the following criteria: ability and ease for others to see and critique every step of the interpretation process, ability and ease for others to replicate the interpretations, explicit connection between interpretations and theories. If I aim to meet those criteria, and if content analysis is better than traditional readings, why do I use it to collect data about only the epistemic aims of the projects I study, and not about other facets, such as research problems, values, etc.?

Though I focus on some aspects of rationales and not on others, the difference in treatment is due to my effort to address this dissertation's specific research questions. The difference isn't due to some intrinsic centrality to the rationales of the features I focus on, nor is it due to some greater ease with which to use empirical methods to study epistemic aims.

Count Data

I report the results of my computational content analyses as count data. A count is the number of words of a given type used in a text. Count data indicate, but don't guarantee, the relative importance of different types of words in texts. Content analysts use count data to indicate the content of a text, especially implicit content in comparison to the explicitly declared content highlighted by authors.[1]

I employed nine types of words. Each type is a general concept, and each concept had its own list of words. Counts for 'Goal' and 'Know' describes the number of times the team used general goal terms or described states of knowledge, respectively. Counts for 'Cause' describe the number of causal words used by the team. All other counts describe the number of words used by the team to describe more specific kinds of goals: Amalgamating phenomena and theories, controlling phenomena, describing phenomena, discovering novel phenomena, explaining phenomena, or predicting phenomena.

The raw data are noisy for several reasons. First, the tools I developed, especially the word lists, are new and need further refinement. Second, natural languages, especially those employed by scientists, are complex, with many homonyms, and no content analysis tools can perfectly systematize such language.[2] As a result, the data includes many false positives in its counts, and it may exclude false negatives.

[1] For instance, an author may state that her text is about only urban topics. But if a count analysis indicates that the author uses more words associated with rural topics than with urban topics, the analysis provides interpreters with reasons to temper the author's explicit statement. In general, content analysists don't judge authors as maliciously hiding information, but simply as not highlighting all of the relevant content.

[2] The formal content analysis methods can't perfectly capture the shades of meaning found in natural language, it remains useful. We use them to make clear and explicit our interpretations of texts, and to show where they are fallible. The methods aren't perfect, but they're better than their alternatives.

I address those issues in several ways. To address potential problems with false negatives, I revised my tools and operational definitions three times throughout the course of the data collection, each time enabling the tools to collect a wider array of terms. To deal with false positives, I removed noisy terms from my word lists as I revised the lists. After each revision, I started the analysis from scratch.

Regardless of those revisions, many false positives remain in the data of Table 4.2. To filter out the remaining noise, I checked each datum and its surrounding sentence(s) to verify that it was relevant to the concept at hand. I discarded any single datum if it had at least one of the following issues:[3]

1. Infelicitous homonym
2. Infelicitous cognate
3. Metadiscourse
4. Describes phenomena, not project rationale
5. Describes layout of graphics
6. Describes research of other teams

For instance, though instances of 'present*' often provide some indication that the team aimed to Describe phenomena, I marked as noisy any instance in a phrase such a '…we present our earlier results in (Johnson and Porter 2000)…', as that instance was both an infelicitous homonym and an instance of Metadiscourse about papers. To ensure

[3] Due to the large amount of data related to the concept of Cause, I didn't clean much of that data. Had I made inferences from that data, I would have needed to clean it more thoroughly.

intracoder reliability, I annotated the data twice and unified my results, as described in the protocol for data collection.[4]

Finally, my codes, and the resultant counts, depend on a couple of assumptions. The first is that the causal terms aren't evidence that the team pursued explanation of phenomena over other aims. If the opposite of that assumption is true, then causal explanation is far and away the dominant aim of the projects. The second assumption is that 'Know' and 'Explain' are distinct concepts. Some use those terms interchangeably, or at least as more akin than I treat them here. I treat terms like 'understand' as providing data for knowledge claims, not for explanation claims, which some may disagree with. Different codebooks could yield different relative rankings for 'Explain'. The most I can do is to publish my codebook (Appendix D) along with my summary data for public criticism.

4.5- Limitations

Content Analyses

The conclusions from the content analysis have several limitations in terms of validity. First, they rely on data generated from only one kind of source: peer reviewed journal articles. Stronger and more well confirmed results would rely on data drawn from several different sources, including grant applications, grant reports, lab notebooks, correspondence between team members, interviews of team members, etc.

[4] My results would be strengthened by checking their reliability across multiple coders. That task must wait for future research.

Second, I eliminated data that seemed to be obvious noise. Stronger conclusions would come from a more precise protocol for identifying noise, from using multiple coders to eliminate noise, and to compare the results of their eliminations.

Third, I need stronger distinctions between epistemic goals and values. I deployed a theoretical concept and operational definition of 'epistemic goals' that partly function to distinguish such goals from values. But I didn't specify a concept and operational definition of 'value'. Stronger conclusions would result from specifying such tools, from using them to collect data about the values pursued in the project, and from comparing the conclusions about the values pursued against the conclusions of the epistemic goals pursued. Doing so would ensure that the concepts are different, and that the operational definitions collect different evidence in support of each concept.

Case Reports

Similarly, the overall case report has several limitations that future research could remedy. First, I used only two kinds of documents as sources of evidence for information about the JPT team and project: published articles and correspondence with the team members. If I'd collected information from more sources kinds of sources, especially structured interviews of the team members, the conclusions would be stronger. Especially weak is the section on research questions pursued by the team.

Second, while the case report describes many aspects of the project, I employed content analysis to collect data for only the sections about epistemic goals. A stronger case description might employ content analyses on other aspects of the project, especially

for the models used by the team, the research questions and problems pursued, and their primary results.[5]

Finally, the case reports are about projects conducted by researchers. A stronger validation process would involve the team members. In such a process, which would involve a semi-structured survey, the team members would review the case descriptions and evaluate their accuracy. I'd originally planned to conduct such surveys, but as my project proceeded, they became increasingly unfeasible. The next step to continue this project would be to conduct such surveys and revise the case descriptions in light of them.

4.6- Assumptions about Models, Mechanisms, and Research Questions

Models

There is a robust debate about how best to understand models as they are used in various fields (Frigg and Hartmann 2012, and references therein). Rather than getting sidetracked by issues in that debate, I posit simply that there are many kinds of models, that among those kinds there are models that represent the mechanical aspects of phenomena, that there are models that clearly do not, and that there is a continuum between the two kinds.

[5] Given the vast literatures on models, on mechanisms, and on questions, it would be difficult to provide operational definitions, for instance of models, that would capture all of the properties of those things. But the vast literature also provides many theoretical concepts from which to build different operational definitions. We might make progress in many of the debates about models, mechanisms, questions, etc., if we operationalized competing theoretical concepts and systematically studied research projects such that we could compare the strengths and weaknesses of competing concepts.

I assume that gene regulatory models represent mechanistic aspects of gene regulation to a good degree, and that evolutionary genetic models by and large don't represent mechanical aspects of evolution. In those assumptions I'm influenced by (Mathewson and Calcott 2011; Davidson 2006; Glymour 2006; Sober 1984; Lewontin 1974). I set aside issues of how best to characterize the representation relation between models and phenomena, and of how best to characterize the relations between models and theories, issues that are traditional topics among philosophers. I hold only that researchers can build, evaluate, and use models to help them achieve some epistemic goals, and that we can empirically investigate those behaviors and goals. I also sidestep issues about whether or not selection is a mechanism.

In this chapter, I use different evidence to support my descriptions of the models employed by research teams than I do to support my descriptions of the epistemic goals pursued by those teams. For epistemic goals, I provide a specific concept of epistemic goal, I operationalize it for use with content analysis, I collect data from texts about the goals pursed, and I reconstruct those goals from the data. For models, I assume two kinds of models and the features common to instances of those kinds, and I represent the models with citations to texts that indicate that the models had the features I represent. The latter method is considerably less systematic and reliable than is the former.

I use different methods for several reasons. First, for the cases I study, the models are more explicit and evident than are the epistemic goals, and there should be little debate about what parts of the projects are the models, or about the features of the models. Epistemic goals, on the other hand, are often tricky to pin down, and a claim about a specific goal for a specific project or team requires a higher degree of empirical

support. Second, my research questions in this dissertation center more on epistemic goals than on models, and so independent verification of my results should scrutinize my results about epistemic goals more heavily than my representations of models.

Projects similar to mine but with different questions and aims might use content analysis to investigate models much as I use it to investigate epistemic goals. I originally planned to do so in my project, and I sketched operationalized concepts of models and used them to collect preliminary data. But I ultimately found that, for my research questions, the data yielded no insights beyond the manifest presentations of models in the research teams' published articles.

Mechanisms

Much like for models more generally, there is a robust debate about how best to understand mechanisms and mechanistic models as they are used in various fields, especially for purposes of explaining phenomena (Anderson 2014a, 2014b, Craver and Tabery 2015; and references therein). Rather than getting sidetracked by issues in that debate, I aim in this chapter merely to describe models. I adopt the list of general features of mechanistic models from the irenic account of (Tabery 2004), and to describe a mechanistic model, I specify the details of those features for the given model. In this strategy, I'm influenced by (Mathewson and Calcott 2011). Though I do not do so here, one might profitably describe gene regulation processes using Salmon's account of conserved marks or quantities (Salmon 1984).

On the irenic account, a mechanism is composed of parts, activities of the parts, and interactions between parts or activities, all of which are organized so as to effect the

phenomenon of the mechanism. Parts are generally conceived as chunks of matter. Activities and interactions are less straightforward.

An activity is some behavior of a part or of a subset of parts within the mechanism. A part or unit has many activities, but not all of them are relevant to all of the mechanisms to which the unit is a part. Researchers evaluate models of mechanisms partly by how well those models represent the activities necessary to the causal operating of the mechanism, and by how well those mechanisms abstract away from unnecessary activities.

An interaction differs from an activity in that it is more explicitly a relation between two parts or processes, whereas activities needn't be. Furthermore, an interaction, which is localized to a spatial subregion of the mechanism, induces changes in the properties of at least one of the things interacting. Again, researchers evaluate models of mechanisms insofar as they represent only those interactions that are necessary to describe the causal operating of the mechanism.

Research Questions

Though I don't focus on a team's research questions to address the questions of this dissertation, I do discuss those questions in my case descriptions. I do so to help contextualize epistemic aims and models as elements of larger research rationales, which include research questions. Sometimes I find it useful to categorize those questions, and I explain that categorization here.

Biologists ask different kinds of questions for different kinds of population studied. There are at least two types of questions that map onto the three types of

populations (simulated, lab, and wild). The two types of questions are respectively about the general possibility of phenomena, and about the actuality of phenomena (Forber 2010). When researchers investigate simulated populations, they ask general possibility questions about those populations. When they investigate laboratory or wild populations, they ask actuality questions about those phenomena.

To answer general possibility questions, researchers construct models and study them with analytic or computer simulation techniques. They do so partly to determine the internal logical consistency of their models, and to determine the bounds and scope of their models.

To answer questions about actual phenomena, researchers collect evidence, either from the lab or from the field, about those phenomena. They do so to develop, confirm, or disconfirm competing theories, hypotheses, or claims about those phenomena. The first team I study, the Johnson Porter, Tulchinsky team, addresses general possibility questions. The second team, the Wray, Garfield, Runcie team, addresses questions about actual phenomena.

I develop my taxonomy of questions from work about related topics by Patrick Forber and Carl Craver. Neither focuses on research questions, but their analyses readily export to that topic. Forber focuses on evolutionary phenomena, on explanations, and he discusses three, not two, types of explanations: global how possibly explanations, local how possibly explanations, and how actually explanations. He maintains that researchers use different kinds of methods to provide those explanations. For global how possibly explanations, researchers simulate evolutionary phenomena under a variety of model parameters. They do so to determine the bounds of the model. For local how possibly,

they provide plausible just-so stories that *could* explain actual or specified evolutionary phenomenon. And for how actually explanations, they collect evidence to confirm or disconfirm competing just-so stories and to pick one as an actual causal explanation of an evolutionary phenomenon. See (Craver 2007) for a similar distinction between how possible, how plausible, and how actual explanations for *mechanistic* phenomena.

I generalize on Forber's and Craver's accounts in several ways. First, I focus on questions and not on explanations. So I posit how-possible questions, how-plausible questions, and how actual questions. The respective explanations would answer their respective type of question. Second, I posit that there are different types of legitimate questions pursued by researchers, types beyond the how-questions that researchers answer with causal explanations. I propose a taxonomy of general possibility questions, local possibility/plausibility questions, and actuality questions. Different questions within each modal type motivate different epistemic goals. For instance, a how-actually question aims for a causal explanation of a phenomenon, but a what-actually question aims for a description of a phenomenon. Third, while I here study projects that themselves study evolutionary or mechanistic phenomena, and thus fit nicely in Forber's and Craver's frameworks, I intend my taxonomy to apply to non-evolutionary or non-mechanistic phenomena.

A future project is to further develop this account of research questions.

CHAPTER 5

CASE 1: JOHNSON, PORTER, TULCHINSKY, AND SIMULATED POPULATIONS

5.1- Research Team

Time period and location of the research team

The research team coalesced in late 1998, published its first paper in 2000, and its most recent papers in 2014 (Johnson 2015, personal communication; Porter 2015, personal communication). It loosely persists at the writing of this description. The team started at, and is mostly based at, the University of Massachusetts, Amherst in Amherst, Massachusetts.

Members of the team

Member 1: Norman Anthony Johnson.

Johnson in one of the team's two founding members. He received his PhD in biology in 1992 from the University of Rochester in Rochester, New York. Chung-I Wu supervised his dissertation. Among other positions, in 1998 Johnson became an adjunct research professor at the University of Massachusetts, Amherst, in a department that later became the department of biology, a position he maintained throughout the period of the project and into 2015. Born in 1966, Johnson was thirty-two years old when he started the project, and he was a member of the project throughout. (Johnson 2015; Johnson 2007; Johnson 1992).

Member 2: Adam Hampton Porter.

Porter is one of the team's two founding members. He received his PhD in biology in 1989 from the University of California, Davis in Davis, California. Arthur Shapiro supervised the dissertation. By 1998, Porter was an associate professor at the University of Amherst, Massachusetts, in what later became the department of biology, a position he maintained throughout the period of the project and into 2016. Born in 1961, Porter was thirty-seven years old when he started the project of which he was a member throughout. (Porter 2015; Porter 1989).

Member 3: Alexander Y. Tulchinsky.

Tulchinsky received his PhD in biology in 2013 from the University of Massachusetts, Amherst. Adam Porter supervised his dissertation, and Norman Johnson was a member of the dissertation committee. Tulchinsky, born in 1982, was a graduate student and researcher on the team through at least 2015. (Tulchinsky 2013).

Member 4: Ward Belfield Watt.

Watt joined the team for a period between 2013 and 2014. He received his PhD in biology in 1967 from Yale University in New Haven, Connecticut. Charles L. Remington supervised the dissertation. While he was a member of the team, Watt was a professor at the University of South Carolina in Columbia, South Carolina. Born in 1940, Watt was 73 years old when he joined the team. (Watt 1967).

5.2- Research Project

Name of Project

For the purposes of this dissertation, I name the project currently being described as the JPT Project, after the last names of the three primary members. I also sometimes label the team as the JPT team.

Outputs of the Project

There are eight primary papers that comprise the output of this project. Five are reports of research, one is a review/ opinion article, one is a journal issue introduction, and one is a dissertation. In chronological order, they are:

Review Articles

1. "Toward a New Synthesis: Population Genetics and Evolutionary Developmental Biology," by Johnson and Porter (Johnson and Porter 2001).
2. "The Micro-evolution of Development," by Johnson (Johnson 2007). Journal issue introduction.

Research Reports

1. "Rapid Speciation via Parallel, Directional Selection on Regulatory Genetic Pathways," by Johnson and Porter (Johnson and Porter 2000).
2. "Speciation Despite Gene Flow When Developmental Pathways Evolve," by Porter and Johnson (Porter and Johnson 2002).

3. "Evolution of Branched Regulatory Genetic Pathways: Directional Selection on Pleiotropic Loci Accelerates Developmental System Drift," by Johnson and Porter (Johnson and Porter 2007).
4. "Hybrid Incompatibility Arises in a Sequence-Based Bioenergetic Model of Transcription Factor Binding," by Tulchinsky, Johnson, Watt, and Porter (Tulchinsky et al. 2014a).
5. "Hybrid Incompatibility Despite Pleiotropic Constraint in a Sequence-Based Bioenergetic Model of Transcription Factor Binding," by Tulchinsky, Johnson, and Porter (Tulchinsky et al. 2014b).

Dissertations

6. "Evolution of Hybrid Incompatibilities in Gene Regulatory Networks," by Tulchinsky (Tulchinsky 2013).

Type of Population Studied

Research teams in biology typically study one of three kinds of populations: simulated, laboratory, or wild (Winther et al. 2015). For its project, the JPT team studies simulated populations of individual organisms.

Research Rationale

According to the map of epistemic relations described earlier in the dissertation, I should be able to describe several aspects of the research project, including research problems, research questions, methods, primary phenomena of study, scientific products,

epistemic aims, and values. In the next section, I focus on the JPT project's scientific products and epistemic aims. In this subsection, I discuss the projects research problems, research questions, methods, and primary phenomena of study.

Primary Phenomena of Study

Throughout the project, the JPT team studied allopatric speciation, the process by which two populations, which share a common ancestor population, become reproductively isolated, such that individuals from the two populations don't interbreed. Those two populations evolve to become distinct species.

There are several routes by which one population can become two populations that are reproductively isolated. The JPT team focused on one such route called post-zygotic isolation. In post zygotic isolation, individuals from one species can mate with individuals from the other, and they can produce zygotes, but those hybrid zygotes are less fit than either the parents or non-hybrid zygotes. The more unfit the hybrids, the more likely the two populations become distinct species.

There are several possible causes that could contribute to instances of post-zygotic isolation, but evolutionary geneticists, including the JPT team, focus on problems in the genomes of hybrid zygotes, called hybrid dysfunctions or incompatibilities. In particular, they focus on hybrid dysfunctions that result from epistasis, in which multiple alleles at different genomic loci interact with each other to produce phenotypes. Evolutionary geneticists argue that in hybrids, the alleles interact with each other in ways that produce less fit phenotypes, compared to non-hybrids. Thus, hybrids on average die earlier than do non-hybrids, so they cannot reproduce to the same level as non-hybrids.

While evolutionary geneticists have developed mathematical models that describe epistasis in abstract, the JPT team proposed that gene regulatory models provided a potential empirical interpretation for epistasis (Johnson and Porter 2000, 528). The JPT team studied gene regulation, the process by which a gene produces a transcription factor protein or some other product, which then, either directly or indirectly (signaling), physically connect to the *cis* regions of other genes, enabling or inhibiting those genes as they produce their own products. Such interactions eventually enable cells to differentiate into specific kinds of cells, to move about within an organism and form tissues, and to produce phenotypes.

Research Problems

Early in their project, the JPT motivated their project with the problem that researchers have a broad goal to understand speciation, but especially for cases of allopatric speciation due to post zygotic isolation and hybrid dysfunction, that goal isn't met. Specifically, researchers little understood how populations evolved such that changes in genes resulted in hybrid dysfunction. Furthermore, while researchers acknowledged the role of epistasis in models of post zygotic evolution, they hadn't studied gene regulation as a possible physical interpretation of epistasis in those models (Johnson and Porter 2000, 528).

Later in their project, the problem that motivated the JPT team became more specific. In the work described in their first few papers, the team developed a model to ameliorate the problem described above, and they tweaked it to work in several different evolutionary scenarios. But the model had its own issues with empirical interpretation. As

described below, the original model employed a concept of binding strength, for which the team had no empirical interpretation, and which limited the empirical testability of their model. The team wanted a model that could be empirically tested, but the model that they built early in the project frustrated that aim (Tulchinsky 2013; Tulchinsky et al. 2014a). So while the team maintained the problem described in the previous paragraph, a more specific problem with their own model motivated their later work.

Research Questions

Except for (Tulchinsky 2014b), the JPT team doesn't use questions, either as logical or as rhetorical devices, in their research reports. Instead, they contextualize their project with research problems and epistemic goals. I don't infer, however, that research questions were absent from the project. If I had used other sources of evidence, such as grant applications, lab notebooks, or interviews, I might have found evidence for specific questions. As questions aren't central to this dissertation project, I didn't collect such evidence.

I provide the following questions as heuristic devices, to help readers better contextualize the JPT project within the map of epistemic relations provided earlier.

1. Johnson and Porter 2000.
    a. How might we model speciation via post zygotic isolation with an evolutionary genetic model that represents some aspects of gene regulation?
    b. Is that model good?

c. How does that model compare to a similar model that represents epistasis, but not gene regulation?

2. Porter and Johnson 2002.
    a. If we alter our model to capture gene flow between the two populations, how does that alteration affect the model's account of speciation?
3. Johnson and Porter 2007.
    a. If we alter our model to include more complex types of gene regulatory pathways for more than one phenotype, can the model also capture cases of developmental systems drift?
    b. If so, how does the evolution of two phenotypes, correlated by a common regulatory pathway, affect representations of speciation?
4. Tulchinsky et al. 2014a.
    a. How might we revise the earlier model to empirically interpret the concept of binding strength?
    b. Is the revised model good?
    c. How does the revised model compare to the previous model?
5. Tulchinsky et al. 2014b.
    a. "How might *cis*-by-*trans* divergence occur if adaptation in *trans* is constrained by pleiotropy?" (Tulchinsky et al. 2014, 1646).
    b. If we alter our revised model to include more complex types of gene regulatory pathways for more than one phenotype, how does the alteration affect representations of speciation?

The questions listed above for the JPT project are general possibility questions. The JPT team built a model to investigate evolutionary phenomena, and they studied it to determine the bounds of possible phenomena to which it could apply (Forber 2010).

General Methods

Throughout their project, the JPT team simulated evolution in populations of individual organisms. The team constructed four models. The code captured those models, and it enabled the team to rerun their simulations under many different starting conditions and parameters. Over the life of the project, the team ran thousands of simulations, changing the number of individuals in the populations, the number of alleles in the population or the number of genes in the GRN, the number of generations, of hybrids, the type of selection, and other aspects of their models (Table 5.1).

The JPT team studied the speciation via hybrid incompatibility, so they constructed or procedure to simulate the evolution of two populations that shared a common ancestor population, and they studied hybrids. To study the evolution of populations, the team labeled a replicate or a run as the complete period of evolution in two populations over the same number of generations. For each run, the two populations started out as copies of each other.

To study hybrid incompatibility, the JPT team constructed hybrids. To do so, they simulated interbreeding, without population mixing, between individuals of the two evolving populations within a run. Within a run, the JPT team generated F1 and F2 hybrids at multiple stages. The team described the fitnesses and phenotypes for the

individual hybrids, and they calculated the mean fitness and mean phenotypic value for the F1 and F2 populations. Eventually, they compared those mean values to the mean values of the evolving parental values. Throughout the project, each time the team generated hybrids, they created 50 F1s and 50 F2s.

Throughout the course of their project, the JPT team manipulated the parameter values of their model and initial simulation conditions. But they manipulated different values and conditions at different stages of the project. In (Johnson and Porter 2000), they manipulated the number of genes in a single regulatory pathway, the number of alleles for genes, the allelic mutation rate, the variability in the phenotypic values at the start of the simulation, the population size, and the number of runs. In (Porter and Johnson 2002), they manipulated the population size, the population structure or migration rate, the number of loci, the number of generations, the number of runs, and the percent change in optimal phenotype or speed of evolution. In (Johnson and Porter 2007), they manipulated the structure of gene regulations along with the number of loci, the pathways that faced different kinds of selection, the optimal phenotype, and the kind of selection studied: stabilizing, directional, and though it isn't selection, neutral evolution.

In (Tulchinsky et al. 2014a), the JPT team manipulated the kind of selection, the direction of selection, the phenotype values at the beginning of the simulations, population size, and the bioenergetics parameters of transcription factor binds to regulatory DNA, parameters that determines the relation between genotype and phenotype and fitness. In (Tulchinsky et al. 2014b), the team manipulated the kind of selection, the strength of selection, the population size, the size of binding structures,

Publication by year	Genetic system: loci # | allele #	Evolutionary processes studied besides hybrid incompatibility	Population size(s)	Number of generations	Number of runs	Aspects manipulated (roughly ranked in order of importance to the study)	Results from simulations on JPT's evo-gen model compared to: (roughly ranked in order of importance)
2000	2, 3 ,4 ,10 | 1, 2	Directional selection	250	2000	40–100	-loci # -allele # -amount of initial variation in populations -mutation rate -population size	1. analytic predictions from theory 2. each other across manipulations 3. results from simulations to traditional evo-gen (multiplicative) model
2002	2, 3, 4 | 2	Directional selection	25, 50, 100, 250, 500, 1000	500, 1000, 2000	150–500	-migration rate -loci # -population size -generation # -runs # -strength of selection	1. results from simulations to traditional evo-gen (multiplicative) model 2. each other across manipulations 3. (background or implicit comparison) analytic predictions from theory
2007	3 (pleiotropic/ branched), 4 (two distinct 2-loci pathways) | 2	-Directional selection -stabilizing selection -strong stabilizing selection -developmental systems drift	500	2000	1000	-pleiotropy or not -type of selection -optimal phenotype (loose binding or tight)	1. each other across manipulations, especially pleiotropy (branched regulatory pathways) to not (2 discrete pathways) 2. analytic (and *a priori*) predictions from theory
2014a	2 | 2	Directional selection (in multiple directions) -stabilizing selection -allelic dominance	25, 50, 100, 200, 400	4000, 40000	200	-Bioenergetic parameter values (ΔG_i, $E_{D@}$, N_{TF}) -population size -type of selection -direction of directional selection (value of optimal phenotype)	1. each other across manipulations 2. (loosely) to the results of the team's **2000** study
2014b	3 (pleiotropic/ branched), 4 (two distinct 2-loci pathways) | 2	-Directional selection (in multiple directions) -stabilizing selection -developmental systems drift	25, 50, 100, 200, 400	4000, 40000	200	-pleiotropy or not -Bioenergetic parameter values (ΔG_i, $E_{D@}$, N_{TF}) -mutation rate for all loci -mutation rate for *cis* regions -population size -type of selection -strength of selection -population size -number of bits on binding motif (12–24)	1. each other across manipulations, especially pleiotropy (branched regulatory pathways) to not (2 discrete pathways); and types of selection across different branches on a pleiotropic pathway. 2. to the results of the team's **2014a** study to establish 4-loci pathway as a control group 3. 2. (loosely) to the results of the team's **2007** study

Table 5.1. Simulations in the JPT project. The table above shows how the JPT team structured and varied their simulations over the course of their project. It also describes which aspects of their models they manipulated, and it lists the comparative methods they used to evaluate their evolutionary genetic model, but also indirectly their model cluster.

mutation rates for all loci, mutation rates for cis regulatory regions, and the bioenergetic parameters of tf/DNA binding.

5.3- Models

The JPT team used at least four distinct models in each stage of their project. The first modeled gene regulation, the second modeled natural selection on quantitative phenotypes, the third modeled speciation due to hybrid incompatibilities, and the fourth modeled reproduction and generational turnover. I describe each below.

Model of Gene Regulation

The JPT team developed a model of gene regulation, and then they tinkered with it over the span of their project. The model described the mechanical aspects of a system in which alleles yielded protein products, some of which bound to the regulatory regions of other genes. To better describe this model, I list the model's system, parts, the activities of the parts, and the interactions between the parts.

System: The explicit general system is a developmental pathway. Implicitly, the pathway is within the cells of organisms. Furthermore, a pathway may traverse cells within an organism. An explicit specific system includes at least two alleles at different loci and the regulatory connection between the output of the first allele and the functioning of the second.

Parts: The explicit parts of developmental pathways are alleles and proteins. Each allele itself has two parts, a *cis* regulatory promoter site, and a *trans* protein production site. There are two kinds of proteins, transcription factors (tfs) and generic, which may be tfs but needn't be. As the organisms are diploid, there are two alleles at a given locus on a chromosome, and for any given individual, the alleles may occur in more than one locus throughout the genome. Implicitly, the system includes all of the parts required to transcribe and translate the DNA within the alleles, to construct and fold proteins from polypeptide chain, and to move tfs through cells.

Activities of parts: In the developmental pathway models of the JPT team, the parts with activities are the *trans* regions of alleles and the proteins. The *trans* region produces tfs or other proteins, and the proteins and tfs move within or between the cells of individual organisms.

Interactions: In the developmental pathway models of the JPT team, the interactions are between the tfs from one allele and the *cis* regions of another allele, and between the *cis* and *trans* regions of single alleles. There are many more interactions implicit in the model, such as those required for protein production, folding, and transport, but the model abstracts away from them. For the first kind of interaction, a tf from one allele attaches to the *cis* region of another, such that the tf changes its property of moving about the cell and becomes affixed to the allele. For the second kind of interaction, the *cis* region, with an affixed tf, promotes the *trans* region of that same allele to produce the protein product of that *trans* region. The promotion depends on how well

the tf binds to the *cis* region, with stronger binds yielding increased promotions, and ultimately more protein produced from the *trans* region. For their early simulations, the JPT team employed a scheme to represent the binding strength between tfs and *cis* regions, discussed in Fig 5.1.

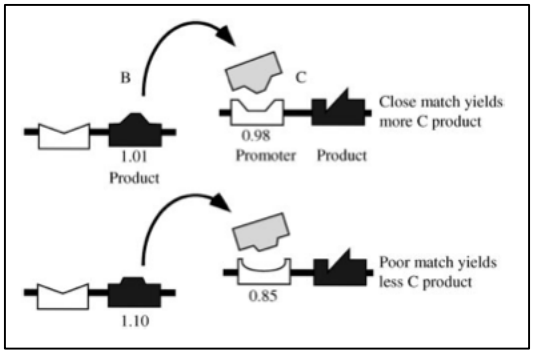

Fig 5.1: Gene regulatory pathway. The figure represents two variants on an interaction between two alleles. The alleles are labeled B and C. In the first and topmost variant, the first allele produces a tf with a shape represented by the value 1.01. That tf binds to the *cis* region of the second allele, which has a shape represented by the value 0.98. The difference in values is 0.03, indicating a tight bond, and thus inducing the second allele to produce a lot of its own protein. In the second and bottommost variant, the same interaction occurs, but the tf of the first allele has a shape represented by the value 1.10, and the *cis* region of the second allele has a shape indicated by the value 0.85. The difference in values is 0.25, indicating a looser bond compared to the topmost variant, and also inducing the second allele to produce less protein than did the topmost variant (Reprinted with permission from (Johnson and Porter 2000, 29)).

Throughout the course of their project, the JPT team studied different specific regulator pathways that employed the basic structure described above. In (Johnson and Porter 2000), the team studied pathways that were connected linear links of 2, 3, 4 or 10 alleles. They also studied pathways of 2, 3, 4, or 10 loci, such that each locus had two alleles, each susceptible to mutation and selection independently of its pair. Such pathways more accurately captured the genetics of diploid organisms, and they enabled

the evolution of dominance relations between loci. In (Porter and Johnson 2002), the team used pathways of 2,3, or 4 four loci. In (Johnson and Porter 2007), the team studied branched pathways of three loci, such that the product from the first locus regulated the *cis* regions of two distinct loci. The team compared evolution on those branched pathways to the evolution of two discrete pathways, each of which had two-loci.

By 2014, the team had revised its mechanistic model. The previous model had at least two major problems. The first was that the concept of binding strength, the central component of the model, referred to nothing in the pathways of actual organisms. The concept lacked an interpretation. The second problem was that the overall model enables evolution only in one direction with respect to the pathways. Such pathways could evolve only from tighter to looser connections between the tfs and the alleles they regulated. They could never evolve from looser to tighter connections. The first problem prevented laboratory and field biologists from applying the JPT team's evolutionary model, which I discuss in the next section, to actual populations of organisms. Even if they could apply that model, the second problem severely limited the range of phenomena they could use it to study.

Tulchinsky joined the team as a doctoral student and aimed to create a mechanistic model of regulatory pathways that would overcome those problems (Tulchinsky 2013; Tulchinsky et al. 2014a). Tulchinsky helped the team develop what they called a lock and key model of interactions between tfs and DNA. The model captured bioenergetics processes that occur when tfs bind to DNA sequences.

The basic mechanism is as follows. Tfs, once produced, exist in the cell and bind to DNA sequences in at least two ways. In the first, the pressure within the cells pushes

them to unbound sections of DNA, where they attach at random. Once attached, they slide over the strand much as a boxcar over a track, but are also able to jump from segment to segment. In the second, tfs sliding over DNA track stick or bind at specific spots, and no longer slide. In those spots, strong hydrogen bonds form between the base sequence of the protein and the nucleotides of the DNA sequence.

A tf sequence binds with different strengths to different sequences of DNA nucleotides, so the stronger a bind, the more likely the tf will stay at the site and perform a regulatory function. Researchers have developed methods to measure the free energy used to form bonds between tfs and DNA sequences. As a tf or a DNA sequence evolves over generations, the free energy used will increase or decrease on a continuous scale, enabling the evolution of stronger or weaker bonds, which enables the evolution of regulatory interactions between tfs and DNA.

Tulchinsky and the rest of the JPT team saw the above mechanism as something that could help them overcome the problems with their earlier model. The above model, though most well developed for prokaryotes, had been empirically established and could provide an interpretation for their abstract notion of binding strength. Furthermore, it enabled the evolution of stronger and weaker binding, not just weaker.

The team developed a heuristic model of a lock and key to represent binds between tfs and DNA sequences. A single tf molecule has a segment that has a sequence of 12 units or bits, and each bit can take the value of either 0 or 1. The same is true of *cis* regions. A tf maximally fits a *cis* region when all of the 0 bits of the tf align with all of the 1 bits of the *cis* region, when all of the 1 bits of the tf align with the 0 bits of the *cis* region, and when neither the tf nor the allele has bits that interpose between the bits in the

sequence of the otherwise exactly aligned molecules. A tf has the bit sequence encoded from its source allele.

The team heuristically visualized the two states in bit value as differences in length on the teeth on a key and in a lock, with keys representing tfs and locks representing *cis* regions. Bits with the value of 1 had long teeth, and those with the value of 0 had short teeth, as in Fig. 5.2 (Tulchinsky et al. 2014a). Long teeth best fit with shorts teeth from the interacting molecule, and vice versa. Keys that fit very well with locks used more free energy, and thus bound to each other with greater strength, than did lock and key combos with poor fits. Each tooth represents the additive effects of the sequences to the total amount of free energy used. The team formalized the above mechanism with several equations, describe in the next section.

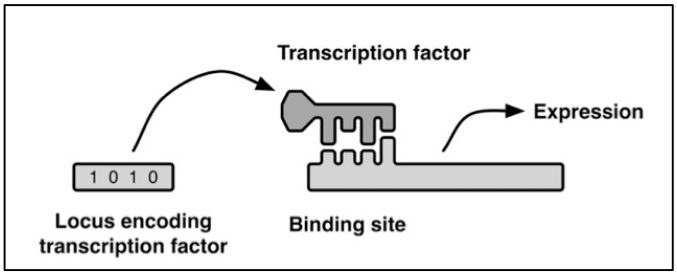

Fig 5.2: Example of a lock and key interaction as a regulatory pathway. The figure represents how a transcription factor (tf) with multiple bits of information binds to a *cis* region that also has multiple bits of information. The light grey rectangle on the left represents the *trans* region of an allele, and the numbers within the rectangle represent the value of the bits of information transcribed to its tf products. The dark gray "key" represents the transcription factor, each tooth represents a bit, and the length of a tooth represents its value, with long teeth encoded as a 1 and short teeth encoded as a 0 in the *trans* region of its source allele. The key fits onto a "lock", which is the *cis* region of the allele represented by the light gray rectangle-oid on the right side of the figure. The lock also has teeth, some that are long and some that are short. Keys fit best with locks when long teeth of one align with short teeth of the other. (Reprinted with permission from (Tulchinsky et al. 2014a)).

Given the language of parts, activities, and interactions, the new model differed from the old in several ways. First, for parts, the new model included alleles and proteins, but also specific nucleotides and tf bases. For activities, the new model implicitly assumed the range of behaviors performed by tfs as they travel from ribosomes to DNA sequences, and slide over DNA strands. For interactions, the new model explicitly focused on the use of free energy between tfs and DNA sequences as the tfs slid over the DNA or as they bound to specific sequences.

With that new model, the team focused on two kinds of genetic systems. The first system included two loci and one kind of tf. The first locus had two alleles, each of which produced its version of the tf. The second locus had two alleles, to which the tf bound, and which produced their own protein products (Tulchinsky et al. 2014a). The second system differed in that it included three loci, two of which were regulated by the tf produced by the first (Tulchinsky et al. 2014b)

Evolutionary Genetic Model

The JPT team developed a model of genetic evolution with their first publication, and they tinkered with that model as they studied different kinds of genetic systems, population structures, and regulatory mechanisms.

The model published in 2000 included four equations, all of which represent distributions assumed to be normal. The first two represent generally binding strength between loci, and more specifically between the tfs of one locus and the regulatory region(s) of other loci. For cases with two loci and haploid organisms (Fig. 5.1),

(a) $\beta(j, j+1) = \exp\{-[product\ (j) - reg(j+1)]^2\}$

where β represents binding strength or degree of match between the two loci, *j* is the locus that produces the tf, and *product*(*j*) is the allelic (shape) value of the tf produced by *j*. Similarly, *j*+1 is the regulated locus, and *reg*(*j* + 1) is the allelic value of the regulatory element. Both *product*(*j*) and *reg*(*j* + 1) are represented by decimal numbers, such that equal numbers represent perfect binding, and yield a maximum binding strength of 1.0. The better the fit, the more product is produced by (*j* + 1). Furthermore, those allelic values randomly mutated at a rate μ over generations of organisms.

For cases with polyploid organisms,

(b) $\beta(j, j+1) = \exp\{-\sum_c [product(j, c) - reg((j+1, c)]^2\}$

where the c represents the average of the shape values across the polyploid copies of the loci.

From those functions, the JPT team defined a phenotype as the amount of product produced by the final gene in a regulatory pathway, an amount compounded by interactions upstream in the pathway.

(c) $P = \prod_{j=1}^{n-1} \beta(j, j+1)$

such that *P* represents the phenotype value of the pathway, and the product operator multiples the binding strengths of all of the upstream interactions in the pathway.

From that function, the team developed a fitness function for changing environments as

(d) $W_i = exp\left[-\left(\frac{P_i - P_{opt}}{\Omega}\right)^2\right]$

where W_i is an individual's fitness, P_i is its phenotype, P_{opt} is the optimal phenotype for the environment, and Ω is a scaling constant that enables comparisons across runs in which the team manipulated the effects of mutation and the rates of change to P_{opt}.

The team used the same model in their 2002 report. But in addition to the random mutation rate μ, the team introduced allelic variation into populations of organisms via migration, represented as *m*, a decimal assigned to each organism that represented their probabilities of moving to the other population.

For their 2007 report, the team tweaked their model. They studied simulated pathways of three loci, in which the first loci regulated the latter two, which didn't regulate each other. Representing (j + 1) now as k, the team defined the amount of product made by k as

(e) $\text{expression}_k = \text{expression}_j * \text{binding}_{jk}$

and they fixed expression_j at 1.0 and didn't allow it to change in value. Given (e), the phenotypic value for each two-locus pathway became

(f) $P = \text{expression}_k$

and given the branched pathways, the fitness function became

(g) $W(P) = exp\left[-\left(\frac{P_i - P_{opt}}{2\Omega}\right)^2\right]$.

When the team changed their mechanical model to incorporate more bioenergetic properties, as first reported in 2014, they had to change their mathematical explication of the binding process. They replaced their abstract concept of binding strength with that of equilibrium fractional occupancy, θ, or how well a tf bound to a DNA sequence.

(h) $\theta = \frac{N_{TF}}{N_{TF} + e^{\Delta G_{TF} - \Delta G_b}}$

in which N_{TF} is the number of tfs in a cell, ΔG_{TF} is the free energy of binding between the tf and the site in question, ΔG_b is the background free energy of binding between all other sites and any tfs. Furthermore, with m as the number of mismatched bits between a tf and its binding site, ΔG_1 the free energy contribution of a single bit to an overall bind, and ΔG_{match} as the free energy of a perfect bind, the team could represent the effect of mutations to alleles on θ, such that $\Delta G_{TF} = \Delta G_{match} - m\Delta G_1$. They held the mutation rate at 0.001 per locus. Furthermore, they defined a parameter, E_{diff}, which captured the difference between background free energy and a perfectly bound tf, such that $E_{diff} = \Delta G_b - \Delta G_{match}$. They used the above two definitions to recast (h) in terms of mismatches between tfs and their binding sites.

(i) $\theta = \frac{N_{TF}}{N_{TF}+e^{-m\Delta G_1 - E_{diff}}}$

To account for dominance of one allele over another at a common locus, and other allelic interactions, the team scaled (i) so that they could use it to account for the number of tfs in a cell specific to a single allele.

(j) $\theta' = \frac{N_{TF}}{N_{TF}+\alpha e^{-m\Delta G_1 - E_{diff}}}$

The details of the scaling factor α aren't essential here for redescription, other than to note that the factor captures, for a competing allele, the number of its tf molecules, the number of mismatches between one of those molecules and its regulatory site, and E_{diff} for those molecules. The JPT team calculated θ' for each of the four kinds of allelic interactions, and they assumed that any two competing alleles produced the same amount of tfs. The team built (j) from (h) and (i), and they used only (j) for their simulations.

While the team still studied expression levels as the phenotypes of interest, they had to modify their mathematical treatment of those phenotypes. They defined the rate of expression, r, for an allele as the sum of its θ's multiplied by a constant k, such that $r = k\,\Sigma\theta'$. The final phenotype for a locus was the sum of the output of each of its alleles.

(k) $P = \Sigma r$

The team focused on tfs that were activators, and they scaled the *r*'s so that a perfect match between a tf and its target site yield a phenotype of $P = 1.0$.

For a fitness function, the team used equation (g), replacing the symbol Ω with the symbol σ^2_s to more standardly represent the variance of fitness around the optimal phenotypic value. For their second 2014 study, the team used equations (i), (j), and (k), and they altered their simulation assumptions.

Throughout their project, the JPT described their evolutionary models as sub collections of the above series of equations (a) through (k). We might instead say that equations (a), (b), (e), (h), (i), (j), and perhaps (c), (f), and (k) are instead mathematical explications or descriptions of the mechanical models described in the previous section. Researchers could, and do, use those equations to study gene regulation without studying the evolution of regulatory networks, so those equations aren't necessarily evolutionary. On this line of classification, only equations (d) and (g), as well is the hybrid incompatibility model described in the simulation structure, represent the actual evolutionary components of JPT's models.

I follow JPT, however, and describe all equations (a) through (k) as properly components of their evolutionary models. First, the fitness functions have no meaning without the phenotype functions, and the phenotype functions have no meaning without their foundational binding or bioenergetic models. For the whole evolutionary model to work, the equations travel as sets, though some of those equations have uses outside of evolutionary studies. Second, as the JPT team classifies sets of those equations as evolutionary models, they indicate the integrative, and not just complementary, flavor of their project.

Model of Speciation

The JPT team built their research around the Bateson-Dobzhansky-Muller (BDM) model of speciation (Coyne and Orr 2004, 269). I've already reviewed aspects of the model in earlier sections about the simulation structure. Here, I collect that information into a short description of the model.

The model represents how speciation occurs between two populations, which begin as one, but later become distinct (Fig. 5.3). The organisms in the root population share a genetic structure. For whatever reason, the two populations become distinct, and as the populations respond to selection pressures, the genetic structure, once common to all organisms in the two population, evolves into two distinct structures, one common to each population. The longer the two population evolve, the more their genetic structures become distinct.

To assess the distinctness of the two populations, or how 'speciated' they are at a given time, researchers study hybrids. For a given time, they a subset of organism from the two populations and interbreed them. They compare the fitnesses of the hybrids to those of the organisms in the distinct populations.

If the hybrids have the same average fitness as do the organisms of the distinct populations, then researchers conclude that the two populations, while distinct, aren't reproductively incompatible. If the two populations are considered distinct species due to their relative reproductive isolation, but if they cause of their isolation, say a glacier, were to disappear (melt), and if the two populations again intermixed, then we would reclassify them again as a common species.

If the hybrids are less fit than the organisms in the distinct populations, then their genetic structure forestalls them from, upon interbreeding, producing offspring likely to perpetuate the lineage. The less fit the hybrids, the more irrevocable the speciation.

The JPT team used the BDM model with a novel twist. Previous researchers deployed the BDM model mostly to study genetic structures of two loci and multiple alleles for each locus. While many used the model to study genes hypothesized to interact with each other in the development of phenotypes, the model doesn't require it. The JPT team used the model to study genetic structures with multiple loci and multiple alleles. Most importantly, while they also studied loci that interacted with each other, the team deployed explicit and mechanistic models of gene regulation for those interactions.

Throughout the JPT project, a "run" or a "replicate" had the same meaning. A single run involved the simulated evolution of two populations that began with similar features, evolved separately for a set number of generations, and for specific time slices produced first and second generation hybrids solely as a means to compare the fitnesses of the hybrids to those of the organisms in the evolving populations.

Model of Reproduction and Generational Turnover

The team never identified the above model as a model. But biologists often describe reproductions/generation representations as models, the team manipulated aspects of the representation (e.g. number of generations) just as they did aspects of the other models, and the representation fits the hallmarks of mechanistic models described earlier. As such, I describe the representation as a mechanical model with a system, parts, activities, and interactions.

System: The model represents a system of at least two discrete populations, such that one population descends from the other via organismal reproduction, and that all of the organisms in the parental population die before those in next generation mature to adults.

Parts: The parts of the systems fall into two levels. For the first level, the parts are the populations and gene pools. In the second level, the parts are the items that make up the populations. For the second level, those parts are individual hermaphroditic organisms, their ova, and their sperm. Each organism has a developmental pathway, the end molecular product of which provided the phenotype studied.

Activities: Adult organisms contribute gametes to a gene pool. Each adult provides sperm and two ova. Parents die after contributing gametes to an offspring pool, and couldn't contribute offspring to more than one generation. Some zygotes matured to be viable adults. If the number of zygotes exceeds the population cap, zygotes die at random. For adults, if the value of their phenotypes exceeds a number, selected anew and randomly every generation, they survive and contribute gametes to the gene pool; otherwise, they die (selection).

Interactions: In the gene pool, sperm randomly fertilize ova, yielding zygotes.

5.4- Epistemic Aims

This section describes the results for my content analysis of the JPT teams' published articles, for which I focused on the team's epistemic aims. I discuss three kinds of results below: sentences collected via open coding; counts of relevant words/ sentences, based on the general procedure and operational definitions developed in the previous chapter; and overall conclusions.

Open Coding

Using open coding, I collected the following statements about epistemic goals from the team's published papers. The quotes provide a (subjective) collection of sentences in which the JPT team explicitly mentions its epistemic goals. The quotes enable comparisons with the count data presented afterwards. I exclude analysis of (Johnson 2007) for lack of content.

Review Articles:

From (Johnson and Porter 2001), eleven passages:

> "Given the ubiquity of regulatory genetic pathways in developmental processes, we contend that study of the population genetics of these pathways should become a major research program" (45).
>
> "The synthesis of population genetics and development would form the basis of a micro-evolutionary theory of adaptation rooted in knowledge of how phenotypes are constructed from genotypes" (45).
>
> "The evodevo studies emphasize what happened deep in the history of life but not how these developmental systems continue to evolve. To accomplish the latter, we need to

incorporate the principles of population genetics into the study of regulatory pathways, and vice versa" (48).

"Given this evolutionary lability, the breeders' equation is too simple to predict long-term evolutionary change…. Ultimately, these predictive models should be grounded in a mechanistic understanding of how G-matrix components evolve" (49).

"Mechanisms are needed for representing how different DNA sequences might translate into phenotypic effects, through physiochemical rules that govern the binding among proteins and nucleic acids" (49).

"While the ambitious goal of a predictive model of phenotypic evolution is well beyond our current capacity, important limited progress can be made using very simple developmental systems" (49).

"Answering these questions, even in very simple developmental systems, will help us understand the developmental and physiological reasons why different traits vary together. This understanding, in turn, will give us the connection we need between evolutionary studies at the level of the gene (and genetic pathways) and the statistical language we use to study the complexity of evolutionary pressures on covarying traits in an ecological setting" (50).

"For these reasons, we believe that the investigation of speciation is also a logical place to build a bridge between population genetic and evolution-of-development studies" (52).

"…we are examining speciation as a possible consequence of micro-evolutionary forces operating on phenotypes determined by interactions of genes in a linear, regulated pathway (Johnson & Porter, 2000; A.H. Porter & N.A. Johnson, manuscript in review)" (52).

"A centerpiece of this synthesis of population genetics and development will be a mechanistic theory of adaptation, one rooted in what we know about how phenotypes arise from genotypes. Such a theory would simultaneously consider quantitative effects of allelic change and population processes" (55).

"On the empirical side, we believe a central goal of such a synthesis should be determining the extent of allelic variation existing in the parts of molecules that have regulatory functions" (55).

Compared to the other papers, the 2001 passages indicate a greater focus on synthesis as a goal of a research program, and that the synthesis will help researchers study complex phenomena across different levels of biological organizations, from genes to populations.

Research Reports:

From (Johnson and Porter 2000), two passages:

> "We explored the proposition that these [gene regulatory] pathways can provide a plausible source of the epistatic variation that has been implicated in the evolution of postzygotic reproductive isolation" (527).
>
> "We propose that regulated genetic pathways are a biologically realistic way to provide the complex epistatic gene interaction seen in empirical studies of hybrid fitness reduction. Here we investigate the plausibility of this proposition" (528).

The passages indicate that the team explicitly conceived of its project as of exploring plausible routes of speciation, and of the role of evolving gene regulatory pathways in those routes. The passages don't indicate, from the table of epistemic goals provided in earlier chapters, specific aims for the paper.

From (Porter and Johnson 2002), seven passages:

> "To study speciation, we use a new class of population genetic models that incorporate simple developmental genetic rules, likely present in all organisms, to construct the phenotype" (2103).
>
> "We believe the study of the evolutionary dynamics of regulated developmental pathways can provide important insights linking models of speciation with more general models of adaptation (Johnson and Porter 2001). Here we use these models to explore the extent to which speciation can occur despite gene flow" (2103).

> "Here we show that, given selection favoring identical but new phenotypes in two populations, postzygotic isolation can evolve even in the face of substantial rates of gene exchange between populations. This process occurs in simple Dobzhansky-Muller models with this type of selection, but more importantly, it also occurs in models of developmental genetic regulation, which have the fundamental epistatic properties of Dobzhansky-Muller models but capture more effectively the mechanisms of physiological gene action" (2104).

> "Developmental genetic models provide a plausible context for Dobzhansky-Muller incompatibilities to occur and bridge this gap to the microevolutionary models used by population and quantitative geneticists" (2104).

> "Our goal, however, was to study the effects of gene flow on speciation due to developmental genetic incompatibility alone" (2105).

> "We expect that developmental genetic models, with increasingly sophisticated mechanisms for translating genotype to phenotype and then to fitness, will continue to capture the fundamental dynamics of Dobzhansky-Muller speciation models (Johnson 2002)" (2110).

> "Developmental genetic models will be useful for thinking about how physiologically based systems of interacting genes are affected by microevolutionary processes that cause population differentiation and adaptation, as well as reproductive isolation" (2110).

The passages indicate that the team continued to take speciation as its primary kind of phenomenon, that evolution to gene regulatory pathways are important to that phenomenon, and that they studied the plausibility of those pathways to affect speciation. The passages indicate that the team aims at a dynamical description of speciation phenomena, but otherwise provides little information about other epistemic aims.

From (Johnson and Porter 2007b), four passages:

> "Here, we use individual-based simulations to study the evolution of traits controlled by branched developmental pathways involving three loci, where one locus regulates two different traits" (57).

> "Thus, a deeper understanding of the conditions under which such [molecular] interactions can diverge seems likely to be of great value in the study of speciation" (58).
>
> "Below, we will discuss how developmental system drift arises in branched pathways. We also find that many evolutionary properties (especially with respect to speciation) of these divergently branched pathways can be understood as extensions of what we have already learned from simple linear pathways" (59).
>
> "We use an individual-based simulation modeling approach to study the evolution of organisms with traits generated under these developmental rules [evolutionary genetic model]" (60).

The passages indicate that while the team still uses a framework of speciation, when they study developmental systems drift, they pay more attention than they previously had to the evolution in gene regulatory networks structures themselves. Again, the passages indicate that the team takes its tools and results to be useful, but the passages aren't explicit about the ends.

From (Tulchinksy et al 2014a), six passages:

> "We explore the evolutionary conditions that promote and constrain hybrid incompatibility in regulatory networks using a bioenergetic model (combining thermodynamics and kinetics) of transcriptional regulation, considering the bioenergetic basis of molecular interactions between transcription factors (TFs) and their binding sites" (1155).
>
> "The present model is a mechanistically explicit case of the Bateson–Dobzhansky–Muller model, connecting environmental selective pressure to hybrid incompatibility through the molecular mechanism of regulatory divergence. The bioenergetic parameters that determine expression represent measurable properties of transcriptional regulation, providing a predictive framework for empirical studies of how phenotypic evolution results in epistatic incompatibility at the molecular level in hybrids" (1155).
>
> "In this study, we investigate the effects of bioenergetic parameters on evolving genetic regulatory interactions and the evolutionary constraints imposed upon their by-product, BDM incompatibilities" (1156).

> "Our results overcome the limitations in the Johnson and Porter (2000, 2001, 2007) models and are well suited to empirical studies of the bioenergetic basis of gene expression…and bioinformatic data characterizing promoter sequences and TF binding …" (1156).
>
> "Thus, to understand at the molecular level how genetic incompatibility evolves between populations, we need a class of models that incorporate the relationship between genotype and phenotype in its bioenergetic context. To this end, we extended the gene-network speciation model of Johnson and Porter (2000, 2007) by incorporating an information-based statistical physics model of transcriptional regulation…" (1161).

The passages indicate that while the team still ultimately targets the phenomena of speciation via hybrid incompatibility, they update their model of gene regulation to be more molecularly mechanistic. They indicate that understanding is partly tied to the mechanisms and partly to the predictive framework that results from using it.

From (Tulchinky et al. 2014b), four passages.

> "We employed a mechanistically explicit bioenergetic model of gene expression wherein parameter combinations (number of transcription factor molecules, energetic properties of binding to the regulatory site, and genomic background size) determine the shape of the genotype–phenotype (G-P) map, and interacting allelic variants of mutable cis and trans sites determine the phenotype along that map" (1645).
>
> "The purpose of this study is to explore this apparent contradiction [*cis* regions as both a source and a constraint of evolution] from the perspective of the molecular basis of pleiotropic regulatory interactions, to determine the bioenergetic properties of these interactions that permit BDM incompatibilities to evolve despite evolutionary constraint" (1646).
>
> "In this study, we examine the conditions under which compensatory evolution between a TF and its binding site may facilitate cis-by-trans regulatory incompatibility despite pleiotropic constraint" (1646).
>
> "Here we examine the effect of the genotype-to-phenotype and phenotype-to-fitness functions on pleiotropically coregulated traits" (1647).

The passages indicate that the team connects its new model of bioenergetic gene regulation to their 2007 interests in developmental systems drift.

Discussion

From the open coding, several general themes emerge. The general phenomenon studied is speciation between two evolving populations. The cluster of models used to study this phenomenon provide a way to plausibly understand that phenomena. Among the cluster of models, the mechanistic model of gene regulation has a special status.

While the open coding indicates that the team aimed to understand the general phenomenon, it rarely revealed what that understanding consisted in. "Understanding", according to the team, has to be mechanistic in some sense. The team focuses on the aim of synthesis more in its opinion/review article, and hardly at all in its research articles. Otherwise, the open coding, and standard readings, doesn't reveal the specific types and instances of aims pursued by the team. For that, I turn to content analysis.

Counts

Table 5.2 collects the raw data for the count analysis. Each row represents one publication by the JPT team, and each column represents one of the nine concepts studied. The numbers in the cells represent, for the given article, the number of its word's—and by proxy the number of sentences—that relate to the relevant concept.[1]

[1] Counts for Cause don't function as proxies for sentences, and their interpretation raises difficulties as described in Chapter 4.

TABLE 5.2

RAW COUNT DATA

Paper	Goal	Know	Amalgam	Control	Describe	Discover	Explain	Prediction	Cause
Research Reports									
2000	36	9	10	11	30	15	10	30	236
2002	18	3	2	6	24	6	6	14	205
2007b	22	17	0	16	45	12	13	35	201
2014a	47	18	5	13	43	22	12	17	251
2014b	30	5	0	19	25	23	7	12	267
Review Papers									
2001	55	25	29	9	33	10	40	30	245
2007a	18	7	13	4	7	11	9	6	54

The data above are noisy for the reasons discussed in Chapter 4. Per the methods discussed in Chapter 4, I removed noisy false positives from Table 5.2 to yield informative data collected in Table 5.3.[2]

TABLE 5.3

INFORMATIVE COUNT DATA

Paper	Goal	Know	Amalgam	Control	Describe	Discover	Explain	Prediction	Cause
Research Reports									
2000	10	5	0	0	21	0	9	30	236
2002	5	3	1	1	13	0	4	14	205
2007b	5	16	0	0	25	5	10	35	211
2014a	7	3	3	0	30	1	7	16	261
2014b	10	2	0	2	14	0	6	12	278
Review Papers									
2001	26	20	29	0	22	1	27	29	245
2007a	3	4	9	0	1	0	0	2	53

[2] As many of the data for 'Control' related to gene regulation, I added those relevant data to the counts for 'Cause'.

Table 5.3 provides an exploratory snapshot, not a final story, about the kinds of epistemic goals pursued by the JPT team. The most we can use counts for is to indicate the relative importance of different kinds of goals within a given paper. We cannot use them, without translating them into percentages of words or sentences per paper, to study the evolution of those goals through the project.[3] The relative ranks are recorded in Table 5.4.

TABLE 5.4

RELATIVE RANKS OF KINDS OF EPISTEMIC GOALS

Research Reports	**Goal 1**	**Goal 2**	**Goal 3**	**Goal 4**
2000	Prediction	Describe	Explain	
2002	Prediction	Describe	Explain	Amalgam/Control
2007b	Prediction	Describe	Explain	Discover
2014a	Describe	Prediction	Explain	Amalgam
2014b	Describe	Prediction	Explain	Control
Other Papers				
2001	Tie: Amalgam/Prediction		Explain	Describe
2007a	Amalgam	Prediction	Describe	

Given Tables 5.3 and 5.4, I make some general notes. First, in the research papers, the goals of predicting, describing, and explaining phenomena were the primary kinds of goals pursued. Second, the goals of predicting and describing phenomena dominated the goal of explaining phenomena. Finally, the team only explicitly indicated its interest in some kind of synthesis in their non-research reports.

[3] That task is another that must wait for future research. Its results wouldn't be directly relevant to my project here.

Discussion

The content analysis results above complement the open coding results from earlier. The above results indicate that the JPT team pursued several consistent types of epistemic aims. Foremost among those types were the goals of predicting and describing phenomena. Less important, but still relevant, was the goal of explaining phenomena. Confirming the results of the open coding, the JPT team discussed aims of synthesis primarily in opinion/ review papers, and sparingly elsewhere.

Further work should be done to tie each of the models studied to the specific types of epistemic goals above. As of yet, I can't do so empirically, and so my analysis has the character of philosophical interpretation/analysis. That work oversteps the descriptive aims of this chapter, and I leave it to Chapter 7.

5.5- Primary Conclusions of the Project

The project as a whole had one primary result, but the stages of the project, represented by the distinct publications, each had a set of primary results. I review those results below. There are myriad results not reviewed here. The results below are often indicated in the research reports as the most important, and they most directly answer the questions heuristically posed in section 2.

For the project as a whole, the primary result was that it was possible, with increasing mechanistic detail of gene interactions, to model microevolutionary phenomena that led to speciation, a macroevolutionary event.

Johnson and Porter 2000

To evaluate their evolutionary genetic model, the JPT team simulated the evolution of populations in BDM speciation scenarios. They simulated thousands of such scenarios. They compared their results to two kinds of information.

First, they calculated from theory (including the model equations) and for different sets of starting assumptions, the percentage of runs that would result in speciation. Then they compared the percentages from their simulated runs to the percentages of predicted runs. Insofar as the simulated percentages matched the predicted percentages, they were in some sense good. (confirmed? plausible?)

Second, the team simulated the evolution of populations via a more traditional multiplicative model. In that model, the team calculated phenotypic values by multiplying the allelic values together, without any consideration for regulation between the alleles. They compared the percentages of runs that resulted in speciation across the traditional model and their model. Insofar as the percentages differed, the results indicated in which empirical situations to apply the traditional and JPT models.

There were four primary conclusions

1. Within an evolutionary genetic model, the team could represent gene regulation via the variable β for binding strength, which they calculated for individuals via equations (a) or (b), which they interpreted mechanistically via the structure depicted in Fig. 5.1, and which partly determined the phenotype and fitness of an individual. Importantly, phenotypes were properties of genetic systems, not of adult anatomy, morphology, behavior, physiology, or life history.

2. For the simulations, the JPT model yielded runs that speciated in the percentages predicted by theory, especially for runs with few (~2,000) generations.
3. For the simulations, the traditional model never yielded runs of few generations (~2,000) in which the populations speciated.
4. For BDM scenarios of speciation with few generations, traditional models can't capture those scenarios, while the JPT model could predict their likelihood, describe them, and causally explain them.

Porter and Johnson 2002

The JPT team modified their model to include a parameter for the rate of migration between the two populations evolving in a BDM scenario. To evaluate that model, they simulated hundreds of such scenarios under different initial conditions, calculated the hybrid fitnesses of the two populations at the end of each run, and averaged the results across runs with similar starting conditions. They compared the results yielded by simulations on their evolutionary genetic model to the results yielded by simulations on a traditional model.[4] The comparisons indicate which model applies to different kinds of phenomena.

[4] The team had shown in their previous study that the speciation rates between their evolutionary model and traditional models differed widely. To ensure that they could meaningfully compare the the models when they studied migration, they tweaked the values of parameters in the traditional model. Given those tweaks, the traditional model, without migration, yielded similar speciation rates to the evolutionary genetic model.

There were four primary conclusions.

1. For the traditional model, speciation occurs generally at the same rate regardless of population size, but it also decreases the higher the proportion of migrants across the two populations. While small populations with high migrant proportions couldn't speciate, larger populations with the same proportions could.
2. For their regulatory model in which one locus regulated another, the same proportion of migrants lessened the chance of speciation in small population compared to large, and the higher the proportion, the less chance of extinction for those small populations. As the migration rate increased and population size increased, the chance of speciation decreased. Similarly, as the strength of selection decreased while the migration rate increased, the chance of speciation decreased.
3. For multi-locus runs, regardless of migration rate, populations with organisms that had four loci regulatory pathways were less likely to speciate than similar populations with three loci pathways, and those populations were less likely to speciate than were populations with two loci pathways. Regardless, speciation still occurred at significant minority of runs for all populations.
4. Researchers should prefer the JPT evolutionary model to the traditional model when they observe speciation despite even small rates of gene flow between populations, or when they study speciation in populations due to selection in known gene-regulatory structures or pathways that yield quantitative, not qualitative, variation in phenotypes.

Johnson and Porter 2007

The JPT team modified their 2000 model to study gene regulatory pathways other than the linear one studied previously. They studied pathways in which the transcription fact from one locus regulated two otherwise distinct loci. To evaluate how well their model captured speciation due to evolution to those pathways, they simulated evolution in two diverging populations thousands of times. They calculated hybrid fitnesses and averaged them over runs with similar initial conditions.

The team compared their simulated results about hybrid fitness (speciation) between populations with branched regulatory pathways to:

1. Predictions from theory detailed in the 2000 paper and analytically applied to branched pathways.
2. Simulated results from similar populations with branched pathways, but that experienced different kinds of selection.
3. Simulated results from populations with four loci, in two distinct (unbranched) pathways.

There were four primary conclusions.

1. For two BDM populations with two independent gene regulatory pathways, and as predicted by theory, the populations didn't speciate when both pathways received stabilizing selection, they speciated in half of the runs when one pathway received directional selection and the other received stabilizing selection, and they speciated in 75% of the runs in which both pathways received directional selection.

2. For two BDM populations with branched regulatory pathways of three loci, and as predicted by theory, the populations didn't speciate when both branches received stabilizing selection, they speciated in half of the runs when one branch received directional selection and the other received either stabilizing selection or strong stabilizing selection.
3. Contrary to predictions from theory, the populations speciated in 55.4% of the runs in which both branches received directional selection, a result that was statistically significant within a 95% confidence interval.
4. In runs on populations that had branched gene regulatory pathways, developmental systems drift (DSD) resulted when one branch received directional selection and the other received either stabilizing or strong stabilizing selection.[5] DSD didn't evolve in runs for which both branches received stabilizing selection. For runs in which both branches received directional selection, DSD didn't result, but if researchers analyzed the data without knowing the genetic architecture of the organisms in the first generations of a run, as they often don't in field studies, they might conclude that DSD had occurred.

Tulchinsky et al. 2014a

The JPT modified their 2000 model to provide a more empirically tractable mechanism of gene regulation, and they updated their evolutionary genetic equations to reflect that mechanism. To evaluate how well their new model captured speciation, they

[5] Despite its name, developmental systems drift has little to do with drift in the sense of evolution by accumulated random genetic mutations. Rather, it describes the phenomenon in which a phenotype's underlying genetic architecture can evolve while the phenotypes itself remains fairly constant.

simulated evolution in two locus runs hundreds of time for thousands of generations. They compared their simulated results from one set of initial conditions to the simulated results of many other simulated conditions. They also compared their results to their 2000 results.[6]

There were six primary conclusions.

1. In an evolving population with two locus pathways, as the number of mismatches in bits between the *cis*-regulatory region and the transcription factor increased, the amount of final protein produced (phenotype) decreased, as did average fitness. For different values of the bioenergetic parameters across simulations, lineages with higher N_{TF} or E_{Diff} or lower ΔG_1 values yielded organisms that produced higher values of the phenotype, despite as many as eight mismatches, compared to lineages with lower N_{TF} or E_{Diff} or higher ΔG_1 values.
2. Regardless of bioenergetic parameter values, stabilizing selection led to hybrid incompatibility (and speciation) most when populations and the effects of mutations were small rather than large. In large populations, stabilizing selection eliminated small and most large mutations from the population, while in small populations stabilizing selection eliminated mostly large mutations.
3. For directional selection, the optimal phenotype in both populations evolving in parallel changed from high to intermediate expression, and thus from fits of tf to *cis* region with no mismatches to fits with some mismatches; or from intermediate

[6] While in previous parts of the project the JPT team studied hybrid incompatibility and speciation in F_1 hybrids, starting with their 2014a study, they focused on F_2 hybrids. Due to the altered mechanism and models, the F_2 hybrids more clearly revealed hybrid incompatibility than did F_1 hybrids.

to high expression, and thus from fits with some mismatches to fits with none. Hybrid incompatibility (speciation) resulted more often with populations evolved from high to intermediate expression than from intermediate to high expression.

4. Regardless of the direction of selection, directional selected speciated the populations quicker and more severely than did stabilizing selection. The more substitutions to the DNA sequence between the derived and ancestral populations, the more extreme the change in fitness and the stronger the incompatibility of hybrids.
5. By definition, one allele for the tf dominated the other when the first produced more tfs than did the second one. For directional selection simulations, as allelic dominance decreased and misregulation increased, F_1 hybrids exhibited hybrid incompatibility at the same rates as did F_2 hybrids. For stabilizing selection, decreases in dominance also decreased hybrid incompatibility in F_1 and F_2 hybrids. In all other cases, F_2 hybrids had more hybrid incompatibility than did F_1 hybrids.
6. Researchers should prefer their new model to their previous models not only because it was more empirically tractable, but also because it enabled researchers to study directional selection in more than one direction, whereas their previous enabled such studies for only one direction, from stronger to looser binding.

Tulchinksy et al. 2014b

The JPT team deployed their 2014a model to study the branched pathways they had studied in 2007. To evaluate how well their model captured the evolution of such

pathways, the simulated thousands of generations for hundreds of runs across many manipulations to the the values of the parameters in their evolutionary genetic model. They compared their simulated results for populations with 3-loci branched pathways to their simulated results for populations with two discrete 2-loci pathways, as a control. They also compared their simulated results for directional selection on traits to their simulated results for stabilizing selection on traits. They also compared their results to the simulated results from their previous 2014a study on 2-loci pathways.

There were five primary conclusions.

1. For simulations on populations in which organisms had two discrete 2-loci pathways, regardless of stabilizing or directional selection, the model yielded hybrid incompatibility (F_2 misregulation) in the same patterns as predicted by their 2014a results. Thus, the two discrete pathways (4-loci) genetic structure provided a good control for the other simulations.
2. For populations with 3-loci in branched pathways, in which the phenotype from one branch received directional selection and the phenotype from the other received stabilizing selection:
    a. directional selection on the phenotype of one branch resulted in less hybrid misregulation than directional selection on one phenotype in 4-loci structures, across all population size and bioenergetic parameters.
    b. across population sizes, the branch that received directional selection resulted in more hybrid misregulation in smaller, rather than larger populations. For different bioenergetic parameter values, the larger the

population, the more likely selection would eliminate organisms with unstable phenotypes likely to arise due to bioenergetic parameters that quickly led to misregulation.

c. across population sizes and bioenergetic parameter values, the branch that received stabilizing selection resulted in hybrid misregulation roughly in the same patterns as the branch that received directional selection.

d. but the branch that received directional selection resulted in more overall hybrid misregulation compared to the branch that received stabilizing selection, except in larger populations for bioenergetic parameters that lead to misregulation quicker. In those cases, hybrid incompatibilities were roughly the same across both branches, trending towards zero. In large populations, compared to smaller, selection eliminated organisms with mutations that quickly led to unfit misregulations.

3. Compared to 4-loci cases, in three loci cases, the branch that received directional selection led to less hybrid misregulation, with the difference increasing as the bioenergetic parameter values tended towards unstable mutations. The branch that received stabilizing selection resulted in more hybrid misregulation than did branches that received stabilizing selection in 4-loci cases.

4. If the researchers manipulated the mutation rate, smaller rates at all loci of, or just the *cis* region of, the directionally selected trait led to more hybrid misregulation than did higher rates. Similar manipulations on 4-loci cases or in cases of stabilizing selection yielded little variation is hybrid misregulation.

5. In populations with branched (pleiotropic) gene regulatory pathways, there is a "sweet spot" of conditions most likely to evolve hybrid incompatibilities. Those conditions are: smallish populations; weak but directional selection; moderate to high amounts of transcription factors (tfs) in the cell (dominance); moderately strong binds between tfs and *cis* regions; relatively long bit-lengths for binding motifs between tfs and *cis* regions; bioenergetic parameter values that yield moderate, and neither extreme, fitness landscapes and genotype-phenotype maps.

CHAPTER 6

CASE 2: WRAY, GARFIELD, RUNCIE, AND WILD-LABORATORY POPULATIONS

6.1- Research Team

Time period and location of research team

The research team coalesced in the late 2000s, published its first paper in 2009, and its most recent papers in 2013. The team no longer persists, as most of its members have scattered across the world to pursue other projects in other laboratories. The team started and existed in Gregory Wray's laboratory in the department of biology at Duke University in Durham, North Carolina.

Wray's lab was, and continues to be, a midsized lab with Wray as a PI, a lab manager, several postdocs, and several grad students. The team studied here comprised one subset of the overall lab for the period studied. As the lab pursued several projects, members worked on different projects, with many members spanning several projects. But not all lab members worked on all the lab's projects.

The team and project studied here is one of many one could study from Wray's lab. This project built on work completed in Wray's lab in the early 2000s (Ramano and Wray 2003; Wray et al. 2003; Balhoff and Wray 2005). It also prompted additional but later projects in Wray's lab (Wygoda et al. 2014; Israel et al. 2016). One could study a larger time period and set of team members from Wray's lab. Here, I focus only on the team that studied sea urchins and published papers between 2009 and 2013. Compared to Wray's other teams, only the one studied here published experiment articles and review

articles, studied both sea urchin development and explicitly studied sea urchin evolution within evolutionary genetic contexts, and featured a core team, the members of which little overlapped with sea urchin teams from earlier or from later in Wray's lab.

Members of the team

Member 1: Gregory Allan Wray

Wray is the principle investigator of his lab, the founding member of the team. He received his PhD in biology, supervised by David R. McClay, from Duke University in 1987 (Wray 1987). At the time when the team formed, Wray was a professor of biology at Duke, a position he maintained throughout the project. He was a member of the project throughout its existence. He (co)authored all six of the articles listed below.

Member 2: David Aaron Garfield

Garfield joined Wray's lab as a graduate student around 2008 after having been at Duke for a couple of years. He received his PhD, supervised by Wray, in biology from Duke in 2011 (Garfield 2011). Afterwards, he became a postdoctoral researcher at the European Molecular Biology Laboratory in Heidelberg, Germany, but he continued to publish papers with the team through 2013. In 2016, he became an assistant professor at the University of Berlin in Berlin, Germany. Born in 1980, Garfield was 28-years old when he joined the team. A member of the team throughout, he coauthored five of the six articles listed below. Two were short reviews, and three were research reports in which he was lead or second author.

Member 3: Daniel Erskine Runcie

Runcie joined Wray's lab as a graduate student after Garfield, also after having been at Duke for a couple of years. He received his PhD, supervised by Wray, in biology from Duke in 2012 (Runcie 2012). Afterwards, he became a postdoctoral researcher at the University of California, Davis, in in Davis, California, but he continued to publish papers with the team through 2013. In 2015, he became an assistant professor at UC Davis. Born in 1983, Runcie was 24-years-old when he joined Wray's lab. He coauthored two of the six articles described below, one as lead author and another as a second author on research reports.

Member 4: Ralph McMillan Haygood

Haygood joined Wray's lab in 2005 as a postdoctoral researcher. He received his PhD, supervised by Michael Turelli, in population biology in 2002 from the University of California Davis in 2002 (Haygood 2002). His postdoctoral position ended in 2009, after which he became an entrepreneur, but he continued to publish articles with the team through 2013. He coauthored all three of the research reports listed below, helping to design the experiments and analyze the data. While in Wray's lab, Haygood primarily focused on other projects related to primate evolution and not this one.

Member 5: Courtney Christine Babbitt

Babbitt joined Wray's lab in 2006 as a postdoctoral researcher. She received her PhD, supervised by Nipam Patel, in biology from the University of Chicago in Chicago, Illinois, in 2005 (Babbitt 2005). Her postdoctoral position ended in 2013, after which she

became an assistant professor at the University of Massachusetts in Amherst, Massachusetts. She coauthored two of research reports listed below, helping to analyze data. While in Wray's lab, Babbitt primarily focused on projects related to primate evolution.

Member 6: William J. Nielson

Nielson was the lab manager in Wray's lab. He co-authored two of the three research reports listed below, for which he helped with experimental procedures.

Member 7: Jennifer Wygoda Israel

Israel joined Wray's lab in 2011 as a graduate student. She received her PhD, supervised by Wray and David McClay, in biology from Duke in 2015 (Israel 2015). Afterwards, she became a postdoctoral researcher at the University of North Carolina in Chapel Hill, North Carolina. She coauthored one of the research papers listed below. After Garfield and Runcie left the lab, Israel's research built on their work with sea urchins, but she focused on different species.

Member 8: Sayan Mukherjee

Mukherjee was a professor in the department of statistics at Duke University, and he joined Wray's team to help with analyses in (Runcie et al. 2012). He received his PhD, supervised by Tomaso Poggio, from the Massachusetts Institute of Technology in Cambridge, Massachusetts in 2001. Hired at Duke as an assistant professor in 2004, he

was promoted to full professor in 2011. He co-authored one of the research papers listed below.

6.2- Research Project

Name of Project

For this dissertation, I name the project the WGR Project, after the last initial for the three primary researchers. I also label the team as the WGR Team.

Outputs of the Project

There are eight items that I take as the primary published outputs of the team. Three are review or commentary articles, three are research reports, and two are dissertations.[1]

Review Articles

1. "Comparative embryology without a microscope: using genomic approaches to understand the evolution of development," by Garfield and Wray (Garfield and Wray 2009).

[1] There is one further paper which looks as if it should be included in my analysis, but which I exclude. It is (Oliver et al. 2010), for which Garfield is second author and Wray and Haygood contributed. While the research is about sea urchin genetics and evolution, the study had differed enough from the others that it seemed not to fit. No one from Wray's lab is a corresponding author, the research is much more explicitly comparative to another species of sea urchins, and it focuses much less on gene regulatory networks and contexts than do the other studied. That said, it does anticipate several of the questions, phenomena, and methods pursued in later studies. I suspect that if I had included the study in the case, my analyses from it would have aligned with those from the other studies.

2. “The Evolution of Gene Regulatory Interactions,” by Garfield and Wray (Garfield and Wray 2010).
3. “Genomics and the Evolution of Phenotypic Traits,” by (Wray 2013).

Research Reports

4. “Population genetics of cis-regulatory sequences that operate during embryonic development in the sea urchin *Strongylocentrotus purpuratus*,” by Garfield, Haygood, Nielsen, and Wray (Garfield et al. 2012).
5. “Genetics of gene expression responses to temperature stress in a sea urchin gene network,” by Runcie, Garfield, Babbitt, Wygoda, Mukherjee, and Wray (Runcie et al. 2012).
6. “The Impact of Gene Expression Variation on the Robustness and Evolvability of a Developmental Gene Regulatory Network,” by Garfield, Runcie, Babbitt, Haygood, Nielsen, and Wray (Garfield et al. 2013).

Dissertations

7. “Selection and Constraint: Population Genetic Approaches to Understanding the Evolution of Sea Urchin Development,” by Garfield (Garfield 2011).
8. “Genetic and Environmental Constraints on Developmental Systems: Towards Predicting Genetic Responses to Climate Change in Sea Urchins,” by Runcie (Runcie 2012).

As there is much overlap between the published articles and the dissertations, I omit analysis of the dissertations to ease the overall analysis. For this case report, I focus on the first six publications listed above.

Type of Population Studied

Research teams in biology typically study one of three kinds of populations: simulated, laboratory, or wild (Winther et al. 2015). For its project, the WGR team studied a population of organisms that I count both as laboratory and as wild. The team studied animals collected from the wild, and they bred the animals in the laboratory for the purposes of their study.[2]

The team studied mostly purple sea urchins *Strongylocentrotus purpuratus*, which live and were collected from off the coast of California. Among other features discussed above and below, the different research reports are part of a shared project partly due to their focus on a common species.

Research Rationale

Given the model of research rationales described in earlier chapter, I here briefly describe several aspects of the WGR Project, including primary phenomena of study,

[2] For a longer project with more cases, I'd class the WGR Project as one that focused primarily on laboratory populations, and not on wild populations. But for the purposes of this dissertation, the WGR Project can stand for both laboratory and wild populations, given the constraints on their study that required laboratory manipulation, and on the scope of their inferences, which are to the wild population. This dual classification is possible due to sea urchins being both abundant in the wild and a model organism for lab studies.

research problems, research questions, and methods. In later sections, I detail the project's scientific products and epistemic aims.

Primary Phenomena of Study

The WGR team primarily studied two related phenomena. The first was variation in the DNA structure of regulatory regions of genes (Garfield et al. 2012). Traditionally, evolutionary geneticists aim to describe the kinds and amounts of genetic variation within populations of conspecific organisms, a phenomenon often called the genetic structure of a population (Roughgarden 1996; Gillespie 2004). For them, "genetic variation" means differences between organisms in the structure of their genes, either in the kinds of alleles at (homologous) chromosomal loci or in the DNA sequences that compose the alleles. So evolutionary geneticists often aim to describe what percentage of the organisms in a population have a given allele or set of alleles, or they describe how much the DNA structure of specific alleles varies across organisms. They often assume that their descriptions, while focused within species, are relevant to variation between species.

The second phenomena the team studied was gene expression, or the amount of RNA molecules produced from segments of DNA. The team measured and analyzed gene expression amounts in purple sea urchins, and they studied how the amounts of expression varied across individuals, across developmental stages, in isolated gene regulatory pathways, and across different kinds of environments.

This second phenomenon differs from the kind of phenomena often studied by evolutionary geneticists. The team was motivated by the hypothesis that much of phenotypic variation, on which natural selection acts, results not from variation to genetic

structure traditionally conceived, but on variation to the amounts of RNA produced from the genes in different developmental stages, developmental pathways, and environmental contexts. Focusing on expression levels, rather than merely on gene structure, enabled the team to study how genes regulate each other, and not just the genes themselves.

Research Problems

For research problems, the team often cited deficiencies in understanding as justifications for its research. For instance,

> "…we know a lot more about how individual genes and proteins evolve than we do about how the interactions between genes evolve, and even less about the effects of changes in regulatory interactions on organismal fitness" (Garfield and Wray 2010, 16).
>
> "…we understand relatively little about the evolutionary forces that shape the *cis*-regulatory elements underlying developmental gene expression" (Garfield et al. 2012, 152).
>
> "An outstanding challenge for both systems biology and evolutionary biology is understanding the molecular mechanisms that allow development to buffer phenotypes while retaining flexibility" (Garfield et al. 2013, 1).
>
> "The extent to which variation in molecular processes such as transcription, splicing, translation, and phosphorylation during early development affects later processes and, eventually, organismal phenotypes remains poorly understood" (Garfield et al. 2013, 1).
>
> "It remains unclear, however, whether…" [or not segregating alleles that influence morphology] "…are common within natural populations and how they are able to influence morphology despite buffering during development" (Garfield et al. 2013, 1).
>
> "We can't fully understand an evolutionary process until we observe it in nature. The casebook is now much larger, but other prominent gaps in our understanding of evolutionary processes in nature remain" (Wray 2013, 56).
>
> "One of the major challenges facing evolutionary genetics is understanding the degree to which interactions between genes and between gene and environment influence trait variation, divergence, adaptation, and speciation…" (Wray 2013, 64).

> "The extent of epistasis, or nonadditive interactions among genes, is not well understood, particularly in natural populations" (Wray 2013, 64).

Those problems are intellectual problems, ones in which knowledge is valued but currently lacking or non-existent. The team also highlights some practical difficulties that hinder research.

> "Traditional approaches to evolutionary genetics are often difficult to apply to natural populations…" (Wray 2013, 55).

> "Unfortunately, finding the causal gene, much less the causal mutation, is generally much harder than identifying a QTL for several reasons…" (Wray 2013, 58).

For one study, the team motivated their research to a larger problem related to ecology and climate change. They noted that the purple sea urchin is a keystone species in the Pacific shoreline region off the coast of North America, and that climate change is likely to raise that ocean's temperature several degrees in the next few decades.

> "While such an increase is unlikely to prevent development of embryos from all but perhaps the most southern populations, it will expose more populations to conditions that cause abnormal development…and create stressful conditions that affect dispersal and reproductive success" (Runcie et al. 2012, 4548).

One implication is, given their role as a keystone species, climate change could affect both the population of sea urchins and the overall ecosystem in which they live. If people could predict the effect of higher temperatures on developing sea urchins, they could also partially anticipate some of the effects of climate change.

Research Questions

The WGR team posed research questions, which differed in function across the two kinds of articles. In their review articles, they posed questions that could help orient and define a field in which researchers integrated tools from regulatory genetics with those from evolutionary genetics to study gene expression.[3] In their research articles, they posed questions that directed, or at least provided a success condition for, the research they conducted and reported.

Review Articles:

1. Garfield and Wray 2009
    a. "How does the genome-wide distribution of selection across development change when closely related species occupy very different habitats or differ markedly in their life history?" (Garfield and Wray 2009, 3).
    b. "Sampling a wider range of species comparisons may solve one of the oldest conundrums in evolutionary developmental biology: why development is so often conserved across vast phylogenetic gulfs and yet sometimes spectacularly diverged among closely related species." (Garfield and Wray 2009, 4).
2. Garfield and Wray 2010
    a. "How common is gene expression variation within and between species?" (Garfield and Wray 2010, 17).

[3] These questions might be described as the higher-level questions in Philip Kitcher's account of significance graphs (Kitcher 2001) or in Alan Love's account of problem agendas (Love 2008).

i. “How often do gene expression profiles differ between related species?” (Garfield and Wray 2010, 17).

ii. “How often are expression differences due to genetic versus nongenetic (i.e., environ- mental) factors?” (Garfield and Wray 2010, 17).

iii. “Do the expression patterns of other genes change at the same time as the gene of interest?” (Garfield and Wray 2010, 17).

iv. “Is there variation between individuals within a species?” (Garfield and Wray 2010, 17).

v. “How often do changes in gene expression evolve in general?” (Garfield and Wray 2010, 17).

vi. “Does this differ among different kinds of genes?” (Garfield and Wray 2010, 17).

b. “What types of genetic changes underlie changes in gene expression?” (Garfield and Wray 2010, 17).

c. “How does natural selection work to shape gene interactions?” (Garfield and Wray 2010, 17).

d. “What kinds of changes in gene interactions produce trait differences?” (Garfield and Wray 2010, 17).

3. Wray 2013.

a. “[W]hy specific mutations in particular genes are involved” [in trait variation]” (Wray 2013, 52).

b. “[H]ow these mutations affect phenotype” (Wray 2013, 52).

c. “[H]ow they [mutations] become established in natural populations” (Wray 2013, 52).

d. “Further, important parameters are sometimes the very items we most need to measure. Should a model of adaptation allow for new mutations or should it be based on standing variation?” (Wray 2013, 56).

e. “What is the effect size of mutations underlying trait evolution?” (Wray 2013, 60).

f. “Where does the genetic variation for adaptation come from?” (Wray 2013, 61).

g. “What is the dynamic of selection during adaptation?” (Wray 2013, 61).

h. “What is the role of hybridization in adaptation?” (Wray 2013, 61).

i. “What kinds of genes are involved in trait evolution?” (Wray 2013, 62).

j. “What kinds of mutations are involved in trait evolution?” (Wray 2013, 63).

k. “Is the genetic basis for adaptation and trait variation predictable?” (Wray 2013, 63).

l. “How extensive is epistasis?” (Wray 2013, 64).

m. “What is the role of cryptic genetic variation?” (Wray 2013, 65).

n. “How important are incompatible gene interactions in speciation?” (Wray 2013, 65).

o. “Do gene-by-environment interactions contribute to adaptation?” (Wray 2013, 65).

Research Reports:

4. Garfield et al. 2012. The team doesn’t use questions, even as rhetorical devices, to contextualize their results in this research report. I provide the questions below as heuristic devices to help my readers better understand the project. The questions are inferred from segments of the text (Garfield et al. 2012, 152–54).
    a. What is the nature of gene expression variation for eight *cis* regulatory regions within a population of sea urchins?
    b. What kinds of selection (positive, stabilizing, negative, etc.) affect those regions?
5. Runcie et al 2012.
    a. What are “the functional and evolutionary implications of environmental change in an important developmental model species?” (Runcie et al. 2012, 4549).
    b. “Here, we investigated the response of *S. purpuratus* embryos to a stressful, but realistic temperature range (12–18 °C), and asked if such temperature variation exposed evolutionary relevant GEIs [gene-environment interactions] by perturbing developmental gene regulatory networks.” (Runcie et al. 2012, 4548).

c. How "buffered" are interactions between genes in regulatory networks despite changes in environmental temperatures? (Runcie et al. 2012, 4549).

d. "Lastly, we tested for GEIs in this network across genetic backgrounds from a natural population, and asked if the response of the network to environmental stress was genetically variable." (Runcie et al. 2012, 4549).

6. Garfield et al. 2013.

a. How much variation in the expression of developmental regulatory genes exists within a natural population? (Garfield et al. 2013, 1).

b. What impact does this variation in gene expression have on downstream genes within a regulatory network? (Garfield et al. 2013, 1).

i. Are "r^2 values between directly interacting genes, on average, stronger than those between active genes with no known regulatory interactions?" (Garfield et al. 2013, 12).

ii. "How does the qualitative nature of regulatory interactions change over development?" (Garfield et al. 2013, 12).

iii. Is "the variation in gene expression encountered in nature buffered or propagated across a gene network during development?" (Garfield et al. 2013, 8–9).

c. [H]ow does expression variation during development influence the morphological phenotypes that lie at the interface between organism

> and environment and are therefore potential targets of natural selection? (Garfield et al. 2013, 1–2).

Methods

Throughout the project, the WGR team used methods commonly employed by developmental geneticists and developmental biologists, and another set of methods commonly employed by evolutionary geneticists.

For their first study, the team studied variation in eight genes across adult sea urchins (Garfield et al. 2012). They used standard developmental genetics methods. They collected tissue from the sea urchins and extracted genomic DNA. They isolated fragments of DNA and via polymerase chain reaction and amplified them via cloning. They used sequencers to describe nucleotide components of the fragments, and they aligned the sequences from different adults on computers with an alignment program. To identify regulatory regions they used transfection assays, to locate regulatory binding sites on genes they used protein–DNA binding assays.

The WGR team also used standard evolutionary genetics methods. To analyze the variation within genes across organisms, they calculated summary statistics and recombination parameters. Given descriptions of genetic variation, they assessed the impact of evolutionary forces on the population using evolutionary models or test statistics such as Tajima's D, Fu and Li's D, Fay and Wu's H, and analogs of the McDonald-Kreitman and Hudson-Kreitman-Aguade tests.

For the second study (Runcie et al. 2012), for methods from developmental genetics and biology, the team raised different organisms in controlled environments,

they fixed some embryos and extracted RNA from them, they photographed some embryos to measure some of their features, and from other embryos they extracted large amounts of RNA, which they sequenced. They also used an assay to target RNA from 72 genes in the gene regulatory network in the purple sea urchin that specifies endomesoderm.

From evolutionary genetics, the team began with a North Carolina II breeding design, a standardized design used to determine the genetic variation among groups or individuals. They began with the eggs from four females and the sperm from four males, they mated each female to each male, creating 16 groups of organisms. They split the fertilized eggs from each group into rearing dishes, each of which they maintained at either 12°, 15°, or 18° C, for a total of 48 distinct cohorts (not counting controls). The team created two replicates of the above design.

When the embryos gastrulated, the team extracted RNA from whole cohorts, not from individuals within the cohorts. After they identified the RNAs and measured their concentrations for each cohort, the WGR team studied the effects of temperature, genetic background, and paternal effects on gene expression levels. To quantify those effects, they developed a Bayesian model of expected effects, fit their data to it, and inferred how much of the variation came from expected effects, and how much came from the other effects. To study the impact of variation within endomesoderm network, they calculated correlation coefficients for expression levels on genes of known regulatory interactions, and for temperature and parental effects on gene expression variation they calculated Pearson correlations. And for temperature effects on spatial expression and skeletal development, they used ANOVA F-tests.

For the third study (Garfield et al. 2013), the team used many of the same methods as in the previous study. But from developmental biology they also studied the urchins at 7 developmental stages, not one, and they measured the features of the juvenile skeletons of the urchins. From evolutionary genetics, the team carefully calculated gene expression variances across cohorts and analyzed those variances into components of additive genetic variance, non-additive genetic variance, and environmental variance. They estimated additive genetic variance for the cohorts and compared those variances across cohorts and across developmental stages. To aid those estimations, the team created a quantitative genetics model akin to, but distinct from, the one created in the previous study.

6.3- Models

The WGR team used several different models throughout their project. I first describe those of their models about mechanisms of genes and gene regulatory networks (GRNs), and then I describe their models about evolutionary processes.

Models of Mechanisms

In its first published research study, the WGR team studied how the regulatory regions of eight genes varied, based on their DNA structure, across individual sea urchins (Garfield et al. 2012). The team didn't explicitly use a mechanistic model of any process. They did use a model that represented the structure of genes and their parts, which included regulatory regions, 5' untranslated regions, start codons, exons, and neutral

regions (Figure 6.1). The model assumed the causal capacities of those parts and of the overall gene, with respect to producing molecules, especially RNA, and ultimately affecting higher-order phenotypes.

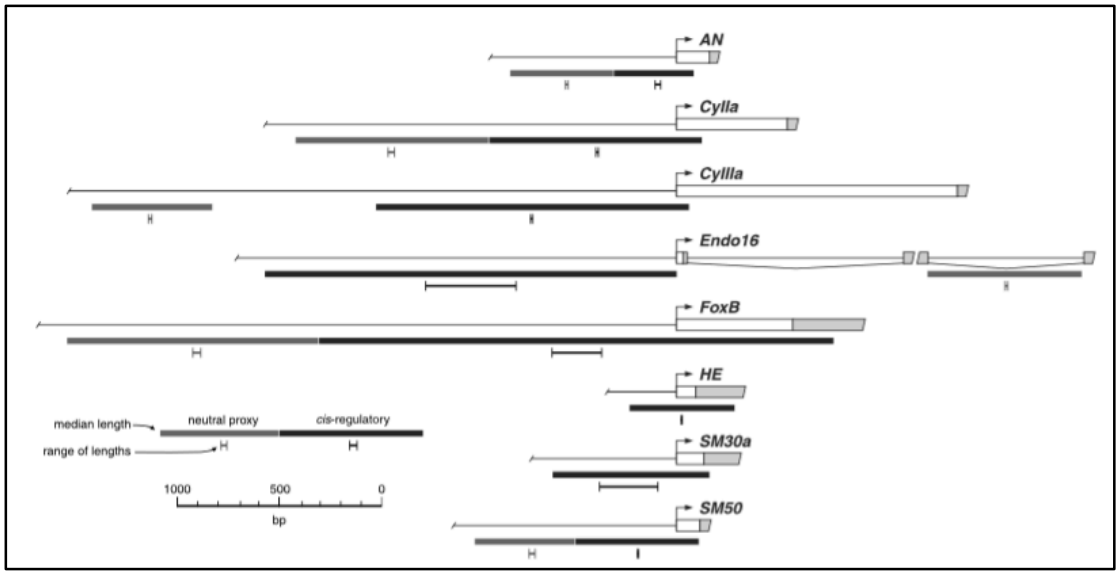

Fig. 6.1. Gene Structure Model. Each line represents a distinct gene. The bent arrows represent the start of the reading frames. White bars to the right of the arrows represent untranslated regions, and grey bars are exons. Known regulatory regions are represented to the left of the bent arrows. (Reprinted with permission from (Garfield et al. 2012)).

But for the second study published in 2012, the team studied genes and their RNA products in the contexts of GRNs and the phenotypes produced by those GRNs (Runcie et al. 2012). They assumed the model in Fig. 6.1 to assess how 14,454 genes varied in the amount of RNA they produced across individuals and breeding groups of sea urchins.

The WGR team next focused on 72 genes that partly comprise a GRN that produces the endomesoderm, ectoderm, and juvenile skeletons in early purple sea urchin embryos. The GRN was basically the one represented in Figure 6.2, with a few fewer genes. That GRN is a model that represents a mechanistic process, and though it abstracts

from many of the details of the physical interactions between the parts, it can still be explicated using the nomenclature of mechanisms.

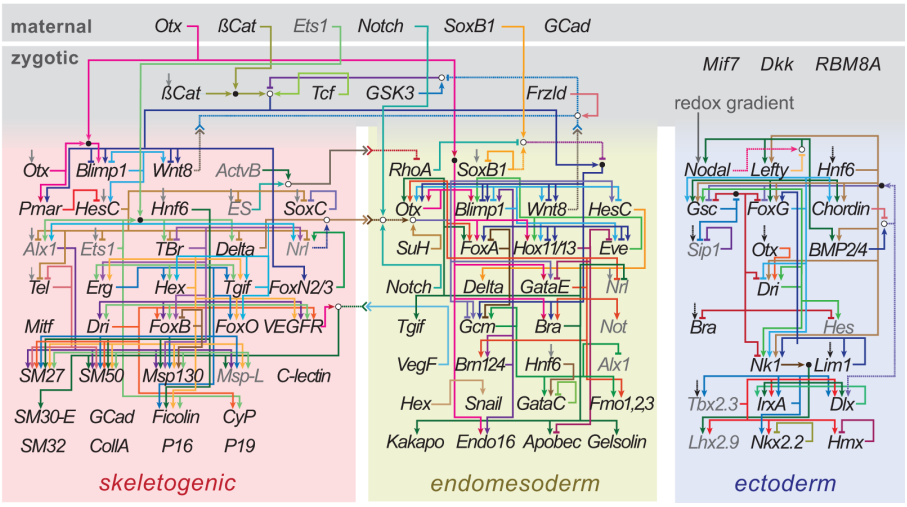

Fig. 6.2. GRN for Sea Urchin Development. The GRN represents the gene regulation that occurs in a sea urchin embryo from 10 hours after fertilization (top of model) to 90 hours (bottom of model) after fertilization. The boxes shows those key interactions in those cells that differentiation into juvenile skeleton cells (red box), endomesoderm (yellow box), and ectoderm (purple box). (Open access image reprinted with minor modifications from (Garfield et al. 2012)).

System: The system is a network of molecular pathways that span many cells in early sea urchin embryos. While the network abstractly represents many causal interactions, the result of those interactions is to produce and differentiate cells of three types: endomesoderm, ectoderm, and skeleton.

Parts: The parts of the systems include the molecules common to genetic and cellular phenomena. Explicitly, they include gene and gene products, especially transcription factor proteins. Implicitly, they include all the kinds of RNA, all the

transport molecules, and all the DNA regions of genes. The parts of the system also implicitly include the cells of early embryos, the materials and parts of those cells like cytoplasm, membranes, maternal proteins, etc.

Activities: In the GRN model, there are three main activities. First, genes produce products, generally transcription factor proteins. Second, those products move through the cells in which they were made, or outside of those cells into the extracellular matrix. Third, cells differentiate into endomesoderm, ectoderm, or skeleton cells. Each of those activities, however, could be the focus of further mechanistic description in terms of systems, parts, activities, and interactions.

Interactions: Explicitly, the GRN model represents two kinds of interactions. In both, the product from one gene binds to the regulatory region of another gene. In the first kind of interaction, the bind forestalls the second gene from producing its own product. In the second kind of interaction, the bind activates or promotes the second gene to produce its own protein. Abstracted from the model, though still implicit within it, are dozens of further kinds of molecular interactions that enable the first two kinds.

For its third published study, the team again used the GRN in Figure 6.2 (Garfield et al. 2013). The 2013 study used basically the same GRN as the one used in the previous study, with a few more genes and a bit more detail the number of interactions studied.

Evolutionary Genetics Models

The WGR team began the project by using a standard set of evolutionary genetic test statistics, but by the end of the project, it had started to develop novel quantitative genetic models.

Garfield et al. 2012

In its first study, the team studied, as many evolutionary geneticists do, the amount of variation in the DNA structures of genes (Garfield et al. 2012). Unlike more traditional studies, the team didn't study variation in the structures of those regions of genes that code for proteins. Instead, they studied variation in the regulatory regions of genes, variation that Wray and others hypothesized to be the genetic source of much variation in phenotypes. Within regulatory regions, the team looked for different kinds of variation, and it looked for differences in variation between binding regions and neutral regions. Once they had measurements for all those variations, the team applied statistical tests to infer what kinds of evolutionary forces might be affected the regulatory regions of the genes. They applied the models to each of the eight genes studied, to the *cis* regulatory regions and to the proxy neutral regions of those eight genes.

To study genetic variation, the team used or created the following data models. The first two are standard models to study variation between nucleotide sequences across individual organisms. The first is π. Start with a population of organisms, and for a given gene common to all of those organisms, describe the nucleotide sequence of each individual's alleles. Next, compare each allele sequence to every other allele sequence. For any two sequences, count the number of differences in nucleotides. That difference is

called the pairwise difference. Calculate each unique pairwise difference, and then sum the differences. The average of the summed pairwise differences is labeled π. To estimate π for a population from a sample, the WGR team used π^n, such that

(a) $\pi^n = \frac{\sum_{i<j} d_{i,j}}{n(n-1)/2}$

where each $d_{i,j}$ represents a single pairwise difference.

The second is θ, which is the π we would expect of a population of organisms if that population evolved only neutrally, and not due to any selective forces. Furthermore, θ represents the diversity in the alleles across a total population of organisms, which for diploid organisms is

(b) $\theta = 4N\mu$

such that N represents the effective population size and μ represents the neutral mutation rate.

When studying samples of populations, researchers estimate θ with θ^W, which is a function on the number of nucleotide sites that differ across sequences of the same allele, called segregating sites, be those sequences from different chromosomes on the same locus within an individual organism, or from different organisms altogether. So

(c) $\theta^W = S/\sum_{i=1}^{n-1} 1/i$

such that S is the number of segregating sites and n is the number of nucleotides in an allele.

The team also used a third model. While π^n represents average estimated pairwise variation, and θ^W captures the pairwise variation we would expect if the sequences came from a neutrally evolving population, the two statistics don't capture everything the WGR team wished to study about sequence variation. To compute those values, researchers must compare instances of alleles to each other, noting any differences in nucleotides across the two instances. Such comparisons detect some kinds of mutations, especially single nucleotide polymorphisms, but they falter when mutations result in major insertions or deletions of nucleotides into allele sequences. In an earlier paper, Wray and Jim Balhoff had developed a statistic to represent so called length variation, a statistic they called π^i (Balhoff and Wray 2005). The WGR team employed it in this study.

To compare DNA sequences from different organisms, researchers display those sequences one on top of the next in computer programs, and then they align the sequences so the similar parts of the sequences are roughly on top of each other. To get the sequences to align, researchers may introduce gaps into one or more of the sequence representations. Across two sequences, a gap represents either that the allele with the gap lost, by mutation, a section of the sequence it should have, or that the allele without the cap gained, by mutation, a section of sequence that alleles from other individuals lack. Mutations of the first kind are called insertions, those of the last kind are called deletions, and together they are called indels.

When calculating the value of π^i for a population, Wray's teams counted the number of gaps/indels, and they weighted, via natural logarithms, each indel by the length of the sequence deleted or added. Ultimately, π^i represented the average weighted variation in indels across all sequences for an allele. The team also calculated the relative incidence of different kinds of length mutations, information not captured in π^i.

Finally, the team also computed ρ, which estimated the amount of sequence divergence between alleles of the purple sea urchin and those for a similar locus from fragile pink sea urchins, a distinct species.[4]

To study selection, the team used five test statistics. The first was Tajima's D, which starts with the π and subtracts θ. The result indicates whether or not the gene studied faces selection forces in the population studied. To estimate D, researchers substitute π^n for π, θ^W for θ, and divides the result by the square root of the sampling variance of the two estimates.[5]

(d) $$D = \frac{\pi^n - \theta^W}{\sqrt{V}}$$

If the result is not significantly different from zero, then the gene faces no selection in the population studied, and evolves only neutrally. If the result is negative, then there is less variation in the population than we'd expect, indicating that the gene faces negative or purifying selection. If the result is positive, then the gene is more variable in structure

[4] It should also be noted that the WGR team used correlational analysis to establish that π^n positively correlated with π^i and with ρ.

[5] The decomposition of √V is beyond the scope of this chapter, so for details see (Tajima 1989).

than we'd expect, indicating that the gene faces positive selection, perhaps through balancing selection.

The next two statistics are similar to Tajima's D, but they estimate θ with information other than that captured by π or θ^W, and they enable researchers to compare data about a population to that from outgroups. For outgroups, the WGR team used DNA sequence data for orthologous genes from fragile pink sea urchins, from red sea urchins, and from green sea urchins. The second statistic is Fu and Li's D,

(e) $D = \frac{S - a_n \eta_e}{\sqrt{V(S - a_n \eta_e)}}$

such that S represents the number of segregating sites, a_n represents $\sum_{i=1}^{n-1} \frac{1}{i}$, and η_e represents the difference in the number of mutations, derived only once, between an ingroup population and an outgroup population.

The third statistic used by the WGR team was Fay and Wu's H,

(f) $\theta_H = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} i^2 \xi_i$

for which n represents the number of sequences, and ξ_i represents the number of times i that researchers detected mutation unique to the ingroup population, but within a gene shared with the outgroup.

The WGR team used Tajima's D, Fu and Li's D, and Fay and Wu's H partly because the tests are sensitive to different kinds of information. Similar to the use of Tajima's D, researchers interpret different kinds of selection from the signed results of Fu

and Li's D and Fay and Wu's H, but the latter two are more sensitive to selective sweeps. The WGR team established the significance of their results using community-standard techniques of repeated simulation of their data hundreds of thousands of times in a neutral model framework.

The WGR team used two final statistics, which also enabled them to compare their data to similar information from outgroup species. For seven of the genes studied, they used a McDonald-Kreitman test and a Hudson-Kreitman-Aguade test, both modified so that the team could apply them to *cis* regulatory regions.

The McDonald Kreitman test applies to the protein coding parts of genes, called exons, which decompose into three-nucleotide-long sets of DNA that, after transcription, code for amino acids and enable cells to collect those amino acids into polypeptide chains. Some exons are synonymous, or code for the same amino acid. A mutation to an exon changes the sequence, which becomes either synonymous with its non-mutated sister exons related by descent, or it becomes non-synonymous with the those sister exons and now codes for different amino acid.

To do MK test, researchers count the number of mutations for a set of exons within a population of organisms, they count the mutations for that set of exons between the population and an outgroup population, and they separate each count into synonymous mutations and nonsynonymous mutations. Finally, they take the ratio of nonsynonymous to synonymous mutations between the population and outgroup (or between species), and they take the same ratio for mutations within the population. If the two ratios are equal, then the evolution between species is neutral. If the ratio for between species is significantly higher than that within species, then evolution is positive between

species. If the ratio between species is significantly lower than the ratio within species, then selection is negative between the two species.

(g)

	Between Populations	Within population
Non synonymous mutations	A	C
Synonymous mutations	B	D

The WGR team modified the MK test to apply to binding sites in *cis* regulatory regions rather than to exons. To do so, they classified mutations to binding sites as either permitting binding or as not, following a strategy from Jenkins (Jenkins et al. 1995).

	Between Populations	Within population
Mutations that prevent TFs from binding to binding site	A	C
Mutations that preserve or enable TFs to bind to binding sites.	B	D

(h)

The WGR team also used a modified Hudson-Kreitman-Aguade test. Similar to the MK test, the HKA test compares data from an in group to that of an outgroup to infer deviations from neutral evolution. The test is a variant of a χ^2 test. The WGR team admitted that the HKA test for *cis* regulatory regions lacked realism or meaning, and all of their analyses returned either nonsignificant or inapplicable results. Given as much, I avoid detailing the model here.

Runcie et al. 2012

In its second research report, the WGR team describes a novel evolutionary genetic model it created to analyze its data and infer additive genetic variance.[6] In matrix format, the team created a model to capture the effects of genetic background, parental effects, and temperature on the measured gene expression levels for 73 genes. The model is:

(i) $\mathbf{y} = \mathbf{X}_Y \begin{pmatrix} \mu \\ \mathbf{u} \end{pmatrix} + \mathbf{E}$,

(j) $\mathbf{u} = \mathbf{X}_U\mathbf{b} + \mathbf{Z}_M\mathbf{a} + \mathbf{Z}_F\mathbf{f} + \mathbf{Z}_D\mathbf{d} + \epsilon$

such that **y** represents the matrix of observed gene expression intensity values for each of the 73 genes in each of the cohorts of organisms, $\mathbf{X}_Y$ is a design matrix that relates those measured intensity values to the expected effects of measurement probes represented by μ and latent transcription represented by **u**. Next, $\mathbf{Z}_M$ and $\mathbf{Z}_F$ are random effect matrices that regress random effects of sires (*M*) and dams (*F*) on temperature and relate the regressions to **u**, while $\mathbf{Z}_D$ does the same thing for parent interactions. The matrices **b**, **a**, **f**, and **d**, all represent fixed or expected effects of factors on observed data: **b** for temperature, **a** for male breeding values, **f** for female breeding values, and **d** for parental interactions. Finally, $\mathbf{X}_U$ is a design matrix that relates fixed effects for temperature to developmental stages, and **E** and ϵ are error terms.

[6] In the supplementary materials for the research report, the team presents the model in two ways. The first is mostly in scalar format, and is too detailed for my purposes here. The second is in matrix format, and is presented above. Readers can suss out the details of the matrix format, details I omit here, by checking the supplementary materials of (Runcie et al. 2012) and seeing how the two presentations relate.

To use the model, the team first assigned priors to fixed effects μ and **b** and to random effects **a**, **f**, and **d**.[7] Next, they fit their data to the model using a Gibbs sampler, and they updated their priors into posteriors for the fixed effects, the random effects, and the expression variances. Depending on different probability density assumptions for each of those random variables, the team inferred how much each cause influenced the measured data by how much the posteriors differed from the priors. They established the significance of their results by simulating, via Markov Chain Monte Carlo, many different priors and sampled posteriors, and accepting only those results that showed substantial convergence.

Garfield et al. 2013

The team estimated the additive genetic variances for traits across cohorts. For a given trait, that variance, when taken across individuals in proportion to their phenotypic variance, provides the narrow sense heritability (h^2) of that trait, which indicates how much selection can change the mean value of the trait in future generations. Because the team compared cohorts, and not individuals, they couldn't calculate heritabilities for traits. But they calculated additive genetic variances (σ^2_A) to help them address the questions they posed, which were about variations in gene expression levels, their sources, and their relations to each other.

For a given trait or phenotype, the additive genetic variance composes part of the overall phenotypic variance (σ^2_P) such that

[7] I ignore here the many details of their strategy to assign priors as beyond the purposes of this paper.

(k) $\sigma^2_P = \sigma^2_A + \sigma^2_I + \sigma^2_E$

and σ^2_I represents non-additive genetic variance, or the variance due to interactions between maternal and paternal effects, and σ^2_E represents variance due to environmental effects. To use the above model, the team needed to estimate all three values on the right side of the equation for every cohort and for every gene studied, and they needed to compare those results across cohorts and genes.

To estimate σ^2_A, the WGR team relied on the breeding design of their experiment. They developed two ways to approximate σ^2_A, one related to sires and the other to dams. The team used a North Carolina II breeding design, in which they mated six males each with six females, yielding 36 cohorts of offspring, with each cohort having hundreds of embryos.[8] They indexed each cohort according to the cohorts' sires and dams. Offspring within a cohort were full siblings (FS), while those in different cohorts but with the same dam were maternal half siblings (MHS), and those in different cohorts but with the same sire were paternal half siblings (PHS). Thus, when offspring are analyzed in relation to their sires, the variances in paternal effects on offspring phenotypes is the covariance of the phenotype values from the PHS,

(l) $\sigma^2_m = \text{cov(PHS)}$

[8] Furthermore, the team completed the experiment in replicate, for a total of 72 cohorts.

and both of those terms estimate the additive genetic variance for the population

(m) $\sigma^2_m = \text{cov(PHS)} \cong \sigma^2_A/4$

Roughly is same for dams, though for them there is the issue of non-genetic maternal effects.

(n) $\sigma^2_f = \text{cov(MHS)} \cong (\sigma^2_A/4) + \sigma^2_{Mat}$

such that σ^2_{Mat} represents variance in maternal effects due either to maternal genetic or maternal environmental factors. That term is estimated by the remainder of σ^2_m subtracted from σ^2_f. Each of $\sigma^2_{m \text{ and }} \sigma^2_f$ estimated about a quarter of σ^2_A, and their sum estimated half.

For σ^2_I, the WGR team decomposed that term as

(o) $\sigma^2_I = \text{cov(FS)} - (\text{cov(PHS)} - (\text{cov(MHS)})$

Due to technical limitations, the WGR team used estimates that differed in structure from the above equations (l) through (o). They used estimates based on the expected mean

squares of paternal factors, maternal factors, and their interactions. Those estimates are based on the models above, but further explanation of them exceeds my aims here.

As for environmental variance, σ^2_E, the team had two options. First, they could ignore it as uniform across all cohorts, given their research design. Or second, they could treat it as part of error variance. The team couldn't estimate error variance as they did above, due to the structure of, and gaps in, their data. Nor could they calculate significance values for the influences of paternal, maternal, or interaction variances on phenotypic variances. For those later two tasks, they employed a model and simulation technique similar to those used in (Runcie et al. 2012). This model was

$$\text{(p)}\ y_{i,j,k} = u + m_i + f_j + I_{i,j} + e_{i,j,k}$$

such that $y_{i,j,k}$ represents the phenotype within a specific cohort k, which was made by breeding male i and female j from the population of starting sea urchins. On the right side of the equation, u represents the mean value for all cohorts for the phenotype in question, m_i represents the additive effects for the male parent i, and f_i represents the additive effects for the female parent j, while $I_{i,j}$ represents the interaction effects for the parents, and $e_{i,j,k}$ captures the error.

As in their previous study, the model was a Bayesian mixed-effects model. The team assigned distributions to the random variables, assigned priors, fit their model to the data, and calculated posterior probabilities over many simulated iterations until they could compute error terms and significance terms.

Given all that statistical machinery, the team could compare additive genetic variances across cohorts, across genes, and across developmental stages. They quantified significant differences across cohorts and stages using standard statistical tests such as Kruskal-Wallace, Spearman Rho, Wilcoxan rank, χ^2, and they could gesture at a co-variance **G** matrix and the information it should contain if they had used an appropriately larger starting population of dams and sires.

6.4- Epistemic Aims

Open Coding

Using open coding, I collected the following statements about epistemic goals from the team's published papers. The quotes provide a (subjective) collection of sentences in which the WGR team explicitly mentions its epistemic goals. The quotes enable comparisons with the count data presented afterwards. In its published reports, the WGR team often explicitly says it aims to address the questions listed earlier. To avoid redundancy, I record below only statements in which the team doesn't explicitly point to questions.

Review Articles:

From (Garfield and Wray 2009), one passage:

> "Discovering which view, or more likely which combination, best explains patterns of conservation and divergence in development is central to understanding the origins of animal diversity… The advent of genome-scale datasets provides an exciting new approach for evaluating these views" (65).

From (Garfield and Wray 2010), three passages:

> "In the remainder of this article, we will discuss how these new technologies can be used to understand the causes and consequences of changes in gene regulatory interactions" (17).

> "Microarrays have also been used in developmental biology studies aimed at inferring how genes interact during development" (17).

> "Detecting and characterizing each of these kinds of molecular interactions requires a different functional assay, which presents a significant practical challenge" (19).

From (Wray 2013), ten passages:

> "Evolutionary genetics has entered an unprecedented era of discovery, catalyzed in large part by the development of technologies that provide information about genome sequence and function. An important benefit is the ability to move beyond a handful of model organisms in lab settings to identify the genetic basis for evolutionarily interesting traits in many organisms in natural settings. Other benefits are the abilities to identify causal mutations and validate their phenotypic consequences more readily and in many more species" (51).

> "The overarching objective—understanding the genetic basis for trait variation, adaptation, and speciation—remains much the same, but the kinds of empirical evidence, methods of analysis, and motivating questions are all changing in ways that few could have predicted even a decade ago. The operational goals of evolutionary genetics have grown substantially beyond characterizing genetic architecture" (52).

> "In short, the goal has shifted from description to comprehension" (52).

> "The primary goal of quantitative genetics is to identify distinct regions in the genome known as quantitative trait loci (QTLs) that influence variation in a trait of interest" (52).

> "Surveying a genome sequence provides a simple way to find mutations of evolutionary interest" (54).

> "Genomic technologies can now reveal differences in many different aspects of molecular function throughout the genome" (54).

> "Functional genomics can reveal the genetic basis for specific differences in molecular function among individuals throughout the genome…" (58).

"Functional assays can reveal that different mutations affecting the function of the same gene can have distinct functional consequences, for instance, raising and lowering the level of expression…" (59).

"This is an area where genomic technologies provide an enormous boost to our empirical understanding and allow theoretical predictions to be tested in some detail" (61).

"One of the most profound changes in evolutionary genetics during the past decade is an emphasis on understanding not just which genetic changes influence phenotype but how they do so…" (62).

Research Reports:

From (Garfield et al. 2012), one passage:

"In this study, we examine the population genetics of the core, upstream cis-regulatory regions of eight genes (*AN*, *CyIIa*, *CyIIIa*, *Endo16*, *FoxB*, *HE*, *SM30 a*, and *SM50*) that function during the early development of the purple sea urchin, *Strongylocentrotus purpuratus*" (152).

From (Runcie et al. 2012), five passages:

"We investigated how stress responses alter the contribution of additive genetic variation to gene expression during development of the purple sea urchin, *Strongylocentrotus purpuratus*, under increased temperatures that model realistic climate change scenarios" (4547).

"Here, we investigated the response of *S. purpuratus* embryos to a stressful, but realistic temperature range (12–18 °C), and asked if such temperature variation exposed evolutionary relevant GEIs by perturbing developmental gene regulatory networks" (4548).

"To explore cellular processes affected by temperature stress, we tested for enrichment of specific functionally related genes sets among the genes either up-regulated or down-regulated at 18 °C" (4552).

"To explore how transcriptional regulatory networks influence GEIs, we investigated the transmission of temperature effects (an environmental perturbation) and genetic effects (a genetic perturbation) through the well-characterized endomesodermal and ectodermal gene regulatory network that controls cell type specification and development in sea urchin embryos" (4553).

"To quantify the effects of the environment (temperature treatment), genetic background, and other parental differences on the expression of each of these genes, we designed a Bayesian hierarchical mixed effect model" (Supplementary Materials 9).

From (Garfield et al. 2013), seven passages:

"We investigated the relationship between robustness and evolvability within the gene regulatory network underlying development of the larval skeleton in the sea urchin *Strongylocentrotus purpuratus*" (1).

"To better understand the relationship between these seemingly opposed properties of robustness and evolvability, we measured how natural variation in gene expression propagates across a network of interacting genes underlying early development in sea urchins" (1).

"Key to understanding both buffering and adaptation is measuring how, and to what extent, variation in developmental gene function impacts downstream phenotypes…" (1).

"…this is the first study we are aware of that has sought to quantify expression variation throughout an extensive gene regulatory network spanning development from embryogenesis to the production of organismal traits, and to relate variation in gene expression throughout a developmental network and across developmental stages to specific morphological trait consequences" (3).

"In order to examine the extent and consequences of variation in gene expression within the gene regulatory network, we set up a 6×6 cross using outbred parents derived from the same wild population" (3).

"In order to understand the impact that variation in the expression of regulatory genes has on downstream targets, we first examined correlation coefficients (r^2) between pairs of genes for which there is experimental evidence of a direct regulatory interaction" (5).

"The proportion of the total phenotypic variation within a population that can be explained, in a statistical sense, by genetic background-independent contributions of genetic variation is the additive genetic variance of a trait, and it is the size of this variance relative to overall phenotypic variance that puts bounds on the efficacy by which selection can change the mean value of a trait" (Supplemental Materials 3–4).

Count Data

Below, I present three tables. Table 6.1 collects raw data from the content analysis. That data includes a lot of noise among the signal, so I checked every datum against its surrounding text and labeled noisy data as one of: infelicitous homonyms/cognate, metadiscourse/figure labelling, not describing project rational/ describing rationales of other teams. I class the data that remained as informative data, and I tabulate in the second table. Finally, to compare results across articles, in the third table I collect the relative ranks of different epistemic aims for each article.

TABLE 6.1

RAW COUNT DATA

Research Reports	**Goal**	**Know**	**Amalgam**	**Control**	**Describe**	**Discover**	**Explain**	**Predict**	**Cause**
2012a	35	33	6	9	50	14	8	18	86
2012b	96	47	9	34	103	61	19	60	448
2013b	226	48	22	42	202	90	39	122	468
Reviews									
2009	8	9	0	6	8	6	9	5	52
2010	34	35	6	14	52	19	25	7	261
2013a	50	81	11	25	53	27	16	8	225

Table 6.2 below has several interesting caveats. First, in the column labeled Goal, there are normal counts and counts in parentheses that include data associated with the search term ‘design*’. The team bred sea urchins in their lab using a breeding design called the North Carolina II breeding design, in which each dam is mated with each sire. The team often mentions the design in 2012b (Runcie et al. 2012) and 2013b (Garfield et al. 2013) research articles, and those mentions are exhausted by repeated use of ‘NCII

design' and similar phrases. Those mentions are informative about the aims of the team, but it also helps my purposes here to indicate the overrepresentation of such data

TABLE 6.2

INFORMATIVE COUNT DATA

Research Reports	**Goal (w/ 'design *')**	**Know**	**Amal gam**	**Contro l**	**Describe (w/o 'correlat *')**	**Discover (w/o 'probe*')**	**Explain**	**Predict (w/ 'correlat *')**	**Cause**
2012a	5	9	1	0	31	8	7	18	86
2012b	9 (20)	14	0	1	67 (29)	58 (2)	12	17 (55)	448
2013b	7 (30)	16	0	0	137 (40)	77 (2)	23	12 (109)	468
Reviews									
2009	1	9	0	0	3	6	8	4	52
2010	10	19	0	0	29	14 (13)	13	7	261
2013a	18	41	3	4	40	22	12	6	225

compared to other possible goal terms. For instance, in the 2013b research report, data from instances of 'design' comprise 23 of the 30 data for Goal.

Second, the columns for Describe and Predict include normal counts and counts in parentheses related to the search term 'correlat*'. For Describe, the parenthetical numbers exclude data for that search term, for 'Predict' they include that data. Correlations serve multiple aims in scientific projects. They describe relations between phenomena, in this case relations between variation in gene expression. But researchers often use correlations as bases for predictions about the future behavior of systems, even if they can't specify the underlying causal system. Here I count instances of 'correlat*' as evidence of describing phenomena, but not for predicting phenomena. I do so for two reasons. First, the team lacked the correlations before the start of their project, and thus could not test them as predictions. Second, in the data from their review papers, in which

the team rarely discussed correlations, there remains a strong signal ranking the importance of describing phenomena over that of predicting phenomena. I argue that the aims of the review papers inform the aims reported in the research articles.

For a third caveat, the column for Discover includes normal counts and counts in parentheses that exclude data associated with the search term 'probe*'. In almost every use of that term, the team described a bit of technology, not a specific aim. The team used the technology to measure the amounts of RNAs in their samples. Some may argue that the term 'probe' for those bits of technology doesn't reflect the aims of the team, and that my data collection techniques would have ignored mentions of such tools had those tools another general name, for instance if some odd acronym like GWAS (for 'genome wide association studies') had become fixed in the research community as the name for what we call probes. I'm sensitive to such criticisms, but I'm ultimately unpersuaded by them. Probes are tools named for their functions, and while researchers could have baptized them differently according to the accidents of history, that's no reason to ignore data about possible epistemic functions/aims when we have such data. That judgment is upheld for this case given the strong signal for Discover in the data for review articles.

Regardless of how one feels about my choices for the counts discussed above, I provide in parentheses what the counts would have been if I'd excluded or included potentially troublesome data. But with the counts in the parentheses above, readers can account for any worries they have and re-rank the aims for each paper.

Table 6.3 below orders different goals according to their rank in overall count data for each article. In most articles, the epistemic aims of describing phenomena and

discovering new phenomena dwarf the importance of predicting phenomena or of explaining phenomena.

TABLE 6.3

RELATIVE RANKS OF EPISTEMIC GOALS

Research Reports	**Goal 1**	**Goal 2**	**Goal 3**	**Goal 4**
2012a	Describe	Predict	Discover	Explain
2012b	Describe	Discover	Predict	Explain
2013b	Describe	Discover	Explain	Predict
Reviews				
2009	Explain	Discover	Predict	Describe
2010	Describe	Discover	Explain	Predict
2013a	Describe	Discover	Explain	Predict

6.5- Primary Conclusions of the Project

Garfield et al. 2012.

For this study, the team described the DNA sequences of the regulatory regions and of select neutral regions for eight genes in several adult purple sea urchins. They calculated the amount of variation in those regions among the population, and then they compared the variation in regulatory regions to the variation in neutral regions, and then they compared both to that of related species. Those comparisons enabled them to use models that suggested the presence or absence of evolutionary forces on those regions. The team's primary conclusions were:

1. Within regulatory regions, variation in the length of the DNA sequences is a legitimate phenomenon, one not captured by most analyses of variation.

2. Summed across the genes, the variation within regulatory regions doesn't significantly differ from the variation in neutral regions. Within regulatory regions, the variation in binding sites doesn't significantly differ from non-binding sites.
3. D and H statistics and McDonald-Kreitman tests indicate that some of the regulatory regions are under directional selection, while others are not. Some face negative or purifying selection, while other face positive selection.
4. D and H statistics also indicated that some putative neutral regions face selection, perhaps indicating that they contain undiscovered regulatory elements.

Runcie et al 2012.

For this study, the team raised purple sea urchin embryos at different temperatures and assessed the impact of higher temperatures on the overall embryos and the expressions levels of genes, especially those within the endomesoderm specification GRN. For each of three temperatures, they measured the amounts of gene expression of thousands of genes in hundreds of embryos in distinct cohorts and at three different temperatures, and they calculated the variation in expression levels within and across temperatures. They also focused in genes in the endomesoderm GRN, measured their expression variation, and correlated such variation with variation in gastrula morphology. Those comparisons enabled them to make the following conclusions:

1. The higher the temperature (12°, 15°, and 18°C), the more embryos developed abnormally, with slower growth rates, smaller embryos, abnormal blastocoels shapes, and higher expression amounts for two chaperone genes. While those

phenomena indicated stressed development under higher temperatures, urchins still developed through juvenile stages without deformities and without dying.

2. The higher the temperature, the more gene expression levels changed (increased or decreased) for more than 2000 genes, more than half of which changed expression levels by at least half, and almost a quarter of which changed expression levels two-fold. These changes affected 4 different kinds of molecular functions.
3. For 72 genes within the endomesoderm GRN, higher temperatures varied the expression amounts of 14 genes. However, for any two genes for which there was an established regulatory interaction, those variations in amounts didn't correlate. That lack of correlation indicates that GRN sub-structure and sub-functions are buffered from environmental perturbations.
4. While parents had strong hereditary effects (parental genetic and maternal) on the expression levels of their offspring, those effects didn't interact with temperature effects.
5. Given the previous two conclusions and projections about the extent of temperature/climate change in the purple sea urchins' Pacific Ocean environment, forthcoming climate change is unlikely to reveal cryptic genetic variation in those urchins.

Garfield et al. 2013.

For this study, the team raised purple sea urchins embryos, measured the expression amounts of 74 genes in the GRN that specifies the endomesoderm, ectoderm,

and juvenile skeletons in the developing organisms. They took measurements at seven temporal points in the process of development. They also measured the lengths of the juvenile skeletons at the seventh point of development. They calculated variations in expressions amounts in skeleton lengths, they compared those variations across cohorts to determine the effect of parental genes on the variations, and they calculated covariances in variations across expressions levels of different genes, across cohorts, and across expression levels and skeleton lengths. Those calculations enabled them to make the following conclusions:

1. Parental effects had detectable and significant effects on the expression levels of the genes in the offspring, but those effects explained at most 15% of the variation in levels among the offspring. Furthermore, parental effects influenced such variation most strongly in early development.
2. For any two genes in which researchers had already established a regulatory interaction, the variation in expression levels for those two genes correlated more strongly than did any two random genes, except for early in development.
3. As development proceeded, the proportion of regulatory interactions that functioned like switches decreased, as the number of interaction that were sensitive to quantitative variations increased.
4. For eight genes studied, variations in their expression levels correlated with variations in lengths of juvenile skeletons. Those genes functioned either at the top (upstream) or the bottom (downstream) batteries in the gene regulatory cascade.

CHAPTER 7

COMPLEX PHENOMENA, RESEARCH SYSTEMS, AND THE PROSPECTS FOR DEVELOPMENTAL EVOLUTION

7.1- Introduction

In this chapter, I analyze and compare the case studies about the Johnson, Porter, and Tulchinsky (JPT) project and about the Wray, Garfield, and Runcie (WGR) project. From them I infer some conclusions about the prospects for evolutionary genetic models with gene regulatory information built into them.

This chapter addresses the questions that drive this project. While the previous two chapters provide descriptions of the research projects I study and enable comparison, this chapter interprets those descriptions in relation to the driving questions. This chapter has four further main sections.

The second section introduces accounts of research systems and of complexity. In the third section I use those accounts to summarize and analyze the cases. In the fourth section I compare the cases to show how they are similar in their epistemic aims and in their implicit conception of *gene*. In the final section, I address the driving question: What are the prospects for developmental evolution?

7.2- Research Systems and Complexity

Research Systems

I propose a concept *research systems* to help me analyze the two cases. The concept is similar to and inspired by Hans-Jörg Rheinberger's concept of *experimental system* (Rheinberger 1997). For Rheinberger, at least two kinds of elements compose experimental systems: epistemic things and technical objects. Roughly, for a specific experimental system, the epistemic thing is the object under study, often a process or a mechanism. Technical objects are the devises, materials, research skills, and other conditions that enable researchers to continually tinker on the epistemic thing to continually engender unexpected events from it. Experimental systems and their parts have histories.

While Rheinberger's concept is suggestive and fruitful, I propose a slightly different concept for several reasons. His concept has many facets, and I'm not sure I can collect and organize all of them. As a result, I find the concept too imprecise to use as a tool for analysis. Furthermore, Rheinberger explicitly says that he aims to use the concept to draw attention to objects and experimental practices, and away from theories. That shift is valuable and important, but ultimately we can't make sense of objects as epistemic things without relating them to theories, models, and other scientific products that have epistemic function in relation to those objects. Given those reasons, it would be misleading and unfair for me to use his term and concept. Regardless, research systems are much like experimental systems.

A research system has three primary kinds of components. The first is the object or phenomenon in the world.[1] Such phenomena can be items or processes. The second is a set of scientific products that have functional relations to the phenomena. Such products can include theories, models, representations of laws, etc. The functional relations include describing, predicting, explaining, discovering, etc., such that the at least one of the scientific products has at least one functional relationship to the phenomena. One product can have multiple functional relationships to a phenomenon, for instance, if a single model describes, explains, and predicts the effects of a process. Furthermore, different scientific products can have different functional relationships to a phenomenon, for instance when one theory describes it, another predicts it without explaining it, and still a third explains it.

The third kind of component in a research system includes the artifacts and materials that researchers use to study the systems. These include devices used to control phenomena, instruments used to measure phenomena, equipment like computers, research space, etc.

Research systems are historical objects, and a given system needn't have all of the same components in all of their historical stages. Rheinberger is right that much research begins when researchers pick a system, not when they pick a theory to test. In many cases, as researchers develop a project, they develop the devises and strategies to control the phenomenon, to study it, to collect data about it, etc., all practices that focus and

[1] Here I treat phenomena and objects as the same kind of thing, though we might fruitfully distinguish them. We might take objects as the things in the world, and phenomena as representations of those things that we construct from data. I avoid that distinction here for ease of presentation, and to sidestep debates about distinctions between data and phenomena, and between objects and phenomena. That said, in general, I'm for such distinctions, though that position is irrelevant to my presentation of research systems.

constrain their studies. Researchers similarly develop theories, models, etc., to describe, predict, and explain that phenomenon. As a result, researcher often, but not always, better conceptualize and isolate the phenomenon as an object in the world. The parts of the system evolve, and as a result, so does the whole system.

Researchers can use research systems in many different studies that comprise longer research projects. For a given study, a team uses a system, and they design their studies, and report their results, in rationales. The rationales evolve over time, and the one used to design a study may differ from the one used to report the results of the study. While researchers develop theories and models to be exportable to other phenomena and research systems, they also develop research systems to be exportable to other labs and research teams.

A research system differs from a research project. The former admits of many possible theories, epistemic functional relations between theories and phenomenon, research teams, rationales, and investigative strategies. The latter, on the other hand, is a socio-epistemic unit in which a specific team in a specific place and time applies specific investigative strategies and specific theories for specific epistemic functional relations to the research system.

From a given project, a team can export at least:

1. scientific products (theories, models, etc.)
2. data
3. devices, instruments, equipment
4. know-how in the form of skilled practitioners or in research protocols

5. the whole research system, either as a spatio-temporal object or as a schematic to develop copies of the spatio-temporal object elsewhere.[2]

Complexity

I use the above account of researcher systems to propose an account of complex phenomena. Many have recently discussed notions of complexity, and often they debate about how to identify complexity. Here, I distinguish complex phenomena from complex theories or explanatory structures. I propose two ways in which phenomena can be complex.

The first notion of complex phenomena is inspired by Morton Beckner (Beckner 1959). Beckner argued that phenomena are complex at least relative to the theories and concepts we use to analyze them. For instance, we use 'flight from predator' as a concept to describe the behavior of an antelope running in front of a puma. But if we aim to describe the behavior of a puma running in front of an antelope, then the concept 'flight from predator' won't apply, the phenomenon will be complex relative to that concept, and we must search for another concept to describe that phenomenon. Similarly, if we aim to casually explain the movement of a molecule from the outer parts of a cell into the nucleus, we would use a mechanistic model of that shows how a chaperone molecule tugs the other molecule to the nucleus. But if we aim to explain how that first molecule was made, that phenomenon remains complex in relation to the mechanistic chaperone model.

[2] I don't discuss research systems as outputs of research projects in earlier chapters. I only saw the difference between research systems and research projects, and the importance of the former, once I completed the case studies in the previous two chapters.

While Beckner provided little more than a few comments about complex phenomena, from his comments I propose a couple of accounts of complexity that we can fruitfully use to analyze research systems. The first account depends on the functional relations between phenomena and the scientific products we use to understand them. While Beckner focused on theories and laws, my account applies to scientific products more generally. Furthermore, while Beckner focused on aims of description and explanations, my account applies to epistemic aims more generally.

A phenomenon is complex$_1$ only

1. relative to a scientific product (theory, model, law, etc.) or a set of such products
2. relative to an epistemic aim or function (description, explanation, prediction, discovery, etc.) or set of such aims or functions
3. if the specified scientific product(s) can't achieve the specified aim(s)

The above account provides necessary conditions for phenomena to be complex. I don't claim that those conditions are jointly sufficient. Such claims are often false and often lead to fruitless disputes over counterexamples. I don't aim to provide the true concept of complexity, but instead to propose concepts that will help analyze cases of research systems. By claiming only that the conditions are necessary, I focus discussion on the analytic functions of the concept, not on fruitless counterexamples. Furthermore, I propose another account of complex phenomena. To avoid confusion or issues with possible sufficiency, I label the two accounts of complexity as distinct, so I can ignore such issues. The account above is about complexity in the first sense, complexity$_1$. The

next account is about complexity in the second sense, complexity$_2$, also shown in Fig. 7.1.

A phenomenon is complex$_2$ only if

1. we can decompose it into parts
2. for at least one part, at least one scientific product has at least one epistemic functional relation (describes, explains, predicts, etc.) to it
3. at least two parts have epistemic functional relations to distinct scientific products
4. no single scientific product achieves all of the epistemic functional relations achieved by the other distinct scientific products
5. each scientific product employed to have an epistemic functional relation to a part is conceptually compatible with every other such product

While both senses describe different kinds of complex phenomena, they have different import for scientists.[3] If a scientist determines that a phenomenon is complex$_1$, that determination might motivate her to find epistemic aims or scientific products for which the phenomenon will no longer be complex$_1$. The situation is different for complexity$_2$. If scientists aim to alleviate the complexity$_1$ of a phenomenon, after years of research, they can end up in a situation in which the phenomenon is comlex$_2$. What should they do then? Insofar as the scientific products help them achieve their epistemic

[3] With the above two concepts, we can understand a few more concepts related to complexity. *Simple phenomena* are those that meet the first two conditions of *complex$_1$*, but fail the third. *Irreducibly complex phenomena* are those that are complex$_1$ for any theory and for any aim, and always fail the first condition of *complex$_2$* (perhaps due to our methods of conceptualizing phenomena).

ends, they can rationally rest content. Some, however, will aim to unify the scientific products so that they can understand the complex$_2$ phenomenon with fewer products or with fewer partitions in the phenomenon.[4] But they needn't.

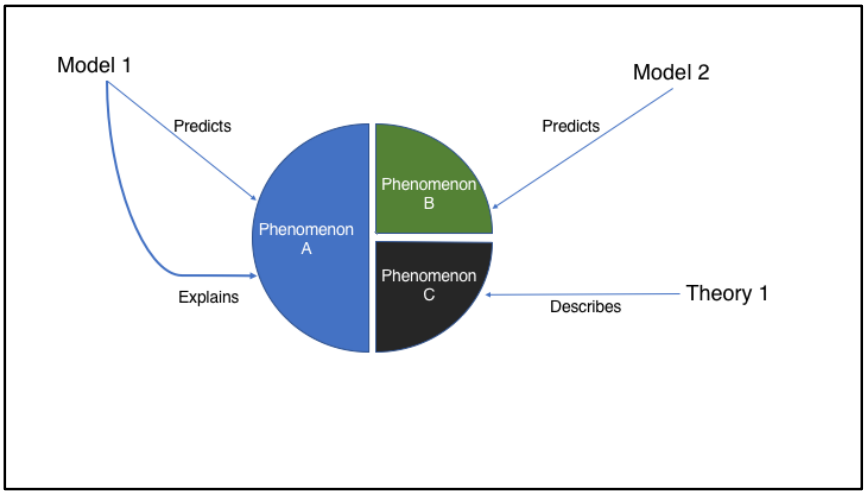

Fig. 7.1. Graphical Representation of a Complex$_2$ Phenomenon. The circle represents a complex$_2$ phenomenon, which is decomposable into three parts: Phenomenon A, B, and C. There are also three scientific products: Model 1, Model 2, and Theory 1. The products have epistemic functional relations (predicting, describing, and explaining) to distinct and parts. No scientific product has an epistemic functional relation to all parts.

The Upshot

In this chapter, I argue that when research teams combine evolutionary genetics and developmental genetics, they partly create research systems in which the phenomena studied are what I call complex$_2$. In doing so, they use models in such a way that they are conceptually compatible, but not necessarily coextensive. Much of their work involves not creating overarching models or theories, but creating bridge concepts that enable

[4] One could argue that Newton lessened or eliminated the complexity$_2$ of motion when he used his laws to describe, predict, and explain both terrestrial motion and extraterrestrial motion, which had previously been distinct parts of the overall phenomenon of motion, and had been epistemically related to distinct scientific products of Galileo's laws and Kepler's laws, respectively.

them to move from (or conceptually staple together) one part of a comlex$_2$ phenomenon to another part.

Furthermore, when people argue that scientific products from evolutionary genetics cannot be synthesized with those from developmental genetics, they sometimes intimate that researchers can't lessen or eliminate what I call the complexity$_2$ of their research systems. In other words, there will never be an overarching theory of the evolutionary genetics of gene regulatory networks. That may be true, though it's an empirical question. More importantly, researchers needn't take, and don't currently take, the lessening of what I call complexity$_2$ as a rational condition for epistemic success when they synthesize models from developmental genetics with those from evolutionary genetics. So some of the complaints of their critics are misplaced. On the contrary, researchers create research systems of complex$_2$ phenomena.

To provide examples of complex research systems, and to understand the two cases studied in this dissertation, I use the concepts sketched above to analyze those cases.

7.3- Case Analysis

For both cases, I show how the teams created research systems about complex$_2$ phenomena. I describe the phenomena they studied, how they partitioned them, the models that had functional relationships to the parts, and the devices and strategies and equipment that the teams used to study the complex phenomena to engender new

information about the phenomena. Finally, I draw some similarities from the two analyses.

JPT System

The Johnson, Porter, and Tulchinsky (JPT) team studied the process of speciation. They made the phenomenon of a speciation complex[2] by cleaving it into at least six parts:

1. Gene interactions within an individual organism, the production of a gene product (RNA or protein) as the terminal event of a GRN;
2. Organism life cycle;
3. Population of organisms, their breeding, and their generational turnover;
4. Variation of phenotypes within population;
5. Two distinct populations;
6. Production of distinct species populations that yielded unfit hybrids;

The JPT team used at least five different models, which had epistemic functional relations to the above phenomena. Based on their structures, I also categorize those models as either mechanistic, statistical, or other.

A. Gene regulatory network model. As the project evolved, the team changed the model to capture bioenergetic interactions of transcription factors binding to DNA regulatory elements. (Mechanistic).
B. Custom JPT evolutionary genetic models, which are sets of equations for calculating phenotype values and fitnesses. As the project evolved, the team

changed the equations for the phenotype values to capture bioenergetic processes of transcription factors binding to DNA regulatory elements. (Statistical)

C. Model of organismal reproduction and generational turnover. (Mechanistic)
D. Data models. (Statistical)
E. Bateson, Dobzhansky, Muller (BDM) model of speciation. (Other)

Based on the structural category into which each model falls, we'd expect the models to have the following epistemic functional relations to the six phenomena.

- The GRN model describes, explains, and predicts phenomenon 1.
- Taken as interpreted models, the custom evolutionary genetic models describe parts of phenomenon 1, calculating phenotype and fitness values. The amount of gene product counts as the phenotype value for the organism or its GRN.
- The model of reproduction and generational turnover describes, explains, and predicts phenomenon 2, and describes and explains phenomenon 3.
- The data models describe phenomenon 4.
- The BDM model describes phenomena 5 and 6.
- As a set, and taken purely theoretically, all four models predict the probability of speciation across many runs of evolving populations.

The above functional relations are consistent with the content analysis results (Table 5.4) for JPT's research reports. Table 5.4 ranked the prevalence of different epistemic aims explicitly or implicitly mentioned in those reports. The table showed that, by number of words/sentences related to each epistemic aim, description of phenomena

and prediction of phenomena were the teams most important aims, with explanation of phenomena a distant third. The above relations show that all the models employed described parts of the complex$_2$ phenomenon (speciation), and that the theoretical system as a whole enabled predictions of speciation. Furthermore, while two models mechanistically explain phenomena, the team only appeals to the explanations accomplished by the GRN model. I collect most of the above information in Figure 7.2

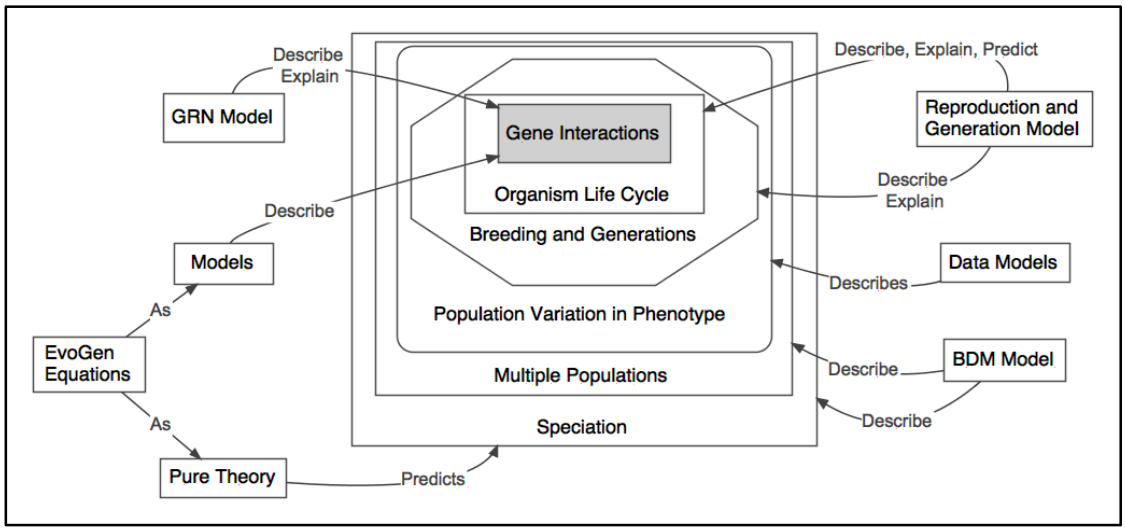

Fig. 7.2. The Complex Phenomenon of the JPT Research System. The figure shows the nested complex phenomenon studied by the JPT team in their project. The Center most grey box represents gene interactions within individual organisms. The white box outside of it represents individual life cycles, and the hexagon outside of it represents breeding and generational turnover in generations. The box outside of it represents variation in phenotype values at the population level, nested within a box representing two multiple distinct populations. The outermost box represents speciation. The various boxes outside the nested boxes represents scientific products, which relate to the phenomena via epistemic functional relations, represented as arrowed lines.

Figure 7.2 remains a hypothesis in need of further validation. While it is consistent with the content analysis results, I've yet to show that it's confirmed or

disconfirmed by them. To do so, I'd need to examine each count datum, code them as related to a specific model, and tally the results. I leave that task for future research.

Finally, the complex$_2$ phenomenon studied by JPT is part of a research system that enabled them to study that phenomenon over many years, and to tinker with it to simulate new events. To study their phenomenon, the team used equipment, devices, and instruments. The equipment they used included the computers on which they wrote their programs and ran their simulations. Those programs were devices to control and simulate phenomena and instruments to measure the simulated phenomena. The claim that the JPT team tinkered with their research system over time is confirmed by Table 5.1, which collects most of the manipulations the team conducted.

WGR System

The Wray, Garfield, and Runcie (WGR) team studied variation in alleles and in gene expression across purple sea urchins. They made those phenomena complex$_2$ by separating them into the following (hierarchical) parts.

1. Within individual organisms, gene structures, gene productions, and gene interactions.
2. Organismal growth and development.
3. Population variation in gene structure, amount of gene products, degree of gene interactions, and other phenotypes such as shape of blastula and time to gastrulation.
4. Evolution to structures of genes.

The WGR used a plethora of models to better understand all of those phenomena.

A. Gene model. (Other)
B. GRN model. (Mechanistic)
C. Developmental stages tables. (Other)
D. Data models such as π, θ, π^i, ρ, and their related estimators. (Statistical)
E. Covariance data models. (Statistical)
F. Custom Bayesian models. (Statistical)
G. Models of evolution: drift only, Tajima's *D*, Fu and Li's *D*, Fay and Wu's *H*, modified McDonald-Kreitman, modified Hudson-Kreitman-Aguade. (Statistical)

Based on the category of the models, we'd expect the above models to have the following functional relationships to the phenomena.

- The gene model describes gene structures and products.
- The GRN model describes gene interactions, and it predicts and describes their outcomes. It also partially predicts and explains organismal growth and development.
- The developmental tables describe organismal growth and development.
- Data models like π, θ, π^i, ρ describe the variation in gene structures across the population.
- Covariance data models describe some of the variation the amount of gene products or in the degree of gene interactions across the population.
- The custom Bayesian models describe and partially explain the variation in the amount of gene products across organisms in the population.

- Standard data models/summary statistics describe the variation in other phenotypes across the population.
- Models/statistical tests of evolution describe the evolution to the structures of genes.

The above functional relations are consistent with the results from Table 6.3. of the WGR team's publications, which ranks the relative importance of different functional relations for each publication. In that table, the highest ranked functional relation for all but one of the publications is that of describing phenomena, a relation included in all of the above bullet points. I summarize most of the information above into Figure 7.3, which remains a hypothesis in need of further validation just like Figure 7.2 and for the same reasons.

But Table 6.3 is not fully captured by Figure 7.3. In the table, one of the primary aims of the WGR team is to discover phenomena, a category not included above. But in Table 6.3, the aim to discover new phenomena was the second most important aim for all but one of the publications. I suggest that the aim to discover new phenomena, while legitimate, is derivative on the aim of describing phenomena. The aim to discover a specific phenomenon is historically sensitive to a research community and to the body of knowledge at a time. Once achieved for a given phenomenon, research system, community, and body of knowledge, it cannot be achieved again. But researchers can still export the scientific products they used to discover that phenomenon to related phenomena, research systems, communities, and bodies of knowledge.

Figure 7.3 also shows that three kinds of models explain their target phenomena: the GRN model, the Bayesian model, and the evolution models. Explanation of

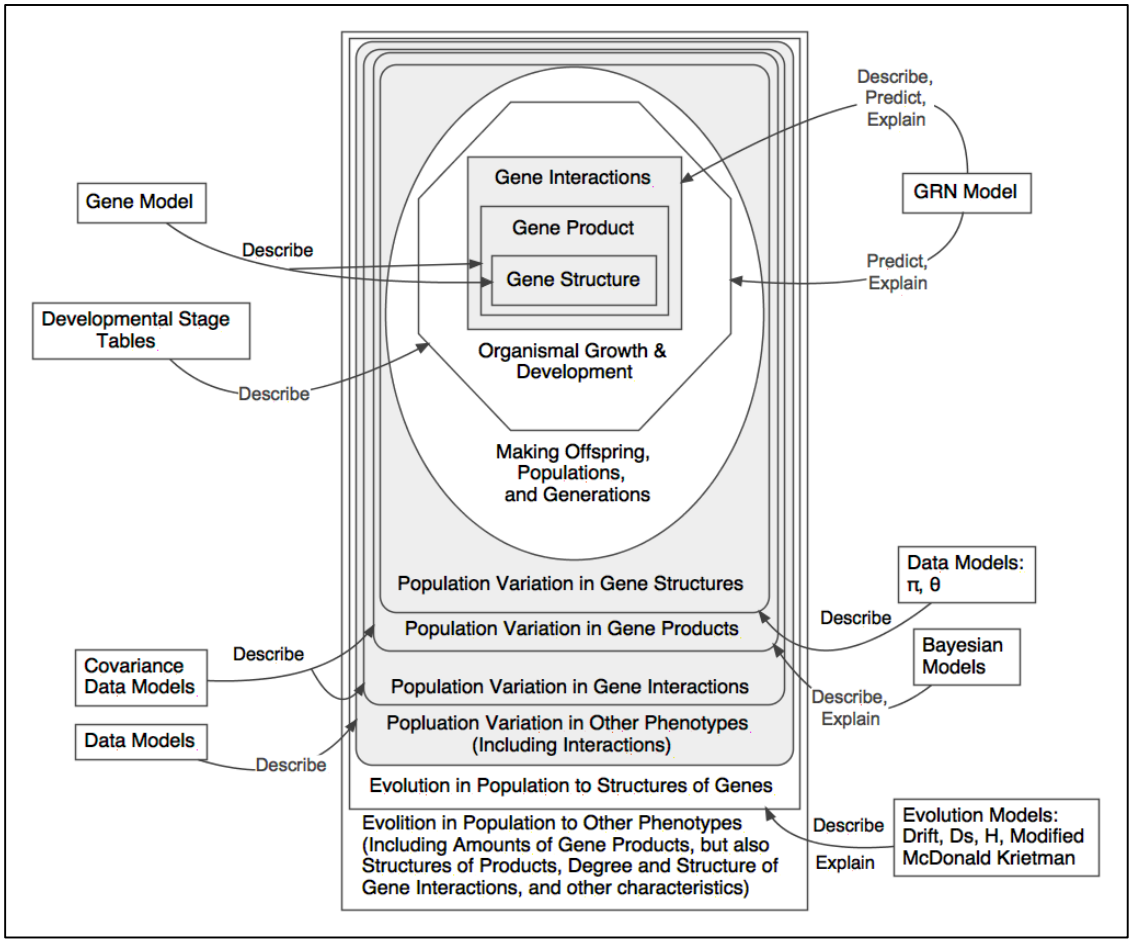

Fig. 7.3. The Complex Phenomenon of the WGR research system. The figure shows the nested complex phenomenon studied by the WGR team over the course of their project. The center-most grey boxes represent three levels of phenomena within individual sea urchins: gene structures, products, and interactions. The white hexagon represents development and growth in individual sea urchins. The white oval represents the phenomena of individuals making offspring, and bridges the individual and population levels of phenomena. The four grey curved-corner rectangles represent different population-level phenomena of variations, while the two white square-cornered rectangles represent population-level phenomena of evolution. The boxes outside of the nested system represent scientific products, and the arrows between those boxes and the nested levels indicate the epistemic functional relations between scientific products and their targeted phenomena.

phenomena is the third highest ranked epistemic relation from Table 6.3. I discuss each of the above model/ phenomenon relations in order.

The GRN model is mechanistic, and as such it explains a phenomenon by describing the subsystem that achieves the phenomenon and the workings of the subsystem such that it achieves that phenomenon. The model decomposes part of organismal growth and development into gene networks, and it decomposes gene networks into genes and gene products. The activities and interactions in the GRN explain how genes make their products in their cellular contexts, and ultimately how cells differentiate and cause the organism o develop. The GRN is unique among the WGR research system as the only scientific product that bridges multiple levels of the $complex_2$ phenomenon to explain target phenomena, and to do so mechanistically.

The Bayesian models explain, in a statistical sense, population variation in the amount of gene regulatory products. They do so via analysis of variance, by breaking the variance into parts and quantifying the sources of those parts.

The models of evolution as statistical tests explain how populations are evolving by identifying the kind of selection operating on them. Such selection can be positive, negative, balancing, non-existent, etc. Such identifications, however, are tentative, what I call how-likely explanations. Those identifications provide working hypotheses that researchers can study either experimentally in in many kinds of extant populations, or via simulation for extinct species, dead generations, or difficult to study actual populations.

In (Wray 2013), Wray explicitly discusses the evolving epistemic aims of evolutionary genetics. “The operational goals of evolutionary genetics have grown substantially beyond characterizing genetic architecture. In short, the goal has shifted

from description to comprehension" (Wray 2013, 52). While Wray discusses the ways researchers now get functional information about genes in relation to each other, he at best gestures at meanings for the quoted text. I use Figure 7.3 to explicate that quote. To comprehend evolutionary phenomena, Wray means that we must give descriptions of the links between the functions of genes and their effects on organisms' development, and that we must causally explain the effects of those links. GRNs both describe the links between those levels of phenomena, and they explain the effects of those links. Such explanations provide one causal component of an analysis of variance on population-level variation. Such an analysis is of genetic variance, which is to be combined with components for environmental variances, interaction variances, and error variances.

Finally, the WGR team used many devices and tools to develop its research systems. In Figure. 7.3, no scientific product relates to the phenomenon of breeding. That said, this phenomenon was central to the WGR team. They controlled this phenomenon with the North Carolina II breeding design, and so that design, while not a scientific product in the normal sense, is a device central to the research system. Furthermore, the team used a plethora of devices to gather data about various levels of their complex$_2$ phenomena, many computer programs to analyze that data, and further programs still to evaluate the significance of their results.

7.4- Cross-Case Comparisons

The projects have three primary similarities as they relate to epistemology: the construction of research systems with $complex_2$ phenomena, strong signals for non-explanatory epistemic aims, and a shared gene concept. I discuss each in turn.

Research Systems

When philosophers and theorists talk about the outputs of evolutionary biology, they tend to focus on models, theories, conclusions, and data. All of those things are important products, and we can do much to better understand how research construct those items and how they diffuse through a research community. But another possible output is the research system, or at least the framework for recreating the system at other locations.

A research system enables researchers to return to a relatively controlled phenomenon, to tinker with it, to ask new questions of it, to develop new scientific products to address it, and to evaluate different functional relationships between those products and the parts of the phenomenon.

Both teams studied here constructed unique research systems. For the JPT team, their research system enabled them to tinker, for more than a decade, with the phenomenon they studied, simulating speciation with roughly the same set of models, but varying the number of genes and alleles, the kinds of selection, the strength of mutation, migration, and ultimately developing a model of gene regulation that was both bioenergetically realistic and tractable with fitness. For the WGR team, their research system enabled them to explore, for several years, the amount by which genes varied in

their expression, and how those variations correlated with variations in different embryonic phenotypes.

The research systems have two commonalities that indicate aspects of the theory of knowledge implicitly shared by the projects. First, the phenomena studied are $complex_2$, for which the teams make no apologies. I infer, then, that the phenomena studied in this burgeoning field will in many cases be $complex_2$.

Second, and as a corollary, no single model or theory will have all desired functional epistemic relations to all parts of the phenomena. Rather, the systems of models used are piecemeal. The teams require no overarching theory to unite all the models. However, they do rely on a bridge concept of gene, discussed in further detail later, to at least ensure that GRN models and evolutionary genetic models pick out the same bits of the genome.

Epistemic Aims

Both of the projects studied here deploy systems of models to meet myriad epistemic aims for the $complex_2$ phenomena they study. As captured by figures 7.2 and 7.3, the models are specific to one or two levels or parts of the phenomena, and some serve only one or two epistemic functions. For instance, with the WGR team, the data models π and θ describe the population variation in gene structure, but they don't explain that variation, nor do they predict, describe, or explain the gene structures themselves.

From the content analyses, the research reports have strong signals for the relative importance of non-explanatory epistemic functions. Especially strong was the signal for the epistemic function of simply describing phenomena, the highest or second-highest

ranked epistemic function for all research reports. Also important were the epistemic aims of predicting phenomena (JPT team) and discovering new phenomena (WGR team), the latter of which I classified as a secondary form, subject to temporal indexing, of describing phenomena.

Nearly all who market evo-devo or devo-evo stress that the new disciplines will enable researchers to causally explain evolutionary phenomena with developmental mechanisms. Given the results from this dissertation, we might note two things. First, evolutionary geneticists often aim to give phenomenological models, or those that enable them to describe and predict phenomena, not necessarily to explain phenomena. Insofar as researcher study phenomena with models of developmental mechanisms and of evolutionary mechanisms, they might take care not to forsake other epistemic aims as without value. A system of models that enables us to explain an evolutionary phenomenon, but not to control it or predict it, may find less use than a system that enables us to do the latter, but not the former.

Second, for complex_2 phenomena, models of development mechanisms can at best be parts of causal stories for such phenomena, but they can't explain all parts of those phenomena. With an account of complexity_2, proponents of evo-devo and devo-evo can now better describe the explanatory role and scope of models of developmental mechanisms in the systems of models constructed to understand complex_2 phenomena.

Gene Concepts

Both teams are somewhat cavalier with how they conceptualize genes. Neither project explicitly defines their gene concepts, but based on how they study genes, we can infer at least a couple of aspects implicit in a concept shared across the two projects.

For both projects, a gene must have at least two features. First, an allele of that gene must be comprised of sections of chromosome (sections that needn't be contiguous nor, nor perhaps even on the same chromosome) that produce a single kind of RNA. While there are one-one relations between alleles and kinds of RNAs, there needn't be one-one relations between alleles and proteins, or between contiguous segments of DNA and kinds of RNA, but there can be such relations for specific alleles. In many cases, each allele will have a regulatory region on the 5′ side of the start codon for the allele's most inclusive coding region, and that region will contain most of the binding site for regulatory molecules. That site will vary in length depending on the allele. A gene, and its alleles, that has the above feature can be represented in GRN models.

Second, a gene that produces an RNA must satisfy the Mendelian laws of segregation, and its alleles must satisfy the Mendelian law of independent assortment. Insofar as those laws are met, researchers can use the models of evolutionary genetics to study how the frequency of that gene changes (or doesn't) in populations. The JPT adopts these Mendelian assumptions throughout its project, and the WGR team assumes them for statistical tests of evolution in (Garfield et al. 2012), and for their breeding design in all their studies.

Many have viewed those two aspects as at best in tension, and perhaps incompatible (Falk 2000; Falk 2010; Griffith and Neumann-Held 1999). The first aspect

of the concept is one commonly associated with a gene concept that evolved, starting in the mid-twentieth century, in molecular biology, while the latter is one that evolved, starting in the second half of the nineteenth century, in studies of heredity and transmission of hereditary material from parents to offspring (Waters 1994; Weber 2005).

One way to understand the tension between those aspects is to consider them as rules for isolating phenomena. Historically, many hoped to relate the two concepts to each other (Weber 2005). Their general strategy was something like this. Use each concept to pick out a set of phenomena. Insofar each set includes the same material entities, then we can establish relations between the concepts, perhaps conceptually reducing one to the other, or perhaps interpreting one with the other. But the strategy went astray at the first step. Each concept appeared to refer to different sets of entities. As the concepts pick out different, though overlapping, sets of referents, many concluded that the relations between the two concepts are at best unclear, and at worst opaque.

The projects studied here overcome that tension in at least two interesting ways. First, for a given species, neither team considers all possible units that the first aspect of their gene concept might pick out, then all possible units that the second aspect might pick out, and then compare the two sets. Rather, they use GRNs to focus their searches. By starting with GRNs, the teams focus on a set of genes known to meet the first aspect of the definition. Next, they assume that those genes meet the second aspect of their gene concept. Given that assumption, the teams study the evolution of those genes within populations, and how that evolution affects speciation, developmental systems drift, phenotypic plasticity, etc. While teams like the JPT team can build that assumption into their models and simulation systems, teams like the WGR team, who study actual

organisms, must eventually substantiate the assumption with empirical work, or flag it for further research. In any case, the teams progressively narrow down the set of putative genes that they study, until they get to a set for which the elements meet both features of their gene concept.

Second, the teams focus on gene products (RNAs and proteins) as the relevant phenotypes for both aspects of their gene concept. That tactic differs from those of earlier strategies to relate the two aspects.

Traditionally for the transmission concept of gene, genes were distinguished from each other by their abilities to affect different adult phenotypes, such as petal color or pea-wrinkledness (Weber 2005, Chapter 7). Alleles were distinguished from each other by their abilities to alter the same adult phenotype, such as the colors of petals. Traditionally for the molecular concept of gene, genes were distinguished from each other by their ability to produce distinct proteins. Researchers assumed both concepts, insofar as they isolated genes, isolated hereditary material that was spatially contiguous. Those who refashioned the molecular concept abandoned that assumption (Waters 1994), and those who refashioned the transmission concept came to see that it was never necessarily part of their concept. But while those who refashioned the molecular concept came to distinguish genes by their ability to produce distinct kinds of RNAs, and maybe even by more precise functions, those who refashioned the transmission concept continued to focus on adult phenotypes, though they sometimes focused on juvenile and embryonic phenotypes related to an organisms physiology or anatomy.

For the two projects studied here, and for both the molecular aspect and the transmission aspect, both teams distinguish genes by their functional relations to RNAs.

The JPT team defines its model of fitness for a GRN on the amount of product produced by the final gene in the network, such that the more product produced by one allele over the other, the more it will dominate the other and impact the developing organism. The WGR team, on the other hand describes the amount of RNA produced in (cohorts of) organisms, it calculates how much those amounts vary across (cohorts of) organisms, and it infers the sources of such variation, including genetic sources. Its GRNs, or at least gene batteries, that have functional relations to embryonic, juvenile, or adult, phenotypes.

So both teams employ a gene concept that features aspects from concepts often thought in tension or incompatible. They do so by distinguishing one gene from another by the kinds of RNAs produced by those genes, and by focusing on genes from known GRNs, and then further focusing on those genes assumed to meet the Mendelian assumptions. For actual organisms, those assumptions are empirically testable. Some have lamented the abstractness of the transmission concept of the gene, but the abstractness here counts in its favor, enabling researchers to find ways to apply it to new systems of study.

This implicit concept of gene is, for both teams, the most direct link between their mechanistic GRN models and their statistical evolutionary genetic models. It enables both teams to construct their $complex_2$ phenomena in such a way that models often thought to be incompatible with each other can have distinct epistemic functional relations to different parts of $complex_2$ phenomena.

7.5- Research Agenda for Developmental Evolution (Devo-Evo)

In this section I address the final driving question of this dissertation: What are the prospects for developmental evolution? I address this question in four different contexts: first as it applies to research projects like the ones studied in this dissertation; second as it applies to integration; third as it applies to debates about the structures and scope of synthesized theories in evo-devo and devo-evo; and finally as it applies to a specific conception of developmental evolution proposed here.

The JPT and WGR Systems and Those Like Them

What are the prospects for the JPT research *system*? There are reasons to think that the system developed by the JPT team will spread to other research teams. First, the team developed it purely theoretically, so experimentalists can interpret it to apply to empirical phenomena. Next, the team showed how their system could relate to many kinds of phenomena that evolutionary geneticists traditionally care about, such us speciation, migration, epistasis, and hybrid incompatibility. Furthermore, they showed how it relates to genetic systems that developmental geneticists increasingly care about, such as gene interactions (physically interpreted epistasis) and actual genetic structures. Finally, the JPT team increasingly specified their system to be empirically applicable in bioenergetic genetic systems.

There are also reasons to think that JPT research system won't spread to other teams. The system depends on researchers' abilities to produce hundred and thousands of generations, or at least enough to present enough lineage evolution to enable hybrid incompatibilities. There are few actual organisms for which biologists have such control,

thus limiting the applicability of the system to organisms like yeast, bacteria, and maybe fruit flies, zebrafish and maybe some quickly germinating plants. Furthermore, research systems often travel to new labs via the migrations of graduated research students. The JPT team has had only one such student, Tulchinsky.

Ultimately, if other teams pick up anything from the JPT team, it likely won't be the whole research system, but just one or two of the models, especially their final model to calculate phenotypes from well-characterized genetic systems. Many researchers study gene structures of just of a few alleles and their interactions, so the JPT evolutionary genetic models likely have more exportability than the whole JPT system. The teams most likely to pick up the whole system are other simulators looking to extend the proof of concept (Servedio et al. 2014).

What are the prospects for the WGR research system? Again, there are reasons for optimism. The JPT team used a lot of common or standardized tools, including the NC-II breeding design, the devices for detecting gene products, and purple sea urchins. As there are large research communities for all of those items, many other teams could adopt most of the WGR system. Furthermore, the Wray lab graduates a decent number of graduate students, who go on to research careers and could take the system with them.

But there are also reasons for pessimism. The system depends on a well-documented GRN, and the best such system is that for sea urchins, which the WGR team already used. Furthermore, their custom Bayesian model, which they used to help estimate and establish the significance effects of different sources of gene expression variation, is complex and requires a research team who has a member trained in Bayesian statistics. Finally, their system relies on access to a facility that can do large scale gene

expression assays. The WGR team had access to one such facility at Duke, but many other research teams won't have that luxury.

Ultimately, the WGR project's results, but not the research system or the models, probably have the most likelihood of being exported. As drosophila GRNs become more accurately described, then the system may export to the drosophila research community. Furthermore, as the sea urchin GRN becomes better described for later developmental stages, and for more embryonic phenotypes, then researchers may again return to the system to study its microevolution. Other teams may also adopt the strategy from (Garfield et al. 2012) of applying statistical tests of selection to the *cis*-regulatory regions of genes, though current citations of the WGR articles indicate that other researchers are more interested in the team's work on gene expression than on their work on selection of regulatory regions. The two WGR articles about gene expression each have three times more citations than the one about selection on regulatory regions.

What are the prospects for *further* projects that develop systems that, like the JPT and WGR projects, use both models of evolutionary genetics and models of GRNs? I think the prospects are good. As a discipline, too many evolutionary biologists use evolutionary genetics for them to stop teaching or using it any time soon. But, as I discuss in the sections below, those biologists would be wise to abandon claims that evolutionary genetics comprises any kind of theoretical core to evolutionary biology. Most evolutionary phenomena are too complex to admit of a single theoretical core.

But my above judgment has flaws. Many argue against the use of evolutionary genetics because of its supposed inability to help us understand most macro-evolutionary phenomena. They note that we understand those phenomena not when we use theories

and models of evolutionary genetics, but when we use models of developmental mechanisms like GRN models with comparative analyses across higher taxa. The most we can use evolutionary genetic models for, some argue, is to describe and predict the same old microevolutionary phenomena they've always described and explained, phenomena like the diffusion of petal colors through a population of geraniums.

That argument, however, undersells the importance of developmental mechanisms and especially GRNs. Those who study macro-evolution may not need evolutionary genetics models, but as projects like the WGR and JPT projects show, those who study microevolution and speciation can use developmental mechanisms and GRNs to better understand microevolutionary phenomena.

Some argue that evolutionary genetics has limited explanatory scope or power (Wagner 2000; Amundson 2005). They argue that the kinds of models deployed in evo-devo or devo-evo, because they apply to higher taxa, have greater scope and therefore more explanatory power, and ultimately, are more valuable. That argument assumes that causal explanations of phenotypes are the only epistemic aims worth valuing, or at least are the most valuable. That assumption at the very least needs an argument to support it, but regardless, there are good reasons to argue against it. For the problems that face farmers, breeders, conservationists, zoos, and sustainability scientists, evolutionary genetics will continue to be highly valued, even if that field never provides a true causal explanation of the phenomena they study, but still enables them to ameliorate problems.

What are the prospects for the gene concept common to the WGR and JPT teams? I think the prospects for the concept are quite good. In many cases, it's likely what many biologists have roughly in mind when they talk about genes. Furthermore, it enables

researchers to connect evolutionary genes to phenotypes, even though those connections will be increasingly mediated between GRN and other developmental genetic models, rather than through traditional genotype-phenotype maps.

How might we interpret the materiality of the gene picked out by the gene concept? Researchers once thought of chromosomes as made of genes arranged in a bead-like manner. Sixty-five years of molecular genetics research destroyed that hypothesis. Perhaps a more fruitful way to think of genes is the reverse. Genes don't comprise chromosomes, rather, genes are comprised of bits of chromosomes, bits that needn't be contiguous. In this sense, chromosomes remain the material units of much of heredity, but genes are the functional units. If a chromosome mutates in such a way that the functional unit partly comprised by that chromosome doesn't follow the laws of assortment and segregation, then that functional unit isn't functional unit of heredity, even if it is a functional unit of development. If an offspring can't inherit that gene, then it can't spread, via descent, in the population.

Integration

Many folks, from biologists to historians to philosophers, have discussed integration of evolutionary and developmental biology (Mitchell and Dietrich 2006; Brigandt 2010; Brigandt and Love 2010). What are the prospects for integrations, and for discussion of integration? I addressed the first part of that question in the previous section. Below, I address the later and indicate how we might fruitfully address the issue of integration.

Much of the discussion about integration has unclear aims. Some discuss how such integrations would be new in some sense (Burian 1986; Brigandt 2010; Wagner 2010), others discuss how they would be retro in some sense (Morange 2011). Regardless of the answer, the theoretical or investigative aims of integration, let alone novel or retro integrations, remain unclear. Often, researchers use the topic of integration to wedge favored phenomena or scientists into emerging fields to legitimate those phenomena or scientists. As a result, evo-devo has increasingly come to mean just about anything to just about anyone.

Those discussions about the newness or retro-ness of integrations in evo-devo have limited theoretical or investigative import. Researchers make broad claims about what questions such integrations will help us answer. But often they stop short of saying which non-intellectual problems they can solve, or what epistemic aims they will achieve, other than that of mechanistic explanation, implicitly taken as the only or best kind of understanding.

Furthermore, insofar as those discussions focus on general questions, they suggest only the most lofty elements of research agendas (Love 2008) or of significance graphs (Kitcher 2001), which are structures of questions that researchers in a field pursue. In research agendas and significance graphs, the most general questions motivate the field, and they imply or otherwise indicate increasingly specific questions, which ultimately motivate specific research projects. Significance graphs form tree or bush-like structures or networks, in which the questions are nodes and the relations between them are edges. But while those in the debate propose an ever increasing list of general research questions, rarely do they provide strategies for inferring experimentally tractable

questions, or of connecting actual experimental or empirical questions their proposed general nodes in a significance graph network.

One source of that shortcoming is that those discussions have become debates in of themselves, divorced from research of recent scientific projects. As such, those debates often operate as vacuum chambers in which folks argue about what questions scientists can or cannot address, rather than what questions they actually aim to address, have addressed, or have attempted to address but have failed.

But if we study actual research projects in depth, we find articles like Greg Wray's 2013 review (Wray 2013). In it, he details no fewer than 15 questions, tying some to proposed general aims for the field, and tying most others to actual research projects and the specific questions they pursued. That review article, and others like it that similarly draw from the aims and conclusions of actual projects, provide excellent sources for those who wish to describe and debate the research agendas for fields that integrate evolutionary biology and developmental biology. They provide the questions actually pursued, and a next step would be to structure those questions into significance graphs, evaluate how well individual questions have been addressed or not, explore why groups of related questions have been addressed while other groups have been ignored or have resisted being answered, and indicate which areas of the significance graph are conspicuously sparse.

A second source of the shortcoming, I suggest, is that many folks focus on theories or concepts as the items integrated, and they do so in an unsystematic way. Insofar as they do so, they focus on general questions that they take theories to address, and not on the specific questions pursued at the leading edge of research. Ron Amundson,

who focused on theories and concepts, acknowledges the limitations of that focus. After arguing that evo-devo is a distinct field (both novel and retro) because its fundamental concepts are incompatible with those of population-focused evolutionary biology, he concludes: "If both evo-devo and population genetics continue to be successful, a way will somehow be found to see them as consistent," (Amundson 2005, 257).

My case studies indicate that, for ongoing projects, there are many kinds of pieces that researchers must integrate to pursue their aims. Those pieces include theories, models, and concepts; but also methods or practices or protocols to prepare phenomena, collect data, and analyze it; problems, questions, values, and epistemic aims; research teams and research locations; communication channels, etc. To understand how researchers integrate theories and other scientific products, which is ultimately an empirical issue, we must systematically study how they integrate all of the above pieces in their research projects.

More importantly, and as I indicated earlier, a fruitful unit by which to study integration is at the level of emerging research systems. Such systems incorporate many of the kinds of pieces listed above. I suggest that study of these units provides more insight into the evolution of ongoing science than do scientific products taken generally and abstractly. At the leading edge of fields that integrate developmental and evolutionary biology, the products are under constant revision, and as Amundson recognized, claims about the possibility or impossibility of integrations between those products face ever changing goal posts.

Ultimately, we can't evaluate such claims without attending to actual scientific projects, their research systems, and their rationales. I suggest that the theories that

survive scientific scrutiny won't be those that we can first integrate and then apply to a range of phenomena. Rather, much of the competition is between research systems, and the theories or models that survive will be those that perform at least one epistemic function in relation to the complex phenomena that partly comprise the objects of those systems. While some scientists and philosophers work to integrate developmental and evolutionary biology at the level of theories, perhaps a more prevalent strategy among scientists is to construct complex$_2$ phenomena in research systems. For them, theories, models, and other scientific products need only be non-contradictory with each other, apply to at least one aspect or component of a complex phenomenon, and perform at least one epistemic function (explanation, descriptions, prediction, etc.) in regards to that aspect.

So how might those of us who study the integration of evolutionary biology and developmental biology fruitfully study integration? First, we should focus on research systems with complex phenomena and multiple scientific products, not just on scientific products. Second, we should study actual research projects, especially at the leading edge of research. But we can also fruitfully address research agendas by identifying and explicating the evolution of significance graphs, not just of questions, but also if non-intellectual problems and epistemic aims. We can identify and explicate practices and methods and the integration of investigative protocols and strategies. And we can address the social phenomena of how scientists integrate research teams, funding streams, and communication channels. If we study those topics empirically, we can describe and explain how scientists actually integrate different aspects of developmental and evolutionary biology.

Structure of Evo-Devo

Many philosophers and theoreticians argue with each other about the structure of evo-devo and of devo-evo in comparison to traditional evolutionary theory as developed in the modern synthesis. Some competing positions argue that evo-devo is:

1. an interfield theory, or a collections of such theories: discussed hypothetically by several authors in (Bechtel 1986).
2. an extension to the modern synthesis: (Müller 2007; Pigliucci 2008; Pigliucci and Müller 2010; Laland et al. 2014; Laland et al. 2015)
3. a theoretical framework distinct from the modern synthesis (in good ways): (Amundson 2005; Craig 2009; Craig 2010; Craig 2011; Craig 2015).
4. a theoretical framework distinct from the modern synthesis and from devo-evo, the last of which is more fruitful: (Hall 2000; Davidson 2011; Laubichler 2010; Wagner 2000; Wagner et al. 2001; Wagner 2014).
5. not distinct from traditional evolutionary theory, and insofar as it is described as such, the descriptions are confused: (Hoekstra and Coyne 2007; Lynch 2007; Wray et al. 2014).

What are the prospects for the above debate? In terms of the debate continuing, there is much grist for the mill, and so the prospects are good. In terms of settling the debate, the prospects are bad. Regardless of the intended aims of the above authors, none of the above claims has been shown to be accurate representations about the theories in the developing field(s). That's not to say that one or more of them can't be true. Rather, the issue is an empirical one, not a purely conceptual one. To address it, we need to better

describe the epistemological structures of fields. While many above discuss supposed relations between theories and models, my conclusions above indicate that, if we want accurate descriptions, we should also look at other kinds of items. Those items include the research systems constructed by those in the field, the $complex_2$ phenomena that party comprise those systems, and the specific epistemic functional relations between different models (and kinds of models) and parts of those $complex_2$ phenomena (and kinds of phenomena).

Rather than continuing the above debate, there is another path when examining models and theories. Love has noted that theorists at the intersection of developmental biology and evolutionary biology offer theories he calls material inference structures (MISs) (Love 2013). These are collections of abstract generalizations that, depending on the research goals of research teams at a given time, lead the teams to develop tools to understand phenomena, tools that might be inconsistent with each other across research projects. I disagree with many of the details of Love's account, but I agree with the overall thrust.[5]

More is to be gained by following Love's strategy. Love highlights the set of generalizations proposed by (Carroll 2008). But other theorists have proposed distinct sets of generalizations. Eric Davidson proposed a set of generalizations specific GRNs (Davidson 2011), and a group of scientists and philosophers proposed one for evo-devo more broadly (Laland et al. 2015). Rather than debating the amount of difference these systems have in relation to the modern synthesis, theorists and philosophers can improve

[5] I think Love's account of MISs and their relations to models is a special case of the theory-model relationship described in (Krakauer et al. 2011).

the study of nature by focusing on these sets of generalizations, clarifying them, and showing that they function (or don't) in successful research.

Developmental Evolution

What are the prospects for developmental evolution (devo-evo)? I conclude this dissertation by arguing that the prospects are quite good, given the account of developmental evolution proposed below. For now, when I talk about developmental evolution, I assume only that it is an emerging field at the interface of developmental biology and evolutionary biology, and that it substantially differs, and will continue to differ, from the other primary field developing at that interface, evo-devo, and from the modern synthesis. By the end of this section, I propose a more specific account of devo-evo.

For its primary proponents, giving an account of developmental evolution has proven tricky. Brian Hall asserted that devo-evo was a distinct field, but said little about how (Hall 2000). Afterwards, proponents distinguished devo-evo from other fields via two related strategies. They focused on kinds of phenomena to be explained, or they focused on the kinds of questions to be answered.

Folks focused on kinds of phenomena, most commonly on evolutionary novelties, especially of morphology (Wagner 2000; Love 2008; Brigandt and Love 2010). Such novelties include the tetrapod limb and avian feathers. Such phenomena occur above the species level of taxonomy, and so their evolution is called macroevolution. Researchers proposed to classify those kinds of phenomena as developmental types (Amundson 2005), a concept that made it easier to conceptualize morphological homologies above

the species level also as products of evolution, and to add those to the explanatory agenda of devo-evo (Wagner 2014). Over the years, folks have posed myriad questions that could be collected into a research agenda for devo-evo (Amundson 2005; Love 2008; Laubichler 2010; Davidson 2011).

From that work, we might characterize devo-evo as:

> a field in which researchers use developmental mechanisms to explain the evolution of developmental types, a class of phenomena that includes morphological novelties and homologies.

The prospects for the field characterized above are good, but proponents of devo-evo should aim for more. Those proponents argue that devo-evo has the potential to significantly differ from evo-devo and from traditional evolutionary biology. Insofar as they set a different research agenda from those fields, as they currently do, then they will ensure that the fields do differ. But in their bolder moments, those proponents argue that devo-evo has the potential to obviate evo-devo and to *supplant* traditional evolutionary biology (Laubichler 2010; Davidson 2011). How might we reframe the research agenda of devo-evo so that it can do so?

Evo-devo faces some epistemological issues. Many folks want it to be a theoretical system that fits everyone's needs, from research to education to healthcare (Moczek et al. 2015). Partly as a result, there is a diminished focus on evolution, both empirically and theoretically, in leading-edge evo-devo research, as many researchers focus increasingly on developmental genetics (Diogo 2016). Perhaps the best route forward for evo-devo is as a trading zone that permits researchers to exchange epistemic

tools, research systems, and the like (Winther 2015). But for evo-devo, it seems no unique theory of knowledge is in the offing. The case is different for devo-evo.

To supplant traditional evolutionary biology, devo-evo proponents might pursue the following tactic. In addition to their agreed agenda about developmental types and macroevolution, they might adopt as much of the agenda of traditional evolutionary biology as possible, and show that the epistemic tools and strategies of devo-evo achieve that agenda better than do those of the traditional field. That is, even for microevolutionary phenomena, when we use models of developmental mechanisms, like GRN models, we understand that phenomena better than if we hadn't used those models.

That tactic has the potential to unite researchers from the five camps listed in the previous section. Some devo-evo proponents are already proposing epistemic tools and strategies that adopt this tactic (Atchley and Hall 1991; Stern 2010; Pavlicev and Wagner 2012). The two projects studied in this dissertation also adopt it. While the JPT team would likely endorse the revolutionary aspect of devo-evo, Wray of the WGR team explicitly refutes it (Wray et al. 2014). But both pursue the same strategy of using developmental mechanisms to better understand microevolutionary phenomena.

Those who promote devo-evo have severely criticized evolutionary genetic theory and models. As a group, they argue that genetic evolution is a poor proxy for phenotypic evolution, that we can't use those models to understand macroevolutionary phenomena besides speciation, that those models can't capture the number of genes found in GRNs, and that they can't capture evolution to genes that causes those genes to change their structure and functions. Therefore, they argue, those theories and models can't provide the foundation for a general theory of evolution. Fair enough.

Those arguments show that we shouldn't privilege evolutionary genetic models and microevolutionary phenomena as the core of evolutionary biology, not that we shouldn't value them at all. Talk of theoretical cores for evolution led to the debate discussed in the previous section. To understand complex evolutionary phenomena, researchers could look to assemble unified or synthetized theories, but for now they're doing what they always do: using whatever theoretical tools they can get to help them understand the phenomena they study.

To take over the research agenda of traditional evolutionary biology, devo-evo proponents can avoid talk of theoretical cores entirely. They can instead focus on the theories that work, that help us understand phenomena, theories that are distinct from each other for different parts of complex evolutionary phenomena. When it comes to integrating or synthesizing models for such phenomena, those scientific products need only be consistent with each other in relation to their specific phenomena.

Proponets of devo-evo needlessly limit the scope of their field, which decreases the likelihood that devo-evo supplants traditional evolutionary biology. They do so by privileging the epistemic aim of mechanistic explanation. They also do so by focusing as much as they do on macroevolutionary phenomena. Finally, they do so by building their research questions and agendas around that aim and those phenomena.

In the long run, if devo-evo adopts the research agenda of evolutionary genetics, it sets itself up to develop a more general theory of genetic inheritance. For such a theory, phenomena once understood by models that assumed Mendel's laws would become special cases of the more general theory. Evolutionary genetics can do little with macroevolutionary phenomena. But it also struggles with a wide range of transmission

phenomena at the microevlotuionary scale. The time is ripe to develop a theory that can account for those phenomena, and for Mendelian phenomena as a special case.

I conclude that devo-evo should broaden its horizons. For epistemic aims, it should countenance projects that pursue many distinct aims, such as discovery, description, prediction, in addition to explanation. Second, for phenomena, it should countenance projects that study macroevolutionary and microevolutionary phenomena, complex or not. And for research agendas, it should countenance as many research problems and research questions as its practitioners can make precise and address.

Given those claims, I characterize developmental evolution as:

> a field in which researchers use scientific products about (at least) developmental mechanisms to understand evolutionary phenomena.

Described as such, devo-evo researchers can absorb many theories and many research agendas from traditional evolutionary biology, and use them in a broad field without a single theoretical core.

It is according to this characterization of devo-evo, which succinctly captures the inclusive spirit of (Wagner et al. 2001), that I think the prospects for a revolutionary biology are greatest.

REFERENCES

Amundson, Ron. 2005. *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo.* New York: Cambridge University Press.

Andersen, Holly. 2014a. "A Field Guide to Mechanisms: Part I." *Philosophy Compass* 9: 274–83. doi:10.1111/phc3.12119.

———. 2014b. "A Field Guide to Mechanisms: Part II." *Philosophy Compass* 9: 284–93. doi:10.1111/phc3.12118.

Ankeny, Rachel A. 2012. "Detecting Themes and Variations: The Use of Cases in Developmental Biology." *Philosophy of Science* 79: 644–54.

———. 2014. "The Overlooked Role of Cases in Casual Attribution in Medicine." *Philosophy of Science* 81: 999–1011. doi:10.1086/677693.

Atchley, William R., and Brian K. Hall. 1991. "A Model for Development and Evolution of Complex Morphological Structures." *Biological Reviews* 66: 101–57. doi:10.1111/j.1469-185X.1991.tb01138.x.

Babbitt, Courtney Christine. 2005. "Developmental Systematics: Synthesizing Ontogeny and Phylogeny in the Malacostraca (Crustacea)." PhD Dissertation: The University of Chicago.

Balhoff, James P., and Gregory A. Wray. 2005. "Evolutionary Analysis of the Well Characterized endo16 Promoter Reveals Substantial Variation within Functional Sites." *PNAS* 102: 8591–96.

Bazerman, Charles. 1981. "What Written Knowledge Does: Three Examples of Academic Discourse." *Philosophy of the Social Sciences* 11: 361–387.

Beatty, John. 2016. "What Are Narratives Good For?" *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 58: 33–40. doi:10.1016/j.shpsc.2015.12.016.

Becher, Tony, and Paul Trowler. 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. London: McGraw-Hill.

Bechtel, William, ed. 1986. *Integrating Scientific Disciplines*. Dordrecht: Martinus Nijhoff.

———. 1986. "The Nature of Scientific Integration." In *Integrating Scientific Disciplines*, edited by William Bechtel, 3–52. Dordrecht: Martinus Nijhoff.

Beckner, Morton. 1959. *The Biological Way of Thought*. New York: Columbia University Press.

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *Philosophical Review* 97: 303–52.

Brigandt, Ingo. 2010a. "Beyond Reduction and Pluralism: Toward an Epistemology of Explanatory Integration in Biology." *Erkenntnis* 73: 295–311. doi:10.1007/s10670-010-9233-3.

———. 2010b. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177: 19–40. doi:10.1007/s11229-009-9623-8.

———. 2012. "The Dynamics of Scientific Concepts: The Relevance of Epistemic AIms and Values." In *Scientific Concepts and Investigative Practice*, edited by Uljana Feest and Friedrich Steinle, 75–104. Berlin: Walter de Gruyter.

———. 2015. "Social Values Influence the Adequacy Conditions of Scientific Theories." *Canadian Journal of Philosophy* 45: 326–56.

Brigandt, Ingo, and Alan C. Love. 2010. "Evolutionary Novelty and the Evo-Devo Synthesis: Field Notes." *Evolutionary Biology* 37: 93–99. doi:10.1007/s11692-010-9083-6.

Bromberger, Sylvain. 1966. "Questions." *The Journal of Philosophy* 63: 597–606. doi:10.2307/2024253.

Brooke, John Hedley. 1981. "Avogadros Hypothesis and Its Fate: A Case-Study in the Failure of Case-Studies." *History of Science* 19: 235–73.

Brown, Cecelia. 2010. "Communication in the Sciences." *Annual Review of Information Science and Technology* 44: 285–316.

Burian, Richard M. 1986. "On Integrating the Study of Evolution and of Development." In *Integrating Scientific Disciplines*, edited by William Bechtel, 209–228. Dordrecht: Martinus Nijhoff.

———. 2001. "The Dilemma of Case Studies Resolved: The Virtues of Using Case Studies in the History and Philosophy of Science." *Perspectives on Science* 9: 383–404.

Butcher, Lee M., Oliver S.P. Davis, Ian W. Craig, and Robert Plomin. 2008. "Genome-Wide Quantitative Trait Locus Association Scan of General Cognitive Ability

Using Pooled DNA and 500K Single Nucleotide Polymorphism Microarrays." *Genes, Brain and Behavior* 7: 435–446.

Caniglia, Guido. 2010. "Post-Experimental Phenomena: An Epistemological Study on Evolutionary Innovations." PhD Dissertation. Firenze, Italia: Università degli Studi di Firenze.

Carnap, Rudolf. 1950. "On Explication." In *The Logical Foundation of Probability*, 2nd ed., 1–18. Chicago: University of Chicago Press.

Carroll, Sean B. 2008. "Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution." *Cell* 134: 25–36. doi:10.1016/j.cell.2008.06.030.

Carter, Ashley J. R., Joachim Hermisson, and Thomas F. Hansen. 2005. "The Role of Epistatic Gene Interactions in the Response to Selection and the Evolution of Evolvability." *Theoretical Population Biology* 68: 179–96. doi:10.1016/j.tpb.2005.05.002.

Cartwright, Nancy. 2007. *Hunting Causes and Using Them*. New York: Cambridge University Press.

Cassirer, Ernst. 1923. *Substance and Function*. Translated by William Curtis Swabey and Marie Collins Swabey. New York: Dover.

———. 1950. *The Problem of Knowledge: Philosophy, Science, and History Since Hegel*. Translated by William H. Woglom and Charles W. Hendel. New Haven: Yale University Press.

Chang, Hasok. 2012. "Beyond Case-Studies: History as Philosophy." In *Integrating History and Philosophy of Science*, edited by S. Mauskopf and T. Schmaltz, 263:109–24. Dordrecht: Springer.

Coyne, Jerry A, and H. Allen Orr. 2004. *Speciation*. Sunderland, Mass.: Sinauer.

Craig, Lindsay R. 2009. "Defending Evo-Devo: A Response to Hoekstra and Coyne." *Philosophy of Science* 76 (3): 335–44.

———. 2010. "The So-Called Extended Synthesis and Population Genetics." *Biological Theory* 5: 117–23.

———. 2011. "Criticism of the Extended Synthesis: A Response to Muller and Pigliucci." *Biological Theory* 5: 395–96.

———. “Neo-Darwinism and Evo-Devo: An Argument for Theoretical Pluralism in Evolutionary Biology.” *Perspectives on Science* 23: 243–79. doi:10.1162/POSC_a_00167.

Crasnow, Sharon. 2011. “Evidence for Use: Causal Pluralism and the Role of Case Studies in Political Science Research.” *Philosophy of the Social Sciences* 41: 26–49. doi:10.1177/0048393110387884.

———. 2012. “The Role of Case Study Research in Political Science: Evidence for Causal Claims.” *Philosophy of Science* 79: 655–66. doi:10.1086/667869.

Craver, Carl F. 2006. “When Mechanistic Models Explain.” *Synthese* 153: 355–76. doi:10.1007/s11229-006-9097-x.

Craver, Carl F. 2007. *Explaining the Brain*. New York: Oxford University Press.

Craver, Carl, and James Tabery. 2015. “Mechanisms in Science.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2015. http://plato.stanford.edu/entries/science-mechanisms/.

Creath, Richard. 1994. “Functionalist Theories of Meaning and the Defense of Analyticity.” In *Logic, Language, and the Structure of Scientific Theories*, edited by Wesley C Salmon and Wolters, 287–304. Pittsburgh: University of Pittsburgh Press.

———. 2010. “The Role of History in Science.” *Journal of the History of Biology* 43: 207–14. doi:10.1007/s10739-009-9208-x.

Crombie, A. C. 1994. *Styles of Scientific Thinking in the European Tradition: The History of Argument and Explanation Especially in the Mathematical and Biomedical Sciences and Arts*. 3 vols. London: Duckworth.

Currie, Adrian. 2015. “Philosophy of Science and The Curse of the Case Study.” In *The Palgrave Handbook of Philosophical Methods*, edited by Christopher Daly, 553–72. London: Palgrave Macmillan.

Darden, Lindley, and Nancy Maull. 1977. “Interfield Theories.” *Philosophy of Science* 44: 43–64.

Davidson, Eric H. 2006. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. San Diego: Academic Press.

———. 2011. “Evolutionary Bioscience as Regulatory Systems Biology.” *Developmental Biology* 357: 35–40. doi:10.1016/j.ydbio.2011.02.004.

Diogo, Rui. 2016. “Where Is the Evo in Evo-Devo (Evolutionary Developmental Biology)?” *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 326: 9–18. doi:10.1002/jez.b.22664.

Donovan, Arthur L, Larry Laudan, and Rachel Laudan. 1988. *Scrutinizing Science: Empirical Studies of Scientific Change*. Dordrecht: Kluwer Academic Publishers.

Douglas, Heather. 2013. “The Value of Cognitive Values.” *Philosophy of Science* 80: 796–806. doi:10.1086/673716.

Duhem, Pierre. 1914. *The Aim and Structure of Physical Theory*. Translated by Philip P. Wiener. 1991st ed. Princeton Science Library. Princeton, N.J.: Princeton University Press.

Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Eisenhardt, Kathleen M. 1989. “Building Theories from Case Study Research.” *The Academy of Management Review* 14: 532–50.

———. 1991. “Better Stories and Better Constructs: The Case for Rigor and Comparative Logic.” *Academy of Management Review* 16: 620–27. doi:10.5465/AMR.1991.4279496.

Elliot, Andrew J., and James W. Fryer. 2008. “The Goal Construct in Psychology.” In *Handbook of Motivation Science*, edited by James Y. Shah and Wendi L. Gardner, 235–50. New York: Guilford Press.

Elliott, Steve. 2016. “Problems and Questions in Scientific Practice.” In . Atlanta, GA: PhilSci Archive. http://philsci-archive.pitt.edu/12555/.

———. “A Budget of Problems for Case Study Research in HPS.” *In Preperation.*

Evans, James A., and Jacob G. Foster. 2011. “Metaknowledge.” *Science* 331: 721–25. doi:10.1126/science.1201765.

Falk, Rapheal. 2000. “The Gene - A Concept in Tension.” In *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, edited by Peter J. Beurton and Raphael Falk, 317–48. Cambridge University Press.

———. 2010. “What Is a gene?—Revisited.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, Culture-bound syndromes, 41 (4): 396–406. doi:10.1016/j.shpsc.2010.10.014.

Fallis, Don. 2007. “Collective Epistemic Goals.” *Social Epistemology* 21: 267–80.

Faust, David, and Paul E. Meehl. 1992. “Using Scientific Methods to Resolve Questions in the History and Philosophy of Science: Some Illustrations.” *Behavior Therapy* 23: 195–211. doi:10.1016/S0005-7894(05)80381-8.

———. 2002. “Using Meta-Scientific Studies to Clarify or Resolve Questions in the Philosophy and History of Science.” *Philosophy of Science* 69: S185–S196.

Fleck, Ludwik. 1935. *Genesis and Development of Scientific Fact*. Edited by Thaddeus J. Trenn, Robert K. Merton, and Frederick Bradley. Chicago: University of Chicago Press.

Flyvbjerg, Bent. 2006. “Five Misunderstandings about Case-Study Research.” *Qualitative Inquiry* 12: 219–45. doi:10.1177/1077800405284363.

Forber, Patrick. 2010. “Confirmation and Explaining How Possible.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 41: 32–40. doi:10.1016/j.shpsc.2009.12.006.

Forrester, John. 1996. “If P, Then What? Thinking in Cases.” *History of the Human Sciences* 9: 1–25. doi:10.1177/095269519600900301.

Frickel, Scott, and Neil Gross. 2005. “A General Theory of Scientific/Intellectual Movements.” *American Sociological Review* 70: 204–32. doi:10.1177/000312240507000202.

Friedman, Michael. 1974. “Explanation and Scientific Understanding.” *The Journal of Philosophy* 71: 5–19. doi:10.2307/2024924.

Frigg, Roman, and Stephan Hartmann. 2016. “Models in Science.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/models-science/.

Garfield, David Aaron. 2011. “Selection and Constraint: Population Genetic Approaches to Understanding the Evolution of Sea Urchin Development.” PhD Dissertation, Durham, North Carolina: Duke University.

Garfield, David, Ralph Haygood, William J. Nielsen, and Gregory A. Wray. 2012. “Population Genetics of Cis-Regulatory Sequences That Operate during Embryonic Development in the Sea Urchin Strongylocentrotus Purpuratus.” *Evolution & Development* 14: 152–67. doi:10.1111/j.1525-142X.2012.00532.x.

Garfield, David A., Daniel E. Runcie, Courtney C. Babbitt, Ralph Haygood, William J. Nielsen, and Gregory A. Wray. 2013. “The Impact of Gene Expression Variation on the Robustness and Evolvability of a Developmental Gene Regulatory Network.” *PLOS Biology* 11: e1001696. doi:10.1371/journal.pbio.1001696.

Garfield, David A., and Gregory A. Wray. 2009. “Comparative Embryology without a Microscope: Using Genomic Approaches to Understand the Evolution of Development.” *Journal of Biology* 8 (7): 1.

———. 2010. “The Evolution of Gene Regulatory Interactions.” *BioScience* 60: 15–23. doi:10.1525/bio.2010.60.1.6.

Garvey, William D., and Griffith, Belver C. 1967. “Science Communication as a Social System.” *Science* 157: 1011–16.

Garvey, William D., and Belver C. Griffith. 1971. “Scientific Communication: Its Role in the Conduct of Research and Creation of Knowledge.” *American Psychologist* 26: 349–62.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Gerring, John. 2004. “What Is a Case Study and What Is It Good For?” *American Political Science Review* 98: 341–54.

———. 2007. *Case Study Research: Principles and Practices*. New York: Cambridge University Press.

———. 2012. “Mere Description.” *British Journal of Political Science* 42: 721–46. doi:10.1017/S0007123412000130.

Giere, Ronald N. 1988. *Explaining Science*. Chicago: University of Chicago Press.

———. 2004. “How Models Are Used to Represent Reality.” *Philosophy of Science* 71: 742–52.

Gilbert, Scott F. 2003. “Evo-Devo, Devo-Evo, and Devgen-Popgen.” *Biology and Philosophy* 18: 347–352.

Gillespie, John H. 2004. *Population Genetics: A Concise Guide*. 2nd ed. Baltimore: Johns Hopkins University Press.

Glymour, Bruce. 2006. “Wayward Modeling: Population Genetics and Natural Selection.” *Philosophy of Science* 73: 369–89. doi:10.1086/516805.

Goldman, Alvin. 2002. “The Unity of the Epistemic Virtues.” In *Pathways to Knowledge*, 51–72. New York: Oxford University Press.

Griffiths, Paul E., and Eva M. Neumann-Held. 1999. “The Many Faces of the Gene.” *BioScience* 49 (8): 656–62. doi:10.2307/1313441.

Grünbaum, Adolf. 1988. “The Role of the Case Study Method in the Foundations of Psychoanalysis.” *Canadian Journal of Philosophy* 18: 623–58.

Gusfield, Joseph. 1976. “The Literary Rhetoric of Science: Comedy and Pathos in Drinking Driver Research.” *American Sociological Review* 41: 16–34. doi:10.2307/2094370.

Haag, Eric S. 2006. “Compensatory vs. Pseudocompensatory Evolution in Molecular and Developmental Interactions.” *Genetica* 129: 45–55. doi:10.1007/s10709-006-0032-3.

———. 2014. “The Same but Different: Worms Reveal the Pervasiveness of Developmental System Drift.” *PLoS Genetics* 10: e1004150. doi:10.1371/journal.pgen.1004150.

Haag, Eric S., and John R. True. 2007. “Evolution and Development: Anchors Away!” *Current Biology* 17: R172–74. doi:10.1016/j.cub.2007.01.015.

Hacking, Ian. 1992. “‘Style’ For Historians and Philosophers.” *Studies in History and Philosophy of Science* 23: 1–20.

Hall, Brian K. 2000. “Guest Editorial: Evo-Devo or Devo-Evo — Does It Matter?” *Evolution and Development* 2: 177–78.

Hansen, Thomas F. 2006. “The Evolution of Genetic Architecture.” *Annual Review of Ecology, Evolution, and Systematics* 37: 123–57.

Hansen, Thomas F., José M. Álvarez-Castro, Ashley J. R. Carter, Joachim Hermisson, Günter P. Wagner, and S. Otto. 2006. “Evolution of Genetic Architecture under Directional Selection.” *Evolution* 60 (8): 1523–36. doi:10.1554/06-093.1.

Hansen, Thomas F., and Günter P. Wagner. 2001a. “Epistasis and the Mutation Load: A Measurement-Theoretical Approach.” *Genetics* 158: 477–85.

———. 2001b. “Modeling Genetic Architecture: A Multilinear Theory of Gene Interaction.” *Theoretical Population Biology* 59: 61–86. doi:10.1006/tpbi.2000.1508.

Hanson, Norwood Russell. 1958. *Patterns of Discovery*. New York: Cambridge University Press.

Hardcastle, Gary. 1999. "Are There Scientific Goals?" *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 30: 297–311.

Harwood, Jonathan. 1993. *Styles of Scientific Thought: The German Genetics Community, 1900-1933*. Chicago: University of Chicago Press.

Haufe, Chris. 2013. "Why Do Funding Agencies Favor Hypothesis Testing?" *Studies in History and Philosophy of Science Part A* 44: 363–74. doi:10.1016/j.shpsa.2013.05.002.

Haygood, Ralph McMillan. 2002. "Three Contributions to Theoretical Population Biology." PhD Dissertation, Davis, California: University of California, Davis.

Hempel, Carl G. 1952. *Fundamentals of Concept Formation in Empirical Science.pdf*. Vol. 2. International Encyclopedia of Unified Science. Chicago: University of Chicago Press.

———. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15: 135–75.

Hermisson, Joachim, Thomas F. Hansen, and Günter P. Wagner. 2003. "Epistasis in Polygenic Traits and the Evolution of Genetic Architecture under Stabilizing Selection." *The American Naturalist* 161: 708–34. doi:10.1086/374204.

Hesse, Mary B. 1966. *Models and Analogies in Science*. Notre Dame: University of Notre Dame Press.

Hoekstra, Hopi E, and Jerry A Coyne. 2007. "The Locus of Evolution: Evo Devo and the Genetics of Adaptation." *Evolution* 61: 995–1016. doi:10.1111/j.1558-5646.2007.00105.x.

Hull, David L. 1988. *Science as a Process*. Chicago: University of Chicago Press.

Humphreys, Paul. 1989. *The Chances of Explanation*. Princeton, N.J.: Princeton University Press.

Hyland, Ken, and Françoise Salager-Meyer. 2008. "Scientific Writing." *Annual Review of Information Science and Technology* 42: 297–338.

Israel, Jennifer W., Megan L. Martik, Maria Byrne, Elizabeth C. Raff, Rudolf A. Raff, David R. McClay, and Gregory A. Wray. 2016. "Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during a Major Life History Switch in the Sea Urchin Genus Heliocidaris." *PLOS Biol* 14: e1002391. doi:10.1371/journal.pbio.1002391.

Israel, Jennifer Wygoda. 2015. "Sea Urchin Body Plan Development and Evolution: An Integrative Transcriptomic Approach." PhD Dissertation, Durham, North Carolina: Duke University.

Jenkins, D. L., C. A. Ortori, and J. F. Y. Brookfield. 1995. "A Test for Adaptive Change in DNA Sequences Controlling Transcription." *Proceedings of the Royal Society of London B: Biological Sciences* 261: 203–7. doi:10.1098/rspb.1995.0137.

Johnson, Norman A. 1992. "Genetics of Hybrid Male Sterility in Three Sibling Species of the Drosophila Melanogaster Species Subgroup." PhD Dissertation, Rochester, New York: University of Rochester.

———. 2007. *Darwinian Detectives*. New York: Oxford University Press.

———. 2007. "The Micro-Evolution of Development." *Genetica* 129: 1–5. doi:10.1007/s10709-006-0028-z.

———. 2015. "Personal Communications."

Johnson, Norman A, and Adam H Porter. 2000. "Rapid Speciation via Parallel , Directional Selection on Regulatory Genetic Pathways." *Journal of Theoretical Biology* 205: 527–42. doi:10.1006/jtbi.2000.2070.

———. 2001. "Toward a New Synthesis: Population Genetics and Evolutionary Developmental Biology." *Genetica* 112–113: 45–58.

———. 2007. "Evolution of Branched Regulatory Genetic Pathways: Directional Selection on Pleiotropic Loci Accelerates Developmental System Drift." *Genetica* 129: 57–70. doi:10.1007/s10709-006-0033-2.

Killcoyne, Sarah, Gregory W. Carter, Jennifer Smith, and John Boyle. 2009. "Cytoscape: A Community-Based Framework for Network Modeling - Springer." In *Protein Networks and Pathway Analysis*, edited by Yuri Nikolsky and Julie Bryant, 219–39. Methods in Molecular Biology 563. New York: Humana Press.

King, Gary, Robert O Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J.: Princeton University Press.

Kinzel, Katherina. 2015. “Narrative and Evidence. How Can Case Studies from the History of Science Support Claims in the Philosophy of Science?” *Studies in History and Philosophy of Science Part A* 49: 48–57. doi:10.1016/j.shpsa.2014.12.001.

Kitcher, Philip. 1984. “1953 and All That. A Tale of Two Sciences.” *Philosophical Review* 93 (3): 335–73.

———. 1989. “Explanatary Unification and the Causal Structure of the World.” In *Scientific Explanation*, edited by Philip Kitcher and Wesley Salmon, 8:410–505. Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press.

———. 1993. *The Advancement of Science*. New York: Oxford University Press.

———. 2001. *Science, Truth, and Democracy*. New York: Oxford University Press.

Klingenberg, Christian Peter. 2008. “Morphological Integration and Developmental Modularity.” *Annual Review of Ecology, Evolution, and Systematics* 39: 115–1132.

———. 2010. “Evolution and Development of Shape: Integrating Quantitative Approaches.” *Nature Reviews Genetics* 11: 623–35. doi:10.1038/nrg2829.

Klingenberg, Christian Peter, Larry J. Leamy, and James M Cheverud. 2004. “Integration and Modularity of Quantitative Trait Locus Effects on Geometric Shape in the Mouse Mandible.” *Genetics* 166: 1909–21.

Krakauer, David C, James P Collins, Douglas Erwin, Jessica C Flack, Walter Fontana, Manfred D. Laubichler, Sonja J Prohaska, Geoffrey B West, and Peter F Stadler. 2011. “The Challenges and Scope of Theoretical Biology.” *Journal of Theoretical Biology* 276: 269–76. doi:10.1016/jjtbi.2011.01.051.

Krippendorff, Klaus. 2013. *Content Analysis: An Introduction to Its Methodology*. 3rd ed. Los Angeles: Sage.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. 1st ed. Chicago: University of Chicago Press.

———. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Lakatos, Imre. 1970. “Falsificationism and the Methodology of Scientific Research Programs.” In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, 91–196. Cambridge: Cambridge University Press.

Laland, Kevin N., Tobias Uller, Marcus W. Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, and John Odling-Smee. 2015. "The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions." *Proceedings of the Royal Society B: Biological Sciences* 282: 20151019. doi:10.1098/rspb.2015.1019.

Laland, Kevin, Tobias Uller, Marc Feldman, Kim Sterelny, Gerd B. Muller, Armin Moczek, Eva Jablonka, and John Odling-Smee. 2014. "Does Evolutionary Theory Need a Rethink? Yes, Urgently." *Nature* 514: 161–64.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno, and Steve Woolgar. 1978. *Laboratory Life: The Construction of Scientific Facts*. Princeton, N.J: Princeton University Press.

Laubichler, Manfred D. 2010. "Evolutionary Developmental Biology Offers a Significant Challenge to the Neo-Darwinian Paradigm." In *Contemporary Debates in Philosophy of Biology*, edited by Francisco J Ayala and Robert Arp, 199–212. Malden, MA: Wiley-Blackwell.

Laubichler, Manfred Dietrich, and Jane Maienschein, eds. 2007. *From Embryology to Evo-Devo : A History of Developmental Evolution*. Cambridge, MA: MIT Press.

Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley: University of California Press.

———. 1984. *Science and Values*. Berkeley: University of California Press.

———. 1987. "Progress or Rationality? The Prospects for Normative Naturalism." *American Philosophical Quarterly* 24: 19–31.

———. 1990a. "Aim-Less Epistemology?" *Studies in History and Philosophy of Science Part A* 21: 315–322.

———. 1990b. "Normative Naturalism." *Philosophy of Science* 57: 44–59.

Leahey, Erin. 2016. "From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration." *Annual Review of Sociology* 42: 81–100. doi:10.1146/annurev-soc-081715-074219.

Lenoir, Timothy. 1997. *Instituting Science: The Cultural Production of Scientific Disciplines*. Stanford, CA: Stanford University Press.

Leonelli, Sabina, and Rachel A. Ankeny. 2015. “Repertoires: How to Transform a Project into a Research Community.” *BioScience* biv061: 701–8. doi:10.1093/biosci/biv061.

Lewontin, Richard C. 1974. *Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. New York: Routledge.

Lloyd, Elisabeth A. 2016. “What a Difference Research Questions Can Make!” *Proceedings and Addresses of the American Philosophical Association* 90: 129–53.

Longino, Helen E. 1996. “Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy.” In *Feminism, Science, and the Philosophy of Science*, edited by Lynn-Hankinson Nelson and Jack Nelson, 39–58. Dordrecht: Kluwer.

Love, Alan C. 2008. “Explaining Evolutionary Innovations and Novelties: Criteria of Explanatory Adequacy and Epistemological Prerequisites.” *Philosophy of Science* 75: 874–886.

———. 2013. “Theory Is as Theory Does: Scientific Practice and Theory Structure in Biology.” *Biological Theory* 7: 325–37. doi:10.1007/s13752-012-0046-2.

Lynch, Michael. 2007. “The Frailty of Adaptive Hypotheses for the Origins of Organismal Complexity.” *PNAS* 104: 8597–8604.

Matthewson, John, and Brett Calcott. 2011. “Mechanistic Models of Population-Level Phenomena.” *Biology & Philosophy* 26: 737–56. doi:10.1007/s10539-011-9277-z.

Maull, Nancy L. 1977. “Unifying Science without Reduction.” *Studies in History and Philosophy of Science Part A* 8: 143–62. doi:10.1016/0039-3681(77)90012-7.

Meehl, Paul E. 1992. “Cliometric Metatheory: The Actuarial Approach to Empirical, History-Based Philosophy of Science.” *Psychological Reports* 71: 339–467.

Mitchell, Sandra D. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Mitchell, Sandra D., and Michael R. Dietrich. 2006. “Integration without Unification: An Argument for Pluralism in the Biological Sciences.” *The American Naturalist* 168: S73–79. doi:10.1086/509050.

Moczek, Armin P., Karen E. Sears, Angelika Stollewerk, Patricia J. Wittkopp, Pamela Diggle, Ian Dworkin, Cristina Ledon-Rettig, et al. 2015. “The Significance and

Scope of Evolutionary Developmental Biology: A Vision for the 21st Century: The Significance and Scope of Evolutionary Developmental Biology." *Evolution & Development* 17: 198–219. doi:10.1111/ede.12125.

Morange, Michel. 2011. "Evolutionary Developmental Biology: Its Roots and Characteristics." *Developmental Biology* 357: 13–16. doi:10.1016/j.ydbio.2011.03.013.

Morgan, Mary S. 2012. "Case Studies: One Observation or Many? Justification or Discovery?" *Philosophy of Science* 79: 667–77.

Morgan, Mary S. 2014. "Resituating Knowledge: Generic Strategies and Case Studies." *Philosophy of Science* 81: 1012–24. doi:10.1086/677888.

Müller, Gerd B. 2007. "Evo–devo: Extending the Evolutionary Synthesis." *Nature Reviews Genetics* 8: 943–49. doi:10.1038/nrg2219.

Müller, Gerd B, and Massimo Pigliucci. 2010. "Extended Synthesis: Theory Expansion or Alternative?" *Biological Theory* 5: 275–76.

Myrvold, Wayne C. 2003. "A Bayesian Account of the Virtue of Unification." *Philosophy of Science* 70: 399–423. doi:10.1086/375475.

Nagel, Ernest. 1961. *The Structure of Science*. New York: Harcourt, Brace, and World.

Oliver, Thomas A., David A. Garfield, Mollie K. Manier, Ralph Haygood, Gregory A. Wray, and Stephen R. Palumbi. 2010. "Whole-Genome Positive Selection and Habitat-Driven Evolution in a Shallow and a Deep-Sea Urchin." *Genome Biology and Evolution* 2: 800–814. doi:10.1093/gbe/evq063.

O'Malley, Maureen A., Kevin C. Elliott, Chris Haufe, and Richard M. Burian. 2009. "Philosophies of Funding." *Cell* 138: 611–15. doi:10.1016/j.cell.2009.08.008.

Pavlicev, Mihaela, and Günter P Wagner. 2012. "A Model of Developmental Evolution: Selection , Pleiotropy and Compensation." *Trends in Ecology & Evolution* 27: 316–22. doi:10.1016/j.tree.2012.01.016.

Pearl, Judea. 2009. *Causality*. 2nd ed. New York: Cambridge University Press.

Peirce, Charles S. 1908. "Abduction and Induction." In *Philosophical Writings of Peirce*, edited by Justus Buchler, 1950th ed., 151–56. New York: Dover.

Peirson, B.R. Erick, A.A. Baker, R. Subramanian, Julia Damerow, and Manfred D. Laubichler. 2015. *Tethne: Bibliographic Network Analysis for Historians* (version v0.6.3.3-beta). doi: 10.5281/zenodo.14813.

Pigliucci, Massimo. 2008. “The Proper Role of Population Genetics in Modern Evolutionary Theory.” *Biological Theory* 3: 316–24. doi:10.1162/biot.2008.3.4.316.

Pitt, Joseph C. 2001. “The Dilemma of Case Studies: Toward a Heraclitian Philosophy of Science.” *Perspectives on Science* 9: 373–82. doi:10.1162/106361401760375785.

Plomin, Robert, and Oliver S.P. Davis. 2009. “The Future of Genetics in Psychology and Psychiatry: Microarrays, Genome-Wide Association, and Non-Coding RNA.” *Journal of Child Psychology and Psychiatry* 50 (1–2): 63–71. doi:10.1111/j.1469-7610.2008.01978.x.

Plomin, Robert, Claire M. A. Haworth, and Oliver S. P. Davis. 2009. “Common Disorders Are Quantitative Traits.” *Nature Reviews Genetics* 10: 872–78. doi:10.1038/nrg2670.

Plomin, Robert, and Yulia Kovas. 2005. “Generalist Genes and Learning Disabilities.” *Psychological Bulletin* 131 (4): 592–617. doi:10.1037/0033-2909.131.4.592.

Popper, Karl. 1935. *The Logic of Scientific Discovery*. 1959th ed. New York: Routledge.

———. 1957. “The Aim of Science.” *Ratio* 1: 24–35.

Porter, Adam H. 1989. “Gene Flow Statistics and Species-Level Taxonomic Decisions in Butterflies.” PhD Dissertation, Davis, California: University of California, Davis.

———. 2015. “Personal Communication.”

Porter, Adam H, and Norman A Johnson. 2002. “Speciation despite Gene Flow When Developmental Pathways Evolve.” *Evolution* 56: 2103–11.

Porter, Michael E. 1996. “What Is Strategy?” *Harvard Business Review* Nov.-Dec.: 61–78.

Potochnik, Angela. 2015. “The Diverse Aims of Science.” *Studies in History and Philosophy of Science Part A* 53: 71–80. doi:10.1016/j.shpsa.2015.05.008.

Quine, W.V.O., and Joseph S. Ullman. 1970. *The Web of Belief*. New York: Random House.

Rescher, Nicholas. 1998. *Predicting the Future: An Introduction to the Theory of Forecasting*. Albany: State University of New York Press.

Rheinberger, Hans-Jorg. 1997. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Romano, Laura A., and Gregory A. Wray. 2003. "Conservation of Endo16 Expression in Sea Urchins despite Evolutionary Divergence in Both Cis and Trans-Acting Components of Transcriptional Regulation." *Development* 130: 4187–99.

Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992: 13–22.

Roughgarden, Jonathan. 1996. *Theory of Population Genetics and Evolutionary Ecology*. Upper Saddle River, NJ: Prentice Hall.

Runcie, Daniel E., David A. Garfield, Courtney C. Babbitt, Jennifer A. Wygoda, Sayan Mukherjee, and Gregory A. Wray. 2012. "Genetics of Gene Expression Responses to Temperature Stress in a Sea Urchin Gene Network." *Molecular Ecology* 21: 4547–62. doi:10.1111/j.1365-294X.2012.05717.x.

Runcie, Daniel Erskine. 2012. "Genetic and Environmental Constraints on Developmental Systems: Towards Predicting Genetic Responses to Climate Change in Sea Urchins." PhD Dissertation, Durham, North Carolina: Duke University.

Ruzzene, Attilia. 2012. "Drawing Lessons from Case Studies by Enhancing Comparability." *Philosophy of the Social Sciences* 42: 99–120. doi:10.1177/0048393111426683.

Salmon, Wesley C. 1981. "Rational Prediction." *The British Journal for the Philosophy of Science* 32: 115–25.

———. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

———. 1989. *Four Decades of Scientific Explanation*. Pittsburgh: University of Pittsburgh Press.

Scott, Mike. 2012. *WordSmith Tools* (version 6). Stroud: Lexical Analysis Software. http://lexically.net/wordsmith/.

Seawright, Jason, and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61: 294–308. doi:10.1177/1065912907313077.

Sellars, Wilfrid. 1953. "Inference and Meaning." *Mind* 62: 313–38.

Servedio, Maria R., Yaniv Brandvain, Sumit Dhole, Courtney L. Fitzpatrick, Emma E. Goldberg, Caitlin A. Stern, Jeremy Van Cleve, and D. Justin Yeh. 2014. "Not Just

a Theory—the Utility of Mathematical Models in Evolutionary Biology." *PLoS Biol* 12: e1002017.

Shannon, C.E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27: 379–423.

Shrader-Frechette, Kristin, and Earl D. Mccoy. 1994. "Applied Ecology and the Logic of Case Studies." *Philosophy of Science* 61: 228–49.

Sober, Elliott. 1984. *The Nature of Selection*. Chicago: University of Chicago Press.

Sommerhoff, Gerd. 1974. *The Logic of the Living Brain*. London: John Wiley and Sons.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2001. *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.

Stern, David L. 2011. *Evolution, Development, and the Predictable Genome*. Greenwood Village, CO: Roberts and Company.

Stewart, Alexander J., Sridhar Hannenhalli, and Joshua B. Plotkin. 2012. "Why Transcription Factor Binding Sites Are Ten Nucleotides Long." *Genetics* 192: 973–85. doi:10.1534/genetics.112.143370.

Stewart, Alexander J., and Joshua B. Plotkin. 2013. "The Evolution of Complex Gene Regulation by Low-Specificity Binding Sites." *Proceedings of the Royal Society of London B: Biological Sciences* 280: 20131313. doi:10.1098/rspb.2013.1313.

Stewart, Alexander J., Robert M. Seymour, Andrew Pomiankowski, and Joshua B. Plotkin. 2012. "The Population Genetics of Cooperative Gene Regulation." *BMC Evolutionary Biology* 12: 173. doi:10.1186/1471-2148-12-173.

Stewart, Alexander J., Robert M. Seymour, Andrew Pomiankowski, and Max Reuter. 2013. "Under-Dominance Constrains the Evolution of Negative Autoregulation in Diploids." *PLoS Computational Biology* 9: e1002992. doi:10.1371/journal.pcbi.1002992.

Tabery, James G. 2004. "Synthesizing Activities and Interactions in the Concept of a Mechanism." *Philosophy of Science* 71: 1–15. doi:10.1086/381409.

Tajima, Fumio. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123: 585–95.

Tenopir, Carol, and Rachel Volentine. 2012. "The Value of Scholarly Reading in the Life Sciences." *Ideas in Ecology and Evolution* 5: 63–73. doi:10.4033/iee.2012.5b.14.f.

True, John R., and Eric S. Haag. 2001. “Developmental System Drift and Flexibility in Evolutionary Trajectories.” *Evolution & Development* 3: 109–19. doi:10.1046/j.1525-142x.2001.003002109.x.

Tulchinsky, Alexander. 2013. “Evolution of Hybrid Incompatibilities in Gene Regulatory Networks.” PhD Dissertation, Amherst, Mass.: University of Massachusetts, Amherst.

Tulchinsky, Alexander Y., Norman A. Johnson, and Adam H. Porter. 2014. “Hybrid Incompatibility Despite Pleiotropic Constraint in a Sequence-Based Bioenergetic Model of Transcription Factor Binding.” *Genetics* 198: 1645–54. doi:10.1534/genetics.114.171397.

Tulchinsky, Alexander Y., Norman A. Johnson, Ward B. Watt, and Adam H. Porter. 2014. “Hybrid Incompatibility Arises in a Sequence-Based Bioenergetic Model of Transcription Factor Binding.” *Genetics* 198: 1155–66. doi:10.1534/genetics.114.168112.

Tuomela, Raimo. 1990. “What Are Goals and Joint Goals?” *Theory and Decision* 28: 1–20.

Fraassen, Bas C van. 1980. *The Scientific Image*. New York: Oxford University Press.

———. 2008. *Scientific Representation*. New York: Oxford University Press.

Wagner, Günter P. 2000. “What Is the Promise of Developmental Evolution? Part I: Why Is Developmental Biology Necessary to Explain Evolutionary Innovations?” *Journal of Experimental Zoology Part B* 288: 95–98.

———. 2010. “A New Synthesis Finally Arriving!” *Evolution* 64: 2486–88. doi:10.1111/j.1558-5646.2010.01004.x.

———. 2014. *Homology, Genes, and Evolutionary Innovation*. Princeton: Princeton University Press.

Wagner, Günter P, Chis-Hua Hiu, and Manfred D. Laubichler. 2000. “Developmental Evolution as a Mechanistic Science: The Inference from Developmental Mechanisms to Evolutionary Processes.” *American Zoologist* 40: 819–31.

Waters, C. Kenneth. 1994. “Genes Made Molecular.” *Philosophy of Science* 61: 163–85.

Weber, Marcel. 2005. *Philosophy of Experimental Biology*. New York: Cambridge University Press.

Winther, Rasmus Grønfeldt. 2015. “Evo-Devo as a Trading Zone.” In *Conceptual Change in Biology*, edited by Alan C. Love, 459–82. Boston Studies in the Philosophy and History of Science 307. Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-9412-1_21.

Winther, Rasmus Grønfeldt, Ryan Giordano, Michael D. Edge, and Rasmus Nielsen. 2015. “The Mind, the Lab, and the Field: Three Kinds of Populations in Scientific Practice.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 52: 12–21. doi:10.1016/j.shpsc.2015.01.009.

Woodward, James. 1989. “Data and Phenomena.” *Synthese* 79: 393–472.

———. 2003. *Making Things Happen*. New York: Oxford University Press.

———. 2014. “A Functional Account of Causation; Or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness.” *Philosophy of Science* 81: 691–713.

Wray, Gregory A. 2013. “Genomics and the Evolution of Phenotypic Traits.” *Annual Review of Ecology, Evolution, and Systematics* 44: 51–72. doi:10.1146/annurev-ecolsys-110512-135828.

Wray, Gregory A., Matthew W. Hahn, Ehab Abouheif, James P. Balhoff, Margaret Pizer, Matthew V. Rockman, and Laura A. Romano. 2003. “The Evolution of Transcriptional Regulation in Eukaryotes.” *Molecular Biology and Evolution* 20: 1377–1419.

Wray, Gregory A., Hopi E. Hoekstra, Douglas J. Futuyma, Richard E. Lenski, Trudy F.C. Mackay, Dolph Schluter, and Joan E. Strassmann. 2014. “Does Evolutionary Theory Need a Rethink? No, All Is Well.” *Nature* 514: 161–64.

Wray, Gregory Allan. 1987. “Heterochrony and Homology in the Evolution of Echinoid Development.” PhD Dissertation, Durham, North Carolina: Duke University.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. “The Increasing Dominance of Teams in Production of Knowledge.” *Science* 316: 1036–39. doi:10.1126/science.1136099.

Wygoda, Jennifer A., Yee Yang, Maria Byrne, and Gregory A. Wray. 2014. “Transcriptomic Analysis of the Highly Derived Radial Body Plan of a Sea Urchin.” *Genome Biology and Evolution* 6: 964–73. doi:10.1093/gbe/evu070.

Yin, Robert K. 2003. *Case Study Research*. 3rd ed. Thousand Oaks, CA: Sage.

APPENDIX A

PROTOCOLS TO FIND RESEARCH PROJECTS IN CORPERA

Goal: If you follow this protocol, you will have an electronic file that represents a population of research projects related to a topic. The protocol assumes the concept of and operationalization of "research project" in Chapter 2. The protocol relies on the Web of Science Database and on the programs Tethne (Peirson et al 2015) and Cytoscape (Kilcoyne et al. 2009; http://www.cytoscape.org/). The protocol has eleven steps.

1. **Develop a set of known cases.** Use informal methods to list a small collection of research projects (4–8) related to the theme of interest.

2. **Form a topic or a collection of papers about a common theme.**
    a. The theme is captured by the questions or problem that motivates your project.
    b. Specify a multi-year time period on which to focus the theme.
    c. Develop a list of keywords relevant to that theme.
    d. On a web browser, go to the Web of Science (WoS) database. It will require access through a paying subscriber, such as a library.
    e. In the Wos database, input your keywords and collect your results.
    f. Refine your results to include documents only from the time period specified in step (2.b).
    g. Refine your results to exclude document types, such as letters to the editor, that aren't wanted in the final collection of papers. This step may take several iterations and refinements.
    h. Export final WoS results for use in Tethne by following the protocol: http://diging.github.io/tethne/doc/0.6.1-beta/tutorial.getting_data.html
    i. For WoS data, you can only download information about 500 papers at a time. If your WoS collection of papers is greater than 500, download the data in files of 500 papers or less. Once you have all of those files, copy and past the data from each of them into a single large text file. Store all of these files in a manageable folder on your computer.

3. **For any any two papers, compare the authors, and place the papers that share an author in a common pile or folder (ORP-4).**
    a. Install Tethne on your computer. Use the following installation guide. http://diging.github.io/tethne/doc/0.6.1-beta/installation.html
    b. Load the WoS data into Tethne according to the following instructions. http://diging.github.io/tethne/doc/0.6.1-beta/tethne.readers.wos.html#module-tethne.readers.wos
    c. Analyze the WoS data such that Tethne creates author coupling networks, in which papers are nodes and edges between the papers represent shared authors. http://diging.github.io/tethne/doc/0.6.1-beta/tethne.networks.papers.html#tethne.networks.papers.author_coupling
    d. Export the networks into a GraphML file for use in Cytoscape. Follow the directions for 'Reading WoS Data', 'Building a Static Network', and 'Export to GraphML', but in the instructions, 'bibliographic_coupling()' with

'author_coupling()'. http://diging.github.io/tethne/doc/0.6.1-beta/tutorial.bibliocoupling.html#building-a-static-network

4. **From the papers in each pile or folder, infer the team members (ORP-3) and the temporal period of the project (ORP-5).**
   a. (This step is easier after the next step. Complete this step on a separate piece of or from a printout of the Cytoscape visualization.)

5. **List each putative research project, its members, and its time period.** Each pile or folder represents a putative research project. The collection of piles or folders represents a population of research projects.
   a. Install Cytoscape 3.1 on your computer. http://www.cytoscape.org/download.php
   b. Load your GraphML file into Cytoscape. http://diging.github.io/tethne/doc/0.6.1-beta/tutorial.bibliocoupling.html#import
   c. Visualize the network and play around with it using the tips described here. http://diging.github.io/tethne/doc/0.6.1-beta/tutorial.coauthors.html#visualize-the-static-graph
   d. Save the network as a .csv file that can be loaded in Cytoscape for later study.
   e. Save the network as an image file to be used in reports.

6. **Check for false negatives in the population of cases.** Use the set of known cases from step (1). Compare the population to that set, and if the population lacks many of those cases, begin again at step (1). Use a new method to collect papers for step (2).

7. **Check for false positives.** Focus on putative projects with many more papers or members, or with much longer temporal periods, than other putative projects in the population. Those putative projects may represent several research projects. To distinguish those projects from each other, use inference methods (V-ORP-6 through V-ORP-9). Furthermore, check that authors with the same name are in fact the same person.

8. **Check further for false positives**. If step six doesn't resolve large putative projects into distinct smaller projects, use V-ORP-10 on the abstracts of the papers.

9. **Check still further for false positives**. For each project, search for other papers by the same authors in roughly the same time period about the same theme. Additional papers further support the existence of that project.

10. **Check once again for false positives**. For each project, search for other kinds of documents that could be outputs of the project. Such documents include grant

applications or reports, white papers or institutional documents, outreach pieces, etc. Additional documents further support the existence of that project.

11. **Validate the results**. For each project, if possible, contact the members of the team. Ask them if, given the concept of research project, they would classify the papers listed as outputs of the same project. Assent provides further evidence for the existence of the project. Dissent provides evidence against the existence of the project, and perhaps a need to modify the set of papers, the relevant team members, or with enough dissent, the concept of research project.

Notes on the Protocol

As each network represents a putative research project, the collection of such networks represents the putative population of research projects related to a topic. The putative population is an approximation to the actual population of projects related to a topic. The above protocols bias the putative population in at least two ways. First, they reconstruct only those projects that had research papers as outputs. Second, they reconstruct only those projects that published papers in journals indexed and sorted by the academic database used, Web of Science in this case. Furthermore, much depends on the sets and structures of keywords used to generate a corpus in the databases, and those who use the protocols will have to tinker with those sets and structures before they get usable results.

Next, there is a trade-off between the specificity of a theme and the size of the population of research projects. The more specific the theme, as represented by the structure of the intersections and unions of the search results of many keywords, the smaller will be the population. The less specific the theme, the larger the population. If the population is small, we face issues of how to select projects for individual case study without biasing our results. I address this issue in Chapter 2. Also, if the population is large, many of the networks will represent clusters of many research projects, and a case study researcher will spend much time distinguishing distinct projects.

Finally, if we study in-depth a research project or projects identified by the above protocol, the results of those studies generalize to the population from which they were selected. Thus, the concept of research project and the protocols enable us to specify an answer to the question raised in the introduction. That case is a case of what, exactly? If we use the concept and the protocols, we can answer: That case x is a case of research projects RP related to topic T. Each individual answer must specify a single case x, a population of cases RP, and T.

APPENDIX B

PROTOCOL FOR CONSTRUCTING CASE DESCRIPTIONS

Goal: If you follow this protocol, you will have a case description of a research project. That description will be in a standard format so that you can systematically compare cases to each other. That description will have five primary sections:

Section 1: Research Team
Section 2: Research Project and Rationale
Section 3: Models
Section 4: Epistemic Aims
Section 5: Primary Conclusions/ Results

The protocol below has eight primary steps. Step 1 must be completed first, and steps 7 and 8 must be completed last. Steps 2 through 6, however, can be completed in any order you find most agreeable.

1. **Collect all of the research papers from which you will construct the description.**
    a. Select a research project as per the protocols in Appendix A.
    b. Collect all of the papers identified as outputs of that project.
    c. Of those papers, focus on ones that share the most common authors, phenomena studied, methods, and scientific products.
    d. Get rid of papers that erroneously identified by the computer program as outputs of the same project.
    e. For the remaining papers, list the co-authors. List the earliest publication date and the latest publication date. Create a timeframe that goes from the five years prior to the earliest date to five years after the latest date.
    f. For each author, collect every paper published in the timeframe. To find all such papers, use at least Google Scholar, Web of Science, and look for the personal or laboratory webpages of each author and their lists of publications.
        i. For Google Scholar, search by: author:"[first name or initial] [last name]"
        ii. Use the Custom Range tool on the left to select the time period.
        iii. Google Scholar: https://scholar.google.com
        iv. Use similar criteria in Web of Science: https:/webofknowledge.com
    g. Read the abstract of each paper collected. Separate those that seem to be outputs of the common project according to step c, and discard the rest.
    h. Of the remaining papers, read them and see if they cite earlier papers in the project, and see if later papers from the project cite them. Also, ensure that the papers meet the aspects listed in step c. If a paper meets many of the above criteria, include it as an output of the project. Otherwise, discard it.
    i. As you add papers, two things will evolve: the cast of team members and the timeframe of the project. repeat steps e–h as necessary as you add new team members to the project and potentially additional years. Do so until

the cast is stable, as is the time frame, and there are no more papers to review.

j. For the final cast, get the dissertations, or at least the dissertation metadata, for each team member who has completed one by the time of your study.
    i. Use the Proquest Dissertations and Theses digital database: http://www.proquest.com/products-services/pqdtglobal.html
    ii. The archive has many full text dissertations, especially those published in the US since 2000, and those documents have lots of information about academic lineage and often short autobiographies.
    iii. Even when the archive lacks full text documents, it has metadata for nearly all dissertations published in the US since the 1950s.
k. Separate those dissertations that were outputs of the research project, and add them to the collection of project outputs.
l. Of the documents in the final collection for the project, separate them into at least the following categories: research reports; review papers; dissertations.
m. For the final cast of members, count how many documents each (co-)authored, and rank the members according to that number, with the highest number at the top of the list and the lowest at the bottom.

2. Write the following subsections of Section 2: Research Project and Rationale.

a. *Name of Project*: Name the project and team. For instance, for the team in which Wray is the most prominent member, followed by Garfield and then by Runcie, name the project the WGR Project and the team the WGR team.
b. *Published Outputs of Project*: List the items from step 1.l above. Have subsections for each category of research report, review articles, and dissertations. For each subsection, list the title of the document and the authors. Within the subsections, list the documents in chronological order by publication date.
c. *Type of Population Studied*: List the kind of populations studied: wild, laboratory, wild/laboratory, or simulated. Also, list the species studied.

3. Write all of the subsections for Section 1: Research Team.

a. *Time period and location of the team and project*: Answer the following questions:
    i. When did the team form?
    ii. When did it end?
    iii. Where did it exist? Was the project one of several from a common lab or network of labs? If so, what are the names of the labs? In what city, state, and country did the lab or project exist? At what institution?

iv. How does the project differ from similar projects pursued by the team or by the lab?

b. *Team Members:* List each member of the team. List them in order according to the rank from step 1.m above. For each person, provide as much of the following information as possible.
    i. Name
    ii. Position in the lab (primary investigator, lab manager, graduate researcher, outside expert, undergraduate researcher, etc.) Look to lab websites for help on this issue.
    iii. Position in the project (designer, data analyzer, materials provider, etc.) For guidance here, look to the author contribution sections of research reports.
    iv. Official position at the institution (professor, graduate student, etc.) Look to lab websites for help here.
    v. When, where, and in what field the person received her highest degree, and what kind of degree it is. If it's a terminal degree, list the advisor.
    vi. Birth date and age when the person joined the team.
    vii. Beginning and end dates during which the person was a member of the team.
    viii. How many of the project documents the person (co-)authored.
    ix. Where the person is now and what position she has.
c. Revise any of the previously written sections as needed.

4. **Finish Section 2: Research Project and Rationale.** Focus on the subsection *Research Rational*, which should have the following subsections.
    a. Primary Phenomena of Study: Focus on the research reports, but also consult the review papers as needed.
        i. Focus on the abstract, introductions, and conclusions.
        ii. Underline text that helps you address the following questions:
            1. What process and items of individual organisms did the team study?
            2. What processes and items of populations did the team study?
        iii. Use that underlined text to write a few short paragraphs that summarizes the primary phenomena studied by the team, that phenomena about which they addressed question.
    b. Research Problems:
        i. Rely on the account of problems provided in (Elliott 2016), according to which problems are states of affairs in which something valued is harmed or obstructed from flourishing.
        ii. Read each document and underline in a distinct color each passage that indicates a problem that motivates the team to pursue the project, or against which they evaluate the significance or success

of the project. Focus especially on abstracts, introductions, and conclusions.
iii. Quote each of these passages on your description document.
iv. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus list across readers.
v. Group the final quotes by kind of publication, if necessary. If necessary or helpful, group the quotes into the kinds of problems they indicate: intellectual problems, research difficulties, other, etc. Within all groupings, list the quotes chronologically by date of publication.
vi. Summarize common themes from the quotes in a short paragraph after the categorized list.

c. Research Questions:
i. Rely on the account of questions provided in (Elliott 2016), according to which questions are abstract objects represented by interrogative statements, with which people indicate a search for information they lack.
ii. Read each document and in a distinct color underline each passage that indicates a question that motivates the team to pursue the project, or against which they evaluate the significance or success of the project. Focus especially on abstracts, introductions, and conclusions.
iii. Quote each of these passages on your description document.
iv. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus list across readers.
v. Group the final quotes by kind of publication, if necessary. If necessary or helpful, group the quotes into the kinds of problems they indicate: intellectual problems, research difficulties, other, etc. Within all groupings, list the quotes chronologically by date of publication.
vi. Summarize common themes from the quotes in a short paragraph after the categorized list.
vii. Revise any of the previously written sections as needed.

d. Methods:
i. Focus on the research reports, and especially on the methods sections therein. But also skim the abstract, introduction, and results sections for relevant information.
ii. In a distinct color, highlight each passage that indicates the methods the team used.
iii. To summarize the team's methods in a few paragraphs on your description document, refer to the underlined sections to address as many of the following questions as you can:
1. What physical models (if any) did the team use?

2. What devices?
3. What items in the object of study did the team manipulate?
4. What processes in the object of study did the team manipulate?
5. What named methods did the team use (GWAS, Western Blot, NCII, etc.)?
6. What methods did they use to control their object of study?
7. What methods did they use to collect data?
8. What methods did they use to refine or clean their data?
9. What methods did they use to analyze their data?
10. How did they tie their data to scientific products like theories, models, hypotheses, etc.?

iv. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus summaries across readers.

v. Revise any of the previously written sections as needed.

5. Write Section 3: Models.

a. Focus only on the research reports (and dissertations, if analyzed).

b. Reread the discussion of models in Chapter 4 of this dissertation.

c. Focus first on mechanistic models.

i. In a distinct color, highlight each passage in each report that indicates the mechanistic models the team used. From those highlighted passages:

ii. List all the mechanistic models used.

iii. For each model, list the mechanism's parts, activities, and interactions. Also, list the mechanisms supraphenomenon, that phenomenon that the mechanism as a whole brings about.

iv. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus summaries across readers.

v. Summarize the results on your description document.

d. Focus on statistical and mathematical models. In a distinct color, highlight each passage in each report that indicates the statistical or mathematical models the team used. From those highlighted passages:

i. Focus first on data models. List all of the data models used, and explain how they're to be interpreted.

ii. Focus next on statistical tests. List all such tests used, and explain how they're to be interpreted.

iii. Focus next on any other mathematical models, advanced functions, etc. Rewrite the models and explain how they're to be interpreted.

iv. For the above steps, you may need to learn more about the models. Look to the works cited pages of the research reports, see which books are cited there, get and read the relevant sections from those texts. Use other texts and online sources as necessary.

v. If possible, use multiple people to do the above four steps, and compare their results. Discuss different results, and make a consensus summaries across readers.
vi. Summarize the results on your description document.

e. As much as possible, organize your summaries chronologically according to the publication date of individual papers within the project.
f. Revise any of the previously written sections as needed.

6. Write Section 4: Epistemic Aims.

a. Write the following subsections.
b. *Open Coding Results:*
   i. Rely on the account of epistemic aims provided in Chapter 3 of this dissertation.
   ii. Read each document and underline in a distinct color each passage that indicates an epistemic aim that motivates the team to pursue the project, or against which they evaluate the significance or success of the project. Focus especially on abstracts, introductions, and conclusions.
   iii. Quote each of these passages on your description document.
   iv. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus list across readers.
   v. Group the final quotes by kind of publication, if necessary. Within all groupings, list the quotes chronologically by date of publication.
   vi. Summarize common themes from the quotes in a short paragraph after the categorized list.
c. *Count Data*:
   i. Use the protocol in Appendix C for doing computerized content analysis.
   ii. Provide tables for Raw Data, Cleaned/Informative Data, and Ranked Aims/Goals.
d. Revise any of the previously written sections as needed.

7. Write Section 5: Primary Conclusions of the Project.

a. Focus only on research reports (and dissertations, if analyzed). Create a subsection on your description document for each report.
b. For each report, focus especially on the results section, and the section titles therein. Also look to the ends of the abstract, conclusion, and discussion sections.
c. Using a distinct color, underline each passage that indicates a primary result.
d. From those underlined passages, provide a summary of the report as follows:
   i. One paragraph to review what the team did

ii. Bulleted list of primary results, which generally includes 3 to 8 items.

e. If possible, use multiple people to do the above three steps, and compare their results. Discuss different results, and make a consensus list across readers.
f. On the description document, list the summaries in chronological order by date of publication.
g. Revise any of the previously written sections as needed.

8. Tidy the Description Document.

a. Reread the document, and correct for typos, style, and inconsistencies.
b. Revise any of the previously written sections as needed.
c. If possible, use multiple people to do the above two steps, and to list sections that need substantive changes or revisions. Discuss different results, and make a consensus description across readers.

9. Validate the Description Document.

a. If possible, have one or more of the team members review the report and send feedback. Prepare a questionnaire for those reviewers to read that has them focus especially on the descriptions sections about models and epistemic aims.
b. Once you receive feedback, revise the document as needed. For suggested revisions you choose not to address, create a second document in which you list the comment and explain why you don't address it.

APPENDIX C

PROTOCOL TO USE CONTENT ANALYSIS TO UNCOVER EPISTEMIC GOALS

Goal: If you follow this protocol, you will have a database and tabulated results. Those results will count the number of times words associated with distinct epistemic goals are used in single texts or a collection of texts.

There are four primary stages, all of them to be done on a computer. The second stage is specific to a single program, Wordsmith Tools. The second stage may be completely replaced by instructions specific to another program. The third stage is specific to data output by that program, but will still function for many other programs.

Stage 1: Prepare texts for content analyses
Stage 2: Prepare texts for, and use, Wordsmith Tools program to get raw data
Stage 3: Inspect and clean data
Stage 4: Tabulate results

Stage 1: Prepare texts for content analysis
Goal: At the end of this stage, you will have a collection of documents in .txt format and organized in a set of folders. They will be in .txt format so that you can load them in any number of content analysis programs. They will be organized in a nested set of folders so that you can retrieve them easily

1. Create a folder and name it something like: Case1
    a. Open the folder and create three sub folders.
    b. Name them: 1Documents; 2Codebooks; 3Data

2. The first folder, 1Document, will house all the documents you want to code
    a. Open 1Documents
    b. Create a subfolder for all the files in their original file formats.
        i. Name it something like: OriginalFormatDocuments
        ii. Put all the original documents into that folder.
        iii. If necessary or helpful, rename the documents so that they share a common name format. For example, year of publication, and article title: (2015)AnInquiryintoMacroevolution.pdf
        iv. For journal articles that have substantial supplemental text documents, save those files here also.
    c. Create a second subfolder for versions of all those documents in plain text format, such that the file names end with the extension: .txt You can edit those files in programs like:
        i. TextWrangler http://www.barebones.com/products/TextWrangler/
        ii. NotePad++ https://notepad-plus-plus.org/
    d. Name the folder something like: RawTexts
    e. For each of the documents in the first folder, make a text version. Use the following methods, ranked in order of ease and quality of text.
        i. Look for the document in html version on the internet. for example, if the document is a recent journal article, go to the journal's website and look for a version of the article you can read

in your internet browser without opening a PDF reader. Copy the text, and paste it into a new .txt file.

ii. For documents in PDF format, use some program that does optical character recognition (OCR).
    1. Adobe Acrobat Pro https://acrobat.adobe.com/us/en/acrobat/how-to/pdf-to-word-doc-converter.html
    2. ABBYY FineReader https://www.abbyy.com/en-us/finereader/
    3. PDF to Text http://pdftotext.com/
    4. Paper Machines plugin on Zotero http://papermachines.org/

iii. For each new .txt file, save it and name it with the same filename as it's associated source document.
    1. For instance, if you created a text file from (2015)AnInquiryintoMacroevolution.pdf
    2. then rename the text file: (2015)AnInquiryintoMacroevolutionRaw.txt
    3. The 'Raw' indicates that file is the direct output of an OCR program.

iv. It may seem pointless to save the raw outputs, but doing so can save you much time if you need to re-do any of the later stages.

f. Create a third subfolder for cleaned versions of the .txt files.
    i. Name it something like: CleanedTextstoCode
    ii. Copy all the documents from the RawTexts folder and paste them into this folder.
    iii. For each new .txt file, comb through each line of text, compare it to the original PDF file, and check for errors. Fix every error in the text file as needed.
    iv. For each new .txt file, ensure that it has all and only the content required by your content analysis scheme for unitizing texts, inference methods, and operational definitions for theoretical concepts/constructs. For instance, if your strategy requires you to code article titles, ensure that the title of the article is somewhere in the .txt file.
        1. For a journal article that has a substantial supplemental text document, copy, paste, and clean that text into the overall .txt file for the document, but at the end of the document.
    v. Consult your content analysis scheme for unitizing texts. For every item that is a different unit, start it on a new line, and separate it from the end of the previous unit by at least two lines.
        1. For this project, units are paragraphs, and the following things each count as distinct paragraphs:

           journal name with article citation data
           title with researcher names

section headings
abstract
normal paragraphs
figure captions
numbered indents (math equations)
whole sidebars

vi. Slightly rename each file to indicate that its contents have been cleaned to indicate.
    1. For instance, if the file name for the raw file was (2015)AnInquiryintoMacroevolutionRAW.txt
    2. Then rename the new file (2015)AnInquiryintoMacroevolutionCLEAN.txt
vii. Clean and rename each file in this folder.

3. Open the folder named 2Codebooks
    a. In it, paste the codebooks with which you will analyze the documents in 1Documents.
    b. For this project, there are 9 codebooks, each collected in Appendix D. Each is a collection of words associated with a concept. For each codebook, make a .txt file with only the words associated with the concept, but not the concept itself. They files should be named:
        1Goal.txt
        2Know.txt
        3Amalgam.txt
        4Control.txt
        5Describe.txt
        6Discover.txt
        7Explain.txt
        8Cause.txt
        9Prediction.txt

4. Save all the files and the folder structure, and back up the folder on a cloud-based tool like Dropbox.

Stage 2: Prepare texts for, and use, Wordsmith Tools program to get raw data
Goal: At the end of this stage, you will have a collection of spreadsheets that have raw data, and a collection of spreadsheets that have data specific to each document analyzed. On each sheet, each row of data will indicate a word from one of the codebooks, the sentence or surrounding text in which it's situated, the document from which the word and sentence came from, the sentence's number, and the paragraph number, and the date of analysis. This stage may be replaced by similar instructions for other programs, as long as those programs return data with the above information.

1. Prepare texts for Wordsmith Tools.
    a. Open the first document in the CleanedTextstoCode folder.
    b. You need to indicate the units in the text, which are paragraphs as detailed in the unitizing scheme.
    c. For each unit, find the beginning of the unit, and just before it type:
    d. For each unit, find the end of the unit, and just after it, type:
    e. So label each unit in the text, and save the text with the current filename.
    f. Note: You could also label the sections and subsections for analysis in Wordsmith Tools, but we don't for this project.
    g. Complete the above steps for each document in the folder.

2. Use a computer that already has Wordsmith Tools installed, or install it on your PC. It won't run on a Mac. http://www.lexically.net/wordsmith/

3. Open Wordsmith Tools. Use mostly the default setting.

4. Prepare the program to recognize paragraphs in the texts.
    a. In the left-side menu, click on the button Language Settings.
    b. In the text format boxes associated to paragraphs, code as follows
        i. To mark the start of paragraphs:
        ii. To mark the end of paragraphs:
    c. One could repeat these steps for sections, though such isn't done for this project.

5. Prepare the program to recognize sentences and specific words.
    a. In the left side menu, click on the Index button
    b. For clusters, change each to 25.
    c. Select the check box for 'Stop at sentence break'
    d. Ensure that a word is to show in the results if it's frequency is at least 1.

6. Load the codebooks.
    a. Click on the Concord button.
    b. Click on the Search Word tab.
    c. Select the upload function under the words 'Ge Search Words from File'.
    d. Select a single codebook.
        i. Navigate to the 2Codebooks folder.
        ii. Select the topmost codebook that has yet to be used.
    e. Click the Load button.
    f. Check that the correct words populate the text box on the right side of the screen.

7. Load texts to be analyzed.
    a. If still in the same Concord session, click on the Texts button.
    b. If starting a new session, click on the Concord button and load the appropriate codebook as in the previous step. Click on the Texts button

c. Click the Change Selection button. A window opens.
d. In the left side of the window, navigate to the folder with texts to code, here named CleanedTextstoCode.
e. In the left side, select all the texts in the folder, and click and drag them to the right side of the window. Remove any other files from the right side of the window.
f. Click the OK button in the top right corner.

8. Save the data.
    a. A new window opens with the data.
    b. Ensure that the sentences and paragraphs are being numbered correctly. If not, look for problems in the paragraph codes in the texts, or in the Wordsmith Tools settings from steps 4 and 5 above.
    c. If the sentence and paragraph numbers look ok, save the file as an excel spreadsheet.
        i. In the box that lets you select the number of characters to save, select some number over 150.
        ii. Name the excel file after the codebook used to analyze the texts.
        iii. Open the file. Sort the rows in the following order
            1. From lowest to highest by sentence number.
            2. From lowest to highest by paragraph number.
            3. From A to Z by article name.

9. Repeat steps 6 through 8 for each codebook.

10. Store the raw data.
    a. Create a folder and name it RawData.
    b. Put all the spreadsheets into that folder.
    c. Put it into the 3Data folder.
    d. In the 3Data folder, create a second folder named RawArticleData

11. Create spreadsheets specific not to codebooks, but to the texts/articles be studied.
    a. Open a new Excel workbook.
    b. Name it for the first article in the folder CleanedTextstoCode.
        i. For instance: (2015)AnInquiryintoMacroevolution.xlsx
    c. Open the first Excel spreadsheet in the RawData folder.
        i. For instance: 1Goal.xlsx
    d. Select all the rows that have data for the relevant article from step b, along with the row that has the column names
    e. Copy those rows, and paste them into the new spreadsheet.
        i. i.e. into (2015)AnInquiryintoMacroevolution.xlsx
    f. In a separate row above the column names, indicate the source of the data to follow.
        i. For instance: Goal
    g. Repeat steps c through e for each spreadsheet in the RawData folder.

h. Save the file and put it in the RawArticleData folder.

12. Repeat step 11 for each article in the CleanedTextstoCode folder.

13. In the 3Data folder, create a new folder and name it CleanedArticleData.
    a. Copy all the files in the RawArticleData folder, and past them into the CleanedArticleData folder.

14. Get rid of unnecessary data.
    a. Open the first file in the CleanedArticleData folder.
    b. Delete all columns, EXCEPT for the one named:
        N
        Concordance
        Set
        Sentence #
        Paragraph #
        File
        Date

    c. To the right of the remaining columns, make new columns and name them:
        Reason to Exclude
        Total Cells
        Noisy Cells
        Clean Cells

    d. Ensure that all the column names are repeated above each of the chunks of data specific to a common codebook.
    e. Save the file, renaming it slightly.
        i. For instance: (2015)AnInquiryintoMacroevolutionCLEAN.xlsx

15. Repeat step 14 for all files in the folder CleanedArticleData.

Stage 3: Inspect and clean data
Goal: At the end of this stage, you will have a collection of spreadsheets for which each of the rows have data have been vetted. The vetting techniques are common to qualitative content analysis studies.

1. Navigate to the CleanedArticleData folder and select a data file.
    a. For instance: (2015)AnInquiryintoMacroevolutionCLEAN.xlsx
    b. Copy the file and save it under a slightly different name.
        i. For instance: (2015)AnInquiryintoMacroevolutionCLEAN1.xlsx

2. Open the new file and check each datum.

a. Starting with the first row, read the relevant Concept ('Know', 'Goal', 'Explain', etc.) and the word picked out for that concept, marked under the Set column.
b. Read the surrounding sentence(s) for the word, marked under the Concordance column.
c. Judge if the datum is irrelevant or infelicitous to the Concept.
    i. If so, code the datum as exhibiting at least one of the following issues, as appropriate.
        1. Infelicitous homonym
        2. Infelicitous cognate
        3. Metadiscourse
        4. Describes phenomena, not project rationale
        5. Describes layout of graphics
        6. Describes research of other teams
    ii. List the reason under the column Reason for to Exclude.
    iii. Move on to the next row

d. If the datum is relevant to the concept, move on to the next row.
e. Repeat for each datum and for each Concept.

3. Repeat the above two steps for every data file.

4. Take a break of at least a week from the files.

5. Repeat steps 1–4 for all data files.
    a. Name each file slightly differently. For instance: (2015)AnInquiryintoMacroevolutionCLEAN2.xlsx

6. Make consensus files.
    a. For a given article, open the two cleaned data files.
        i. For instance:
        (2015)AnInquiryintoMacroevolutionCLEAN1.xlsx
        and
        (2015)AnInquiryintoMacroevolutionCLEAN2.xlsx
    b. For the topmost datum/ row, compare the codes for irrelevancy across the two data files.
    c. If the rows agree on the assignment of code, or on the lack thereof, move to the next row.
    d. If the data files disagree about the (ir)relevancy of the topmost row for the Concept, settle on a single code.
    e. Copy the uncoded data file and save it as a new file.
        i. For instance:
        (2015)AnInquiryintoMacroevolutionCLEAN3Consensus.xlsx
    f. On the new file, enter a consensus code into the Reason to Exclude column.

g. Repeat the steps a–f for all rows and all concepts for a given article. Save the file regularly
h. Repeat the steps a–g for all articles.

7. If possible, repeat steps 1–6 with other researchers as coders.
    a. Make consensus files across coders according as in Step 6.

8. Tabulate the results.
    a. For a given article, open the consensus data file.
    b. For the first Concept, count the rows of all data. List that number under the Total Cells column.
        i. Use Excel function =CountA(range), for which the range is the columnar list of words collected under the Set column for the concept.
    c. For the first Concept, count the rows of irrelevant or infelicitous data. List that number under the Noisy Cells column.
        i. Use Excel function =CountA(range), for which the range is the columnar list of words collected under the Reason to Exclude column for the Concept.
    d. For the first Concept, count the rows of relevant data. List that number under the Clean Cells column.
        i. Use Excel function =Cell1 - Cell2, for which Cell1 is the cell that lists number of Total Cells, and Cell2 is the cell that lists the number of Noisy Cells.
    e. Repeat steps a–d for all Concepts.
    f. Repeat steps a–e for all articles.

Stage 4: Tabulate results
Goal: At the end of this stage, you will have a collection of tables that summarize the results from the previous stages.

1. Copy Table A.1 and paste it into your document.

2. Under the Paper column, rename ‘Article 1’ to be about the first article in the case.

3. For that first article, enter the counts from the corresponding consensus cleaned data files. You can create a table for:
    a. Uncleaned or raw counts.
    b. Cleaned counts.

TABLE A.1

DATA TABULATION

Paper	Goal	Know	Amalgam	Control	Describe	Discover	Explain	Predict	Cause
Research Reports									
Article 1									
Article 2									
…									
Review Papers									
Article a									
Article b									
…									

4. Repeat step 3 for all articles in the case.

5. Copy Table A.2 and paste it into your document.

6. For the first article, and based on the results tabulated above, list the goal with the highest counts under the Goal 1 column.
    a. Consider only: Amalgam, Control, Describe, Discover, Explain, Predict

7. Repeat step 6 for the counts second, third, and fourth highest ranked goals, and list them under the respective columns of Goal 2, Goal 3, and Goal 4.

8. Repeat steps 6–7 for all articles in the case.

TABLE A.2

RELATIVE RANKS OF EPISTEMIC GOALS

Research Reports	Goal 1	Goal 2	Goal 3	Goal 4
Article 1				
Article 2				
…				
Other Papers				
Article a				
Article b				
…				

APPENDIX D

CODEBOOKS FOR CONTENT ANALYSIS

1GOAL	2KNOW	3AMALGAM
accomplish*	apprehen*	amalgam*
achieve*	ascertain*	assembl*
aim*	cogni*	assimilat*
ambiti*	comprehen*	blend*
aspir*	conclu*	bridg*
attain*	deciph*	collect*
attempt*	determin*	combin*
calling*	discern*	connect*
covet*	educat*	consolidat*
design*	elucidat*	coordinat*
desir*	enlight*	dovetail*
end*	fathom*	federat*
focus*	grasp*	fit* together
goal*	illuminat*	fuse*
hope*	infer*	fusing
inspir*	inform*	fusion*
intend*	insight*	harmon*
intent*	instruct*	integrat*
job*	intellig*	interfac*
mean* to	judg*	intermix*
mission*	know*	merg*
motiv*	learn*	mesh*
need*	teach*	put* together
objectiv*	understand*	reconcil*
plan*		synthesi*
point*		systemati*
purpose*		tie*
pursu*		tying
realiz*		unif*
requir*		union*
role*		unite*
seek*		uniti*
striv*		unity
tactic*		
target*		
utilit*		
want*		
wish*		

4CONTROL

administ*
apply*
apllie*
coerc*
command*
compel*
conquer*
contain*
control*
dictat*
domesticat*
dominat*
driv*
drov*
employ*
engineer*
exploit*
harness*
impel*
impos*
manag*
manipulat*
master*
oblig*
order*
overpower*
ply*
proscri*
reign*
rein*
restrain*
subdue*
subjugat*
superinten*
tame*
wield*
wrangl*
wrest*
yoke*

5DESCRIBE

captur*
categoriz*
characteri*
chronicle*
circumscri*
classif*
connot*
correlat*
delineat*
denot*
depict*
describ*
descrimin*
description*
designat*
detail*
display*
document*
exhibit*
identif*
illustrat*
impart*
individuat*
label*
limn*
mirror*
name*
naming
phenomenol*
pick* out
pictur*
present*
record*
recount*
render*
represent*
signify*
signification*
sketch*
specify*
specification*
spotlight*
trace*

6DISCOVER

ascertain*
bare*
comb*
com* across
detect*
determin*
discern*
discover*
encounter*
explor*
expos*
extrapolat*
find*
hunt*
locate*
locating
look* for
ma* known
perceiv*
probe*
pursu*
reconn*
reveal*
scour*
scout*
search*
seek*
spot*
stalk*
survey*
track*
trail*
uncover*
unearth*
unveil*

7EXPLAIN

account* for
demonstrat*
diagnos*
etiolog*
explain*
explan*
how
why

8CAUSE

activat*
adapt*
affect*
alter*
bring* about
catal*
caus*
chang*
coax*
compel*
compuls*
consequen*
disturb*
drive*
driving
drove
dynamic*
effect*
efficac*
elicit*
engender*
ensu*
eventuat*
excit*
forc*
generate*
generating
genesis
giv* rise
impact*
impel*
impress*
induc*
influen*
manipulat*
mechani*
modif*
modulat*
origin*
perturb*
potent*
power*
precipitat*
produc*
prompt*
react*
regulat*
respons*
sourc*
stimul*
trigger*
yield*

9PREDICT

adumbrat*
anticipat*
augur*
bode*
clairvoyan*
conject*
correlat*
divin*
expect*
extrapolat*
forecast*
foreknow*
foresee*
foresight*
foretell*
foretold
future
gauge*
guess*
herald*
look* for
portend*
predict*
presag*
prescien*
prognos*
prophes*
prospect*
reckon*
surmise*
suspect*

APPENDIX E

PERMISSIONS

I received the following permissions to reuse copyrighted figures in this dissertation.

For:	**Source:**
Fig. 3.1. Page 85.	Krippendorff, Klaus. 2013. "Figure 4.2: Components of Content Analysis." In *Content Analysis: An Introduction to Its Methodology*. 3rd edition. Los Angeles: Sage Publishing, 2013. Publisher: Sage College License: 3998910298681
Fig. 5.1. Page 137.	Johnson, Norman A., and Adam H. Porter. 2000. "Fig. 2." In "Rapid Speciation via Parallel, Directional Selection on Regulatory Genetic Pathways." *Journal of Theoretical Biology* 205: 572–42. Page 529. Publisher: Elsevier License: 3995480745332
Fig. 5.2. Page 140.	Tulchinsky, Alexander, et al. 2014. "Figure 1." In "Hybrid Incompatibility Arises in a Sequence‐Based Bioenergetic Model of Transcription Factor Binding." *Genetics* 198: 1155–66. Page 1157. Publisher: Genetics Society of America License: 3995490893335
Fig. 6.1. Page 188.	Garfield, David A., et al. 2012. "Figure 1." In "Population genetics of *cis*‐regulatory sequences that operate during embryonic development in the sea urchin *Strongylocentrotus purpuratus*." Evolution and Development 14:152–167. Page 153. Publisher: John Wiley and Sons License: 3995461118550