

What Can We Know About Past Cultural Transmission?

Extending Behavioral Models to Archaeological Time

Mark E. Madsen
University of Washington
June 6, 2013



Agenda

1. CT modeling in archaeology
2. Enriching our models of CT
3. Coarse graining of CT models for archaeological use

“The future is already here –
it’s just not evenly distributed.”

William Gibson



Wednesday, June 5, 13

Sci Fi author William Gibson usually gets quoted about technology, but his recent novels have actually been about cultural change in the modern world. I like this quote because it contains the core of Darwin's insight into evolution: change is born of variation, with some variants we see around us today becoming the seeds of tomorrow's commonplace, by adoption and diffusion and being selected for performance.

Cultural transmission models are about population thinking: how variation is distributed, and how it leads to changes that we see and describe at the population level.

Talk today about how we model that, what our primitive early models are like, and how we can improve them.

Cultural transmission models
describe the possible outcomes
that occur when we combine:

- *social learning processes*
- *social networks and institutions*
- *innovation and sources of error*
- *social and ecological niches*

Where “outcomes” include:

- *spatiotemporal patterns*
- *frequency histories*
- *richness and evenness patterns*

Wednesday, June 5, 13

Where “combine” means: construct a dynamical model (either stochastic or deterministic, but mainly stochastic)
We then study the behavior of that dynamical model over various combinations of its parameters (i.e., its phase space)
Looking for classes of “outcomes” -- and ways to create empirical tests and tools for applying to real data



Apprenticeship



Individual learning (trial and error)

Wednesday, June 5, 13
What do we mean by social learning?



Formal instruction

Imitation of relatives
or peers



Spread of agriculture

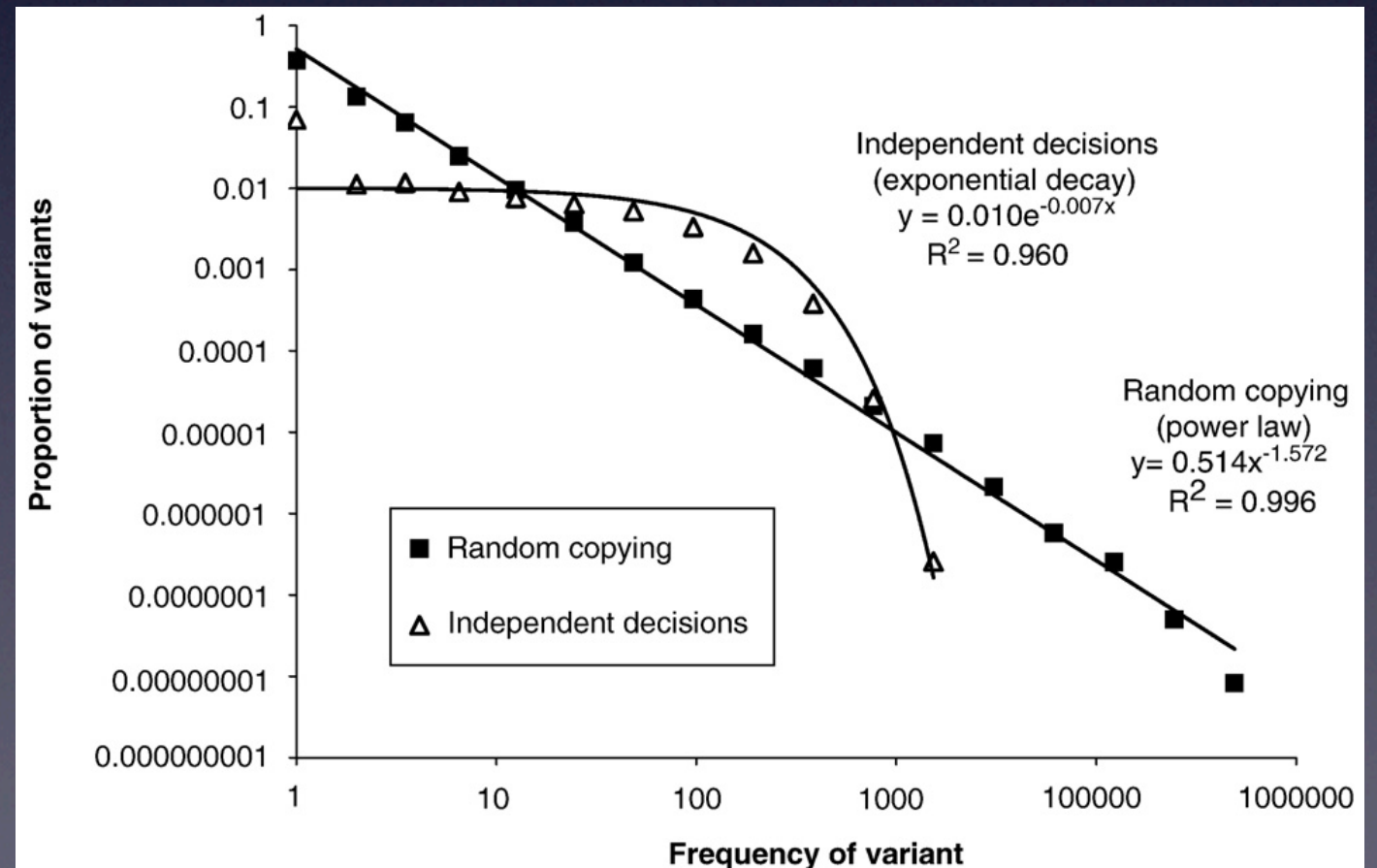
Based on the arrival of wheat
(kyr = 1000 years)



www.eliznik.org.uk/

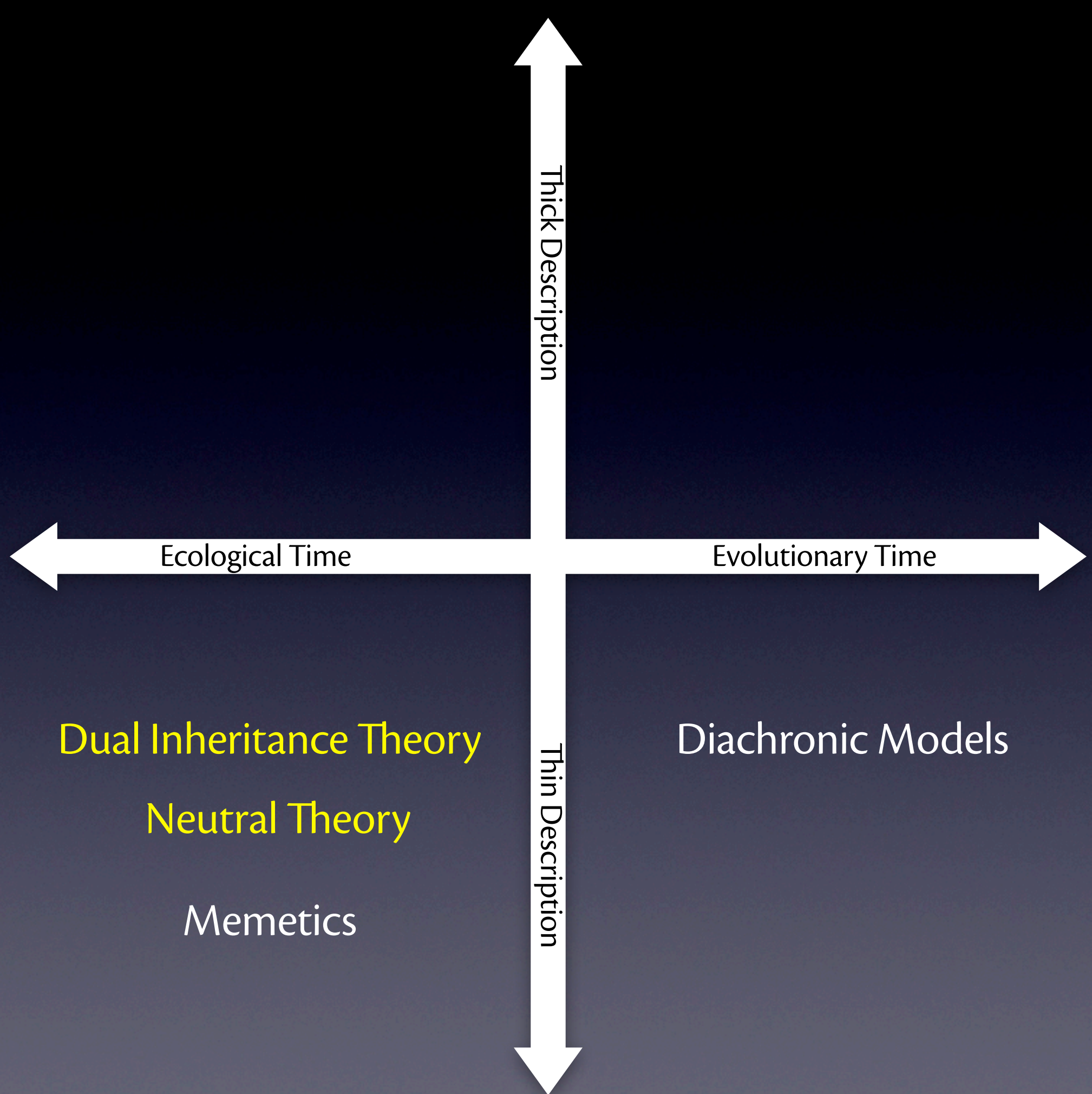
Spatiotemporal patterns: diffusion, migration, exchange

Equilibrium frequency distributions



Frameworks for creating formal CT models

- Wright-Fisher infinite-alleles neutral model
- Boyd and Richerson's dual inheritance/bias models
- Cavalli-Sforza and Feldman's dual inheritance models
- Replicator dynamics
- Epidemiological models (e.g., SIS, SIR)



Wednesday, June 5, 13

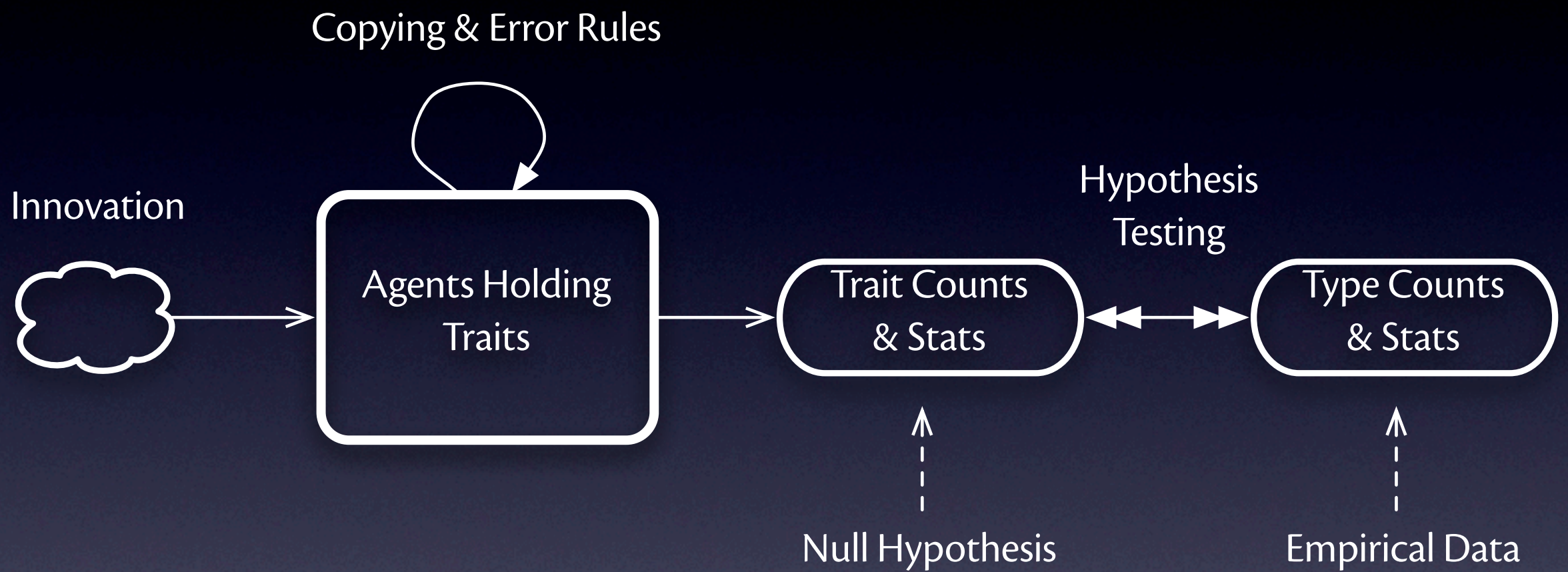
I'm going to use this MAP as an organizing device for the rest of the talk.

We're starting with formalized (i.e., mathematical or simulation) models of CT -- there's a lot more in qualitative, laboratory, or ethnographic/observational description, but my concern here is what the formalized stuff is like.

Explain why DIT and neutral theory fit on the 'ecological time' end of the axis -- equilibrium models, do not in current forms easily explain temporal history, but instead synchronic steady states.

Memetics is "thinner" than DIT because it posits little complexity on the part of humans doing the copying (i.e., imitation is like contagion), whereas DIT at least has some complexity to the biases and mechanisms posited.

Not focusing much on memetics, but a little later, I've got a slide which will allows us to see what we can (and should) essentially ignore memetics as "pop science" cultural transmission.



Schematic of CT models in the B&R or neutrality framework

Wednesday, June 5, 13

This is the schematic structure of a CT model and research project today. This schema covers most of the published applications -- Shennan, Bentley, Steele, Hamilton, Buchanan, my own early work with Lipo, Dunnell, and others.

Note that the null hypothesis here are statistical patterns -- frequency distributions we expect empirical data to match, within error limits.

Unbiased transmission rule:

For each agent, select an agent at random, adopt a random trait from the set of traits they hold. With some probability, a newly invented trait is adopted instead.

Conformist transmission rule:

For each agent, select the most common trait in the population with probability C , otherwise use unbiased copying or innovation as above.

Wednesday, June 5, 13

These are the algorithms used in the two most commonly simulated models of CT

Mention pro-novelty or anti-conformism as the opposite.

Seriously oversimplified. You can just see how oversimplified this is. Pure contact diffusion. And we don't even gain analytical tractability by simplifying this much -- WFIA

“When you first get a theory, you don’t know how it works! You’ve got to spend time playing with it....[there is] a *learning curve* for how the hypothesis works: *after* you understand it, you can falsify.”

William Wimsatt

Wednesday, June 5, 13

good reason for working with oversimplified models....even the very simple rules just shown have complex behavior, especially in structured pops. There’s a learning curve just figuring out how these rules generate different outcomes in different circumstances -- and especially in different kinds of social networks.

And there may be an explanatory role for models this simple -- especially when we have very large scale or highly smoothed data (we’ll get to that later in the talk).

But as fundamental theory designed to accurately describe what happens in social learning, and its population level consequences? We’re going to look back at these models as our "training wheels"

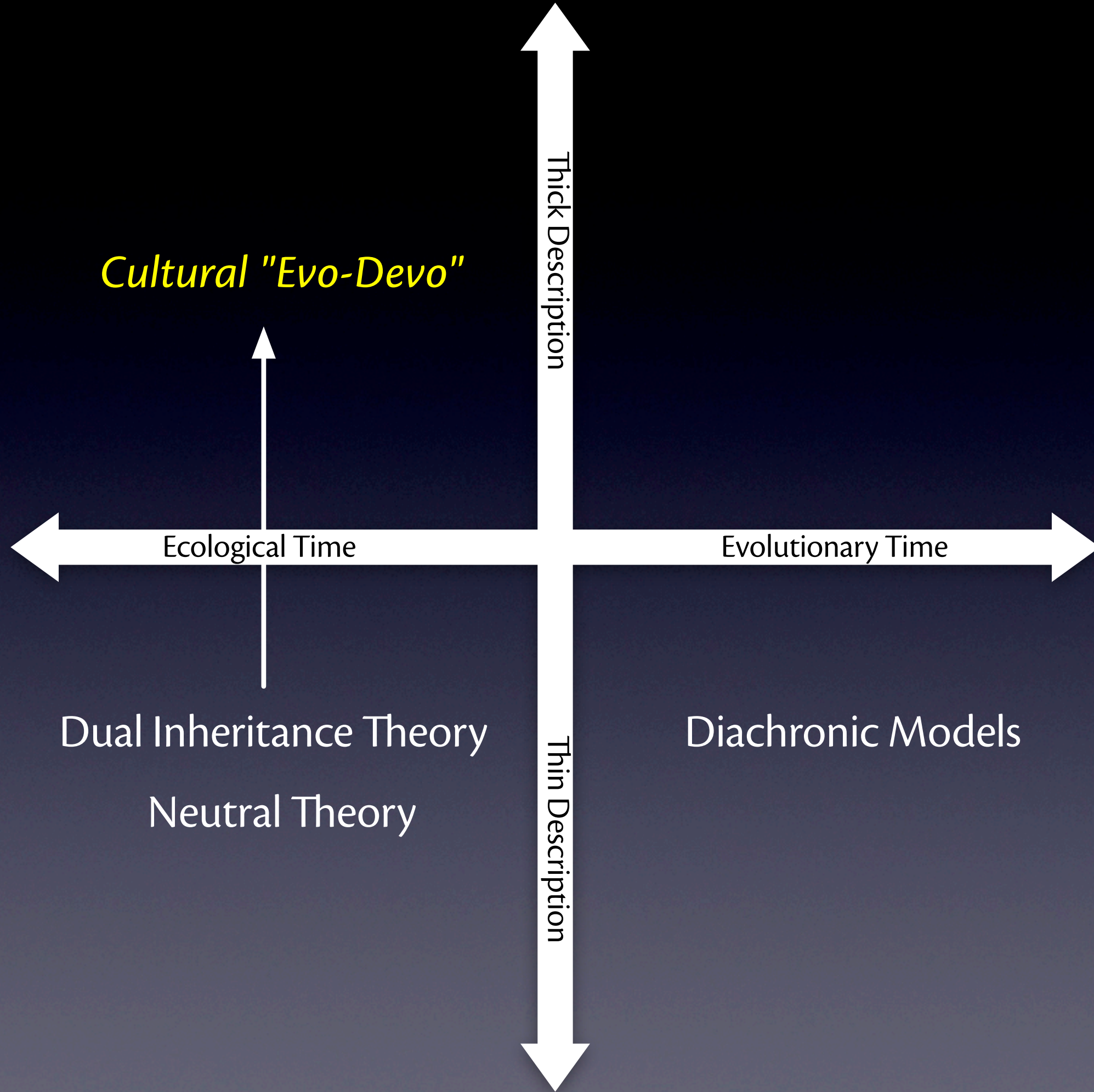
People are conformists....sometimes, and in certain contexts.

People are also novelty seekers, especially at certain ages and in specific contexts.



Wednesday, June 5, 13

The real frontier in building and applying quantitative models of CT is heterogeneity. Real people are mixtures of different learning strategies, and populations are mixtures of people with different kinds of tendencies and aptitudes. Reflecting that in a tractable way in our models is difficult, but necessary if we're goign to explain real situations and data, ESPECIALLY with contemporary populations and high resolution data.



Cultural "Evo-Devo"

Thick Description

Ecological Time

Evolutionary Time

Thin Description

Dual Inheritance Theory

Neutral Theory

Diachronic Models

What's missing is a **formal** way of describing how the **use life** of cultural information affects its spread and distribution.

In other words, the **evolutionary** effects of **development**

cultural “evo-devo”

Wednesday, June 5, 13

We know a lot about the way cultural information is used in specific situations, especially in technological areas where the constraints are clear. Much of this is from ethnography and experiment.

What we haven't done is put this into our quantitative models to see how the patterns we expect can and should change.

I'm not the one who coined this usage -- William Wimsatt and James Greisemer have done the most to influence the thinking in this section, especially in linking evo-devo to our problems of producing "thicker" descriptions of cultural transmission.

Three aspects:

1. Social learning processes themselves
2. Structured cultural information
3. Generative entrenchment and innovation (if we have time)

Most real skills and cultural phenomena involve more than just imitation or “copying”

- Imitation
- Formal instruction
- Individual learning

Often acquired over long spans of time with real expertise taking years



Wednesday, June 5, 13

This is why “memetics” is a poor account of human cultural transmission. We may imitate or copy simple things, but almost nothing else about human culture is simple enough to be accounted for by “culture as virus” or “selfish memes” style models.

Becoming a doctor....it's not done through imitation. Or least, imitation is a small part of it.

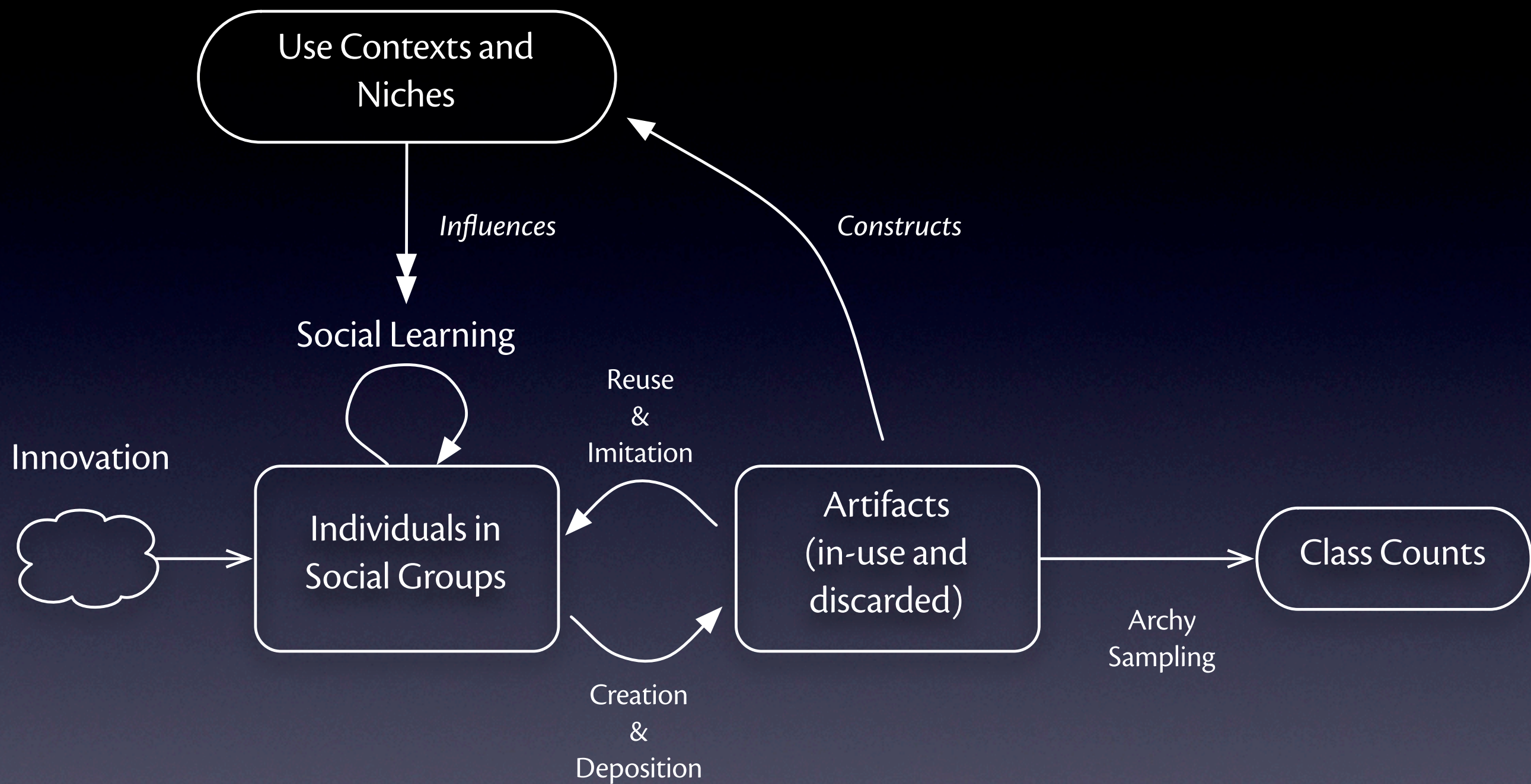
“...the upstream generation structures the learning environment of the downstream generation, so that trial-and-error learning combined with observational learning and (sometimes) explicit instruction results in the reliable reacquisition of expertise.”

Kim Sterelny, *The Evolved Apprentice*

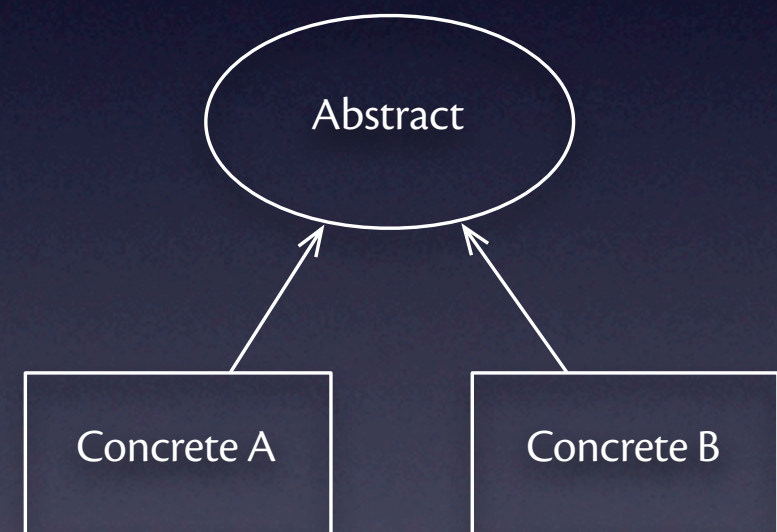
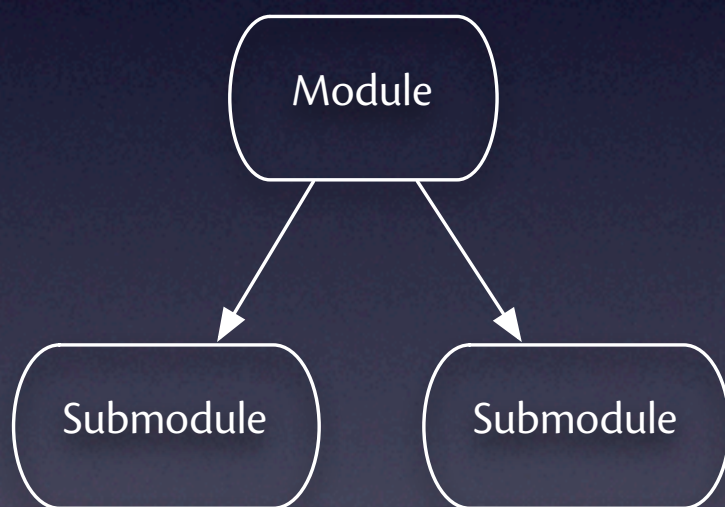
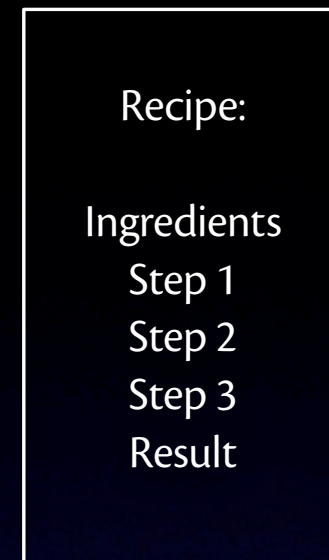
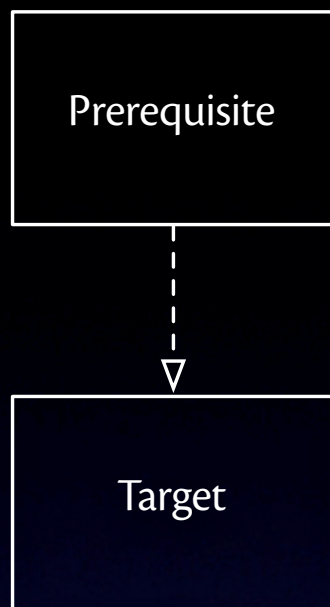
Wednesday, June 5, 13

In other words, cultural information is used by humans to construct niches, and those niches become part of the learning environment for succeeding generations, often with special purpose behaviors (like teaching, apprenticeship) and ways for neophytes in a given skill to learn step by step, safely.

So CT is itself not only part of human niche construction, but it uses constructed niches to structure learning and inheritance, often as simplified or scaffolded versions of the full niche itself.



Schematic of CT models with embodiment and niche construction



Important ways that cultural information is structured

Wednesday, June 5, 13

Give example module/submodule -- almost any modern industrial technology. Atl-atl construction. Ceramics.

Recipes show up anytime there's a series of steps. Each step, of course, can be modular itself.

Prerequisite - target shows up with any cumulative knowledge. Think arithmetic -> algebra -> calculus

Abstract / concrete turns up whenever we realize that two things share common features, and we can teach those common features once....

The important thing in CT models is that if people are learning information in "clumps" like this, these patterns WILL affect the spatiotemporal patterns we should see in data, and the frequency histories and correlations -- the "outcomes" and empirical tests I mentioned at the beginning.

And we haven't yet begun seriously modeling them yet. Mesoudi and O'Brien did some modeling to establish the importance of this kind of structure over structureless CT, but they weren't focused on WHAT YOU'D SEE.

Generative Entrenchment:

Early developmental events structure and constrain later events, becoming “prerequisites”

Genetics and development is rife with GE'd elements

Culture has many GE'd elements as well, and mechanisms for creating stability if not stasis

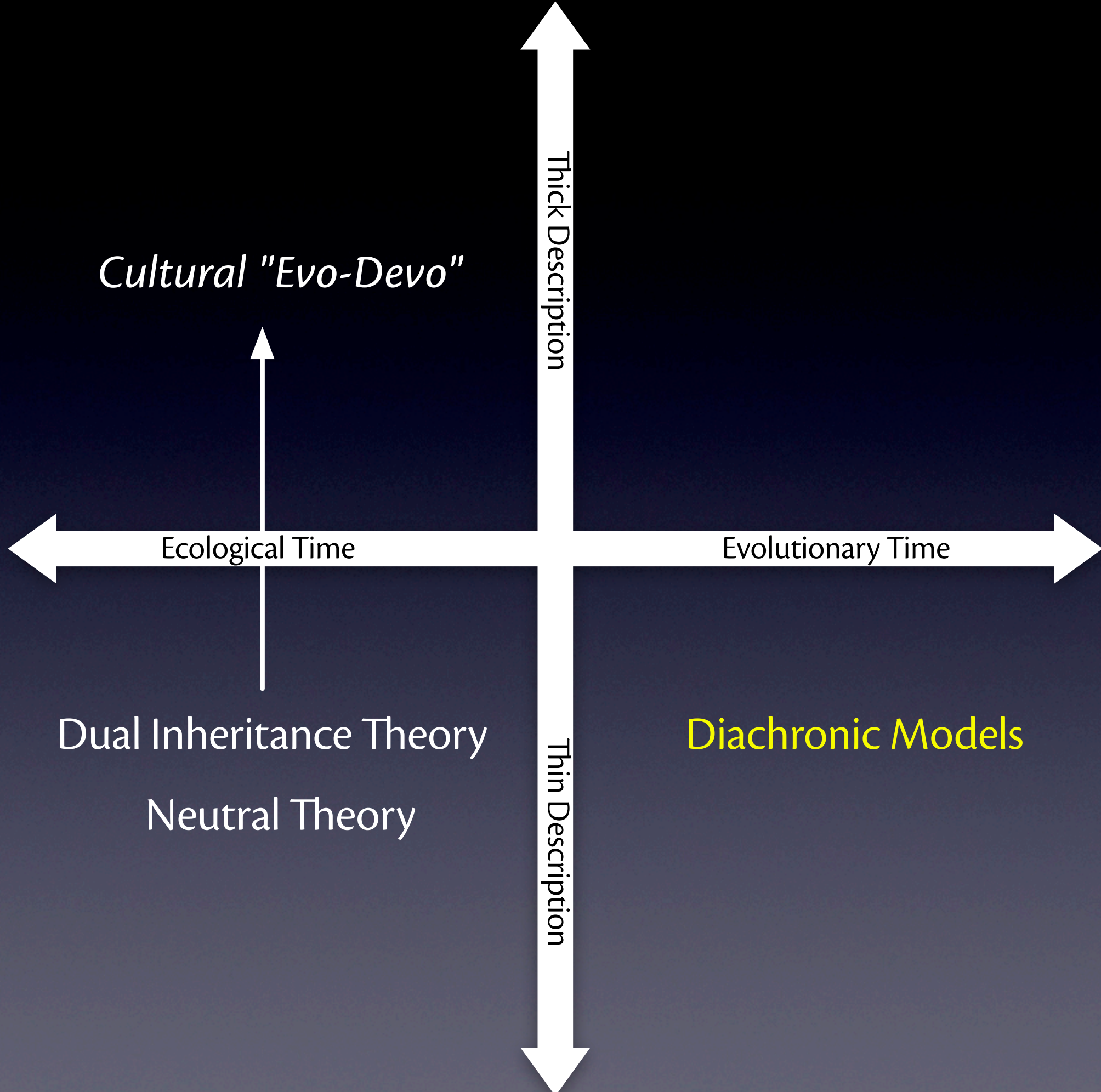
Amish “*Ordnung*” -- Craig Palmer

- Culture provides tools for “loosening” GE
- Behavior is changeable with new learning
- Prototype and test before switching
- Early/mainstream/late adopter cycle

Wednesday, June 5, 13

Adoption cycles allow:

- partial testing in the population without full commitment (i.e., allopatric change is safe change)
- creates a pool of teachers and models
- allows time to construct learning niches and scaffolds for most people to acquire the new skills and knowledge
- laggards and late adopters preserve old knowledge if you need to switch back





Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi



A non-equilibrium neutral model for analysing cultural change

Anne Kandler^{a,b,c,*}, Stephen Shennan^c

^a Department of Mathematical Science, City University London, Northampton Square, London EC1V 0HB, UK

^b Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

^c Institute of Archaeology, University College London, 31–34, Gordon Square, London WC1H 0PY, UK

JTB 330: 18-25

Synchronic \Rightarrow Diachronic version of unbiased CT

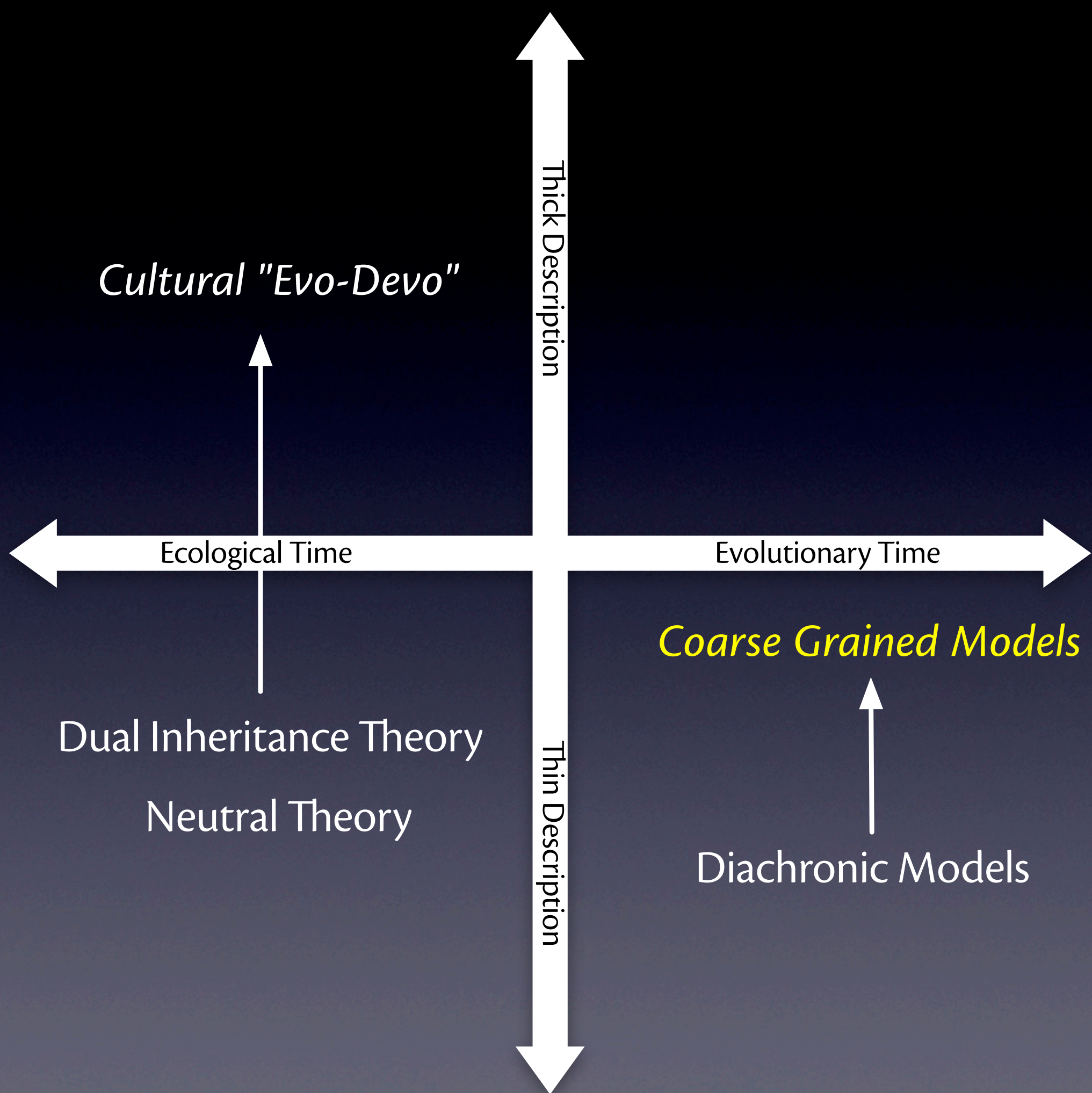
Statistic: expected number of variants at T,
till present at $T+n$

Perfect for archaeological use...?

Wednesday, June 5, 13

I'll be coming back to this example in the next section of my talk, but this is a huge step forward in the construction of mathematical CT models for use with historical or archaeological data.

Talk a bit about synchronic/equilibrium model predictions versus nonequilibrium or diachronic predictions....





Coarse graining: fine details are “averaged over” to produce a less detailed model

Types of Scientific Theories

- Fundamental theory
- Effective theory

*accurate description of phenomena
at a particular scale, without claiming
completeness or full realism*

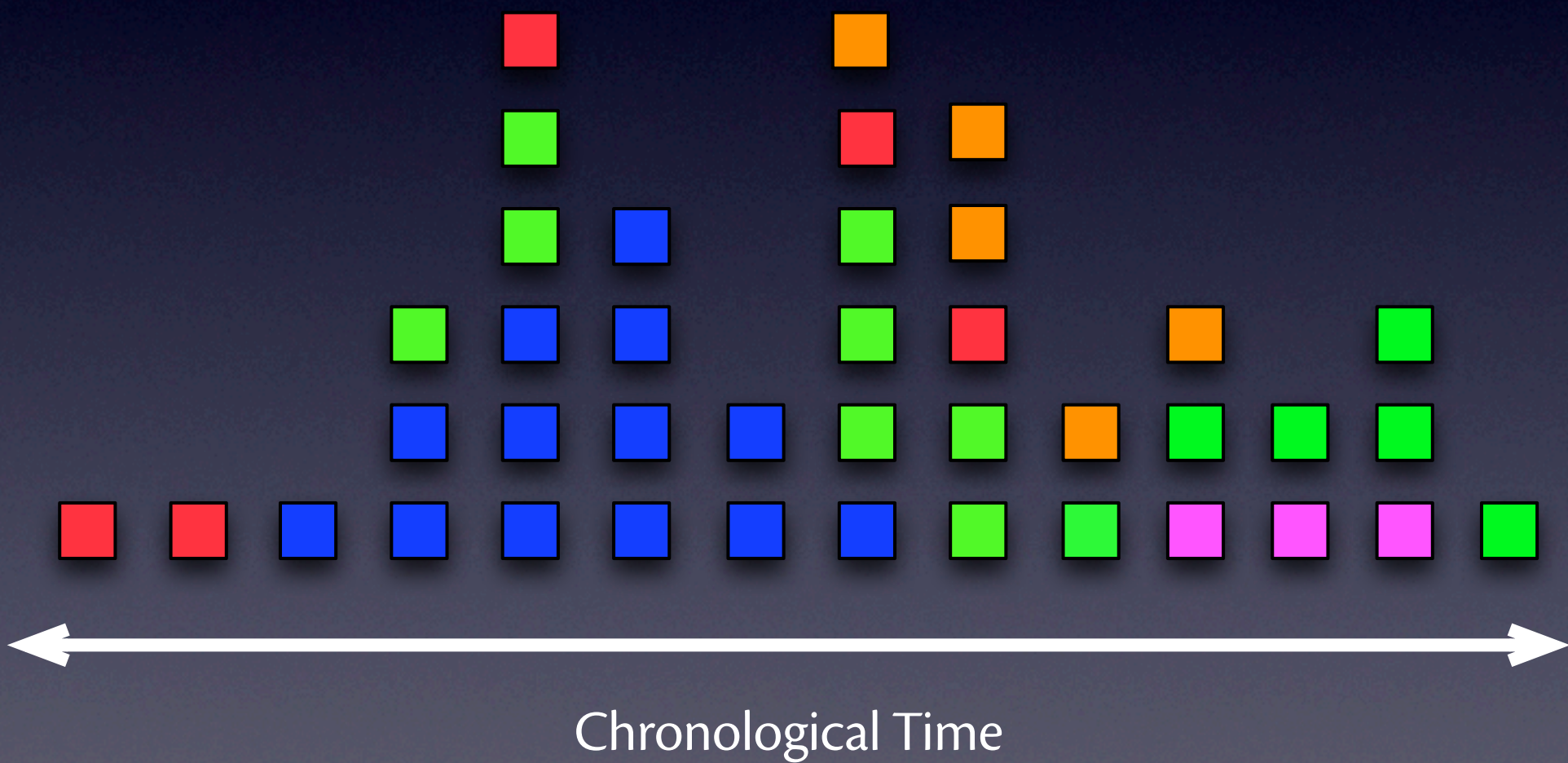
Cultural “evodevo” \Rightarrow **fundamental theory**

Coarse Grained CT models \Rightarrow **effective theory**

Temporal Coarse Graining (*Time Averaging*)

Artifact Deposition (Color = Types)

Internal Structure of an Artifact Assemblage



Wednesday, June 5, 13

When we take a bulk sample of an archaeological deposit (either by excavation or surface collection), we’re seeing many depositional and discard events aggregated together.

Intuitively we understand that many of the items in that sample might not have been contemporaneous, and that the sample represents a DURATION of time, not an instant in time.

The question is, what effect does this have on the kind of metrics we look at in cultural transmission models? Answering this is critical since virtually every testable prediction we make from any CT model -- neutral theory, Boyd and Richerson’s bias models, etc -- are counts or averages.

So comparing model predictions which are derived from “point in time” values, to archaeological data which are aggregated over DURATIONS, isn’t a valid comparison.

Instead, we need to know what the effect of aggregation is, on our models. I undertook numerical experiments to start figuring this out.

Numerical simulation experiments

- Simulated unbiased and conformist CT
- Range of innovation & conformism rates
- Each historical trajectory sampled raw, and 11 levels of temporal aggregation
- Measured richness, neutrality tests, evenness, estimation of “theta”

Full results: <http://arxiv.org/abs/1204.2043>

Wednesday, June 5, 13

Open source software: Transmission Framework

Population size 2000, innovation rate (λ) 0.1 to 100

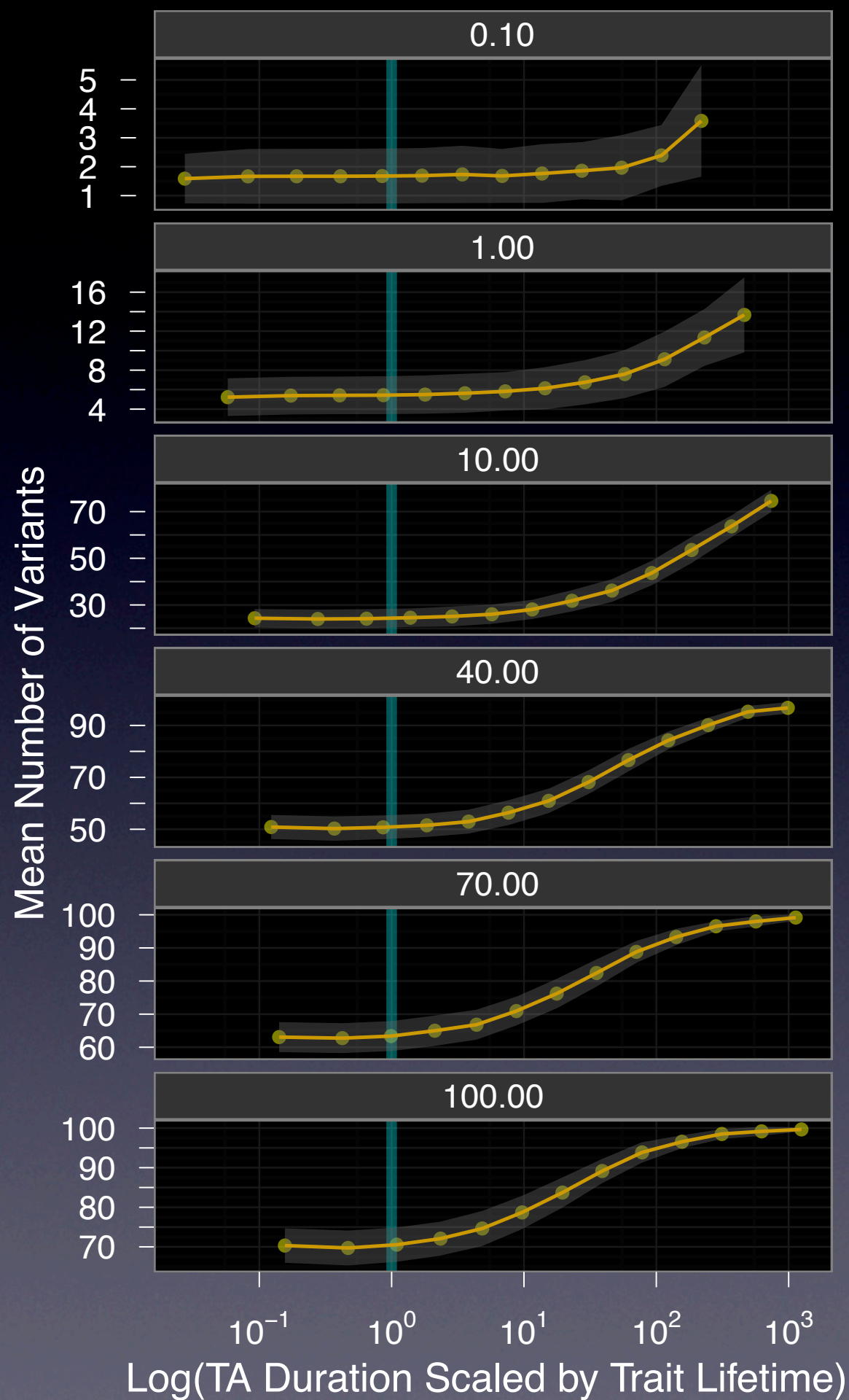
► Sampled 100 individuals in aggregated blocks from 3 and 8000 steps

Measurements

► 3MM+ samples of mean richness (K_n)

► 1.1MM+ samples of normalized t_f and Slatkin Exact tests

► Tracked trait lifetimes for 39.6MM+ traits



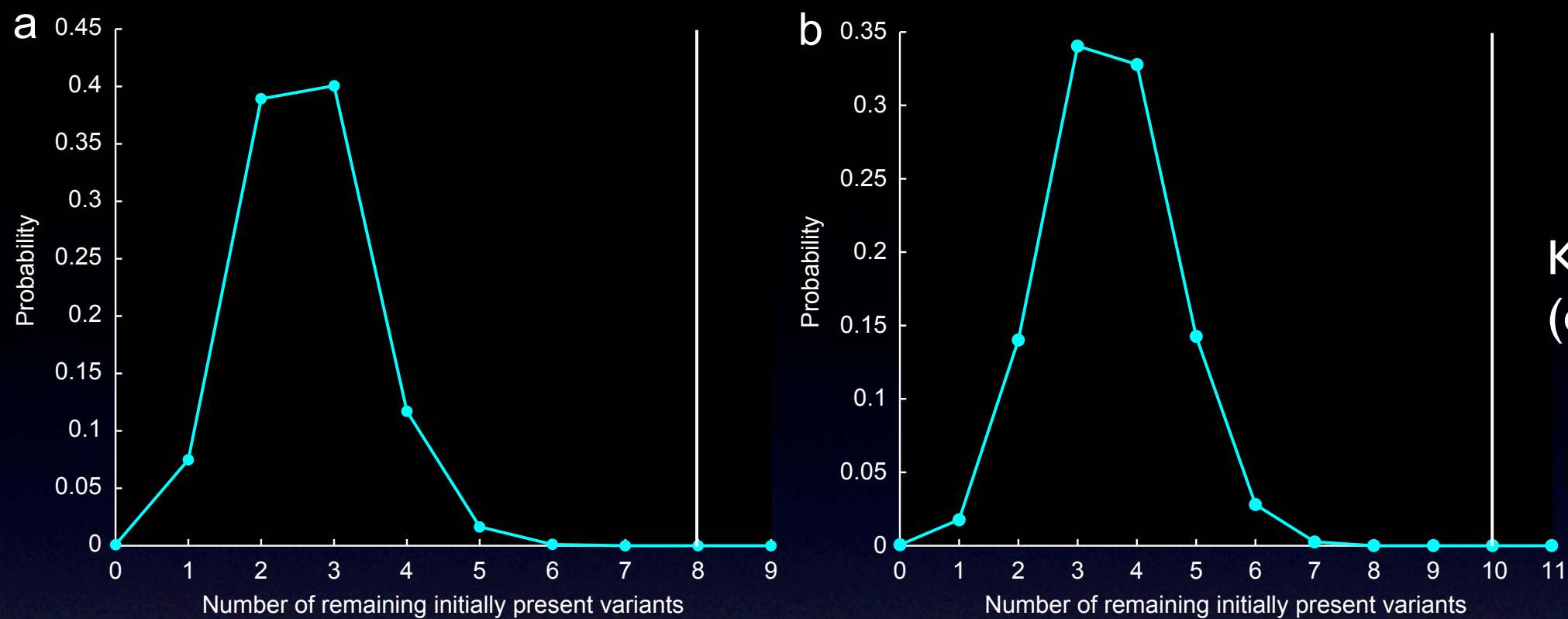
Number of variants is inflated

Effect strong at duration >
Average variant lifetime

Effect strong at high
innovation rates

Wednesday, June 5, 13

This means we have to be careful when we're interpreting the number of variants in an archaeological sample as evidence for or against a model of cultural transmission. And yet we do this *all* the time.



Kandler and Steele 2013
(excerpt Fig. 3)

Diachronic model predicts the surviving number of variants in pairs of LBK ceramic phases

Observed survival is much larger than predicted.

Is this pro-novelty bias or the inflationary effect of time averaging on class richness?

Wednesday, June 5, 13

Kandler and Steele's model, mentioned earlier, produces diachronic predictions by looking at how many traits observed at T1 are expected to survive to T2, given a particular population growth curve and CT process.

This should be right in our wheelhouse as archaeologists, because differences between assemblage-based observations are our bread-and-butter in data analysis.

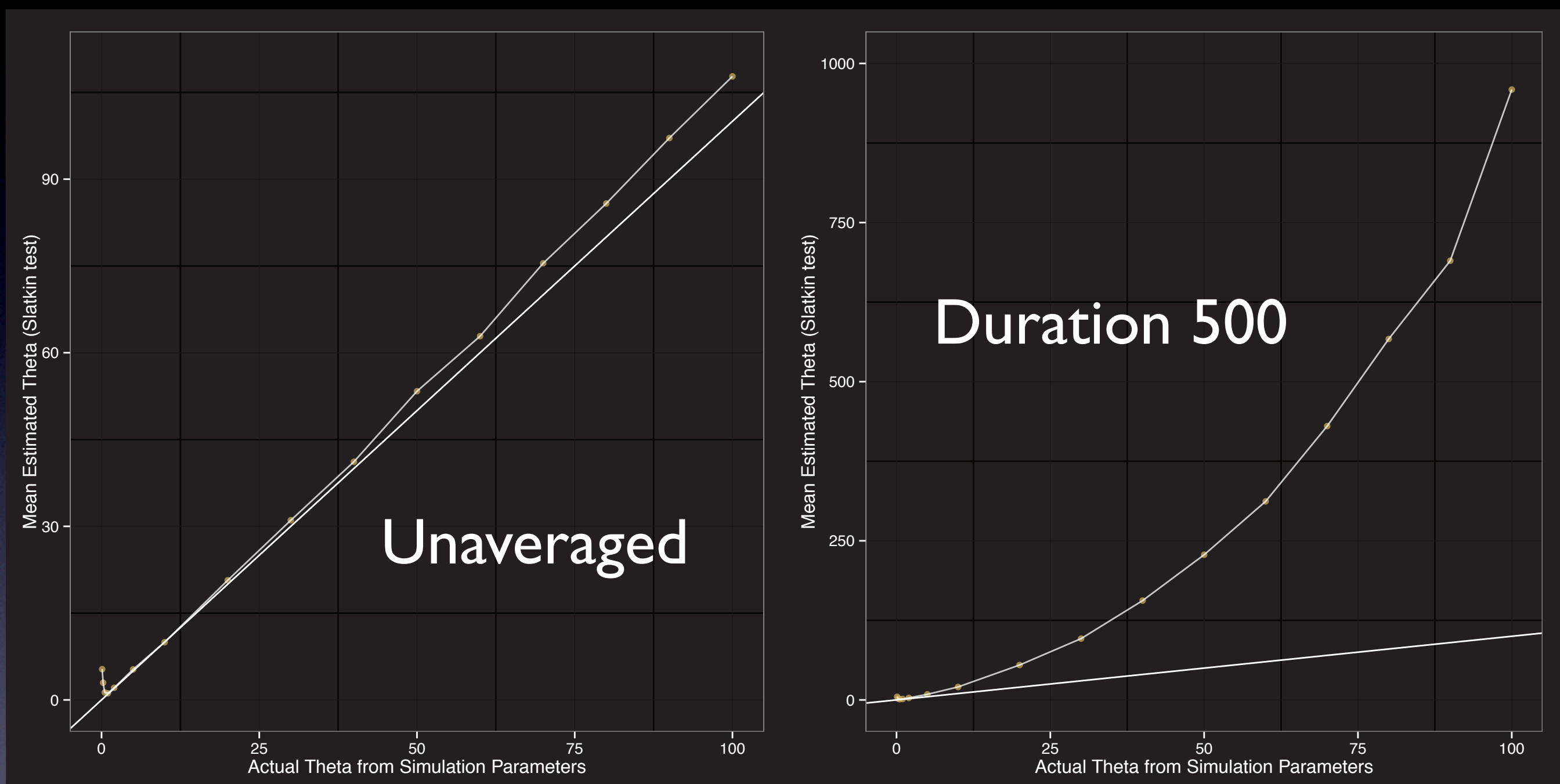
The curve represents the probability density for the number of "surviving" variants. The vertical line represents actual observations from early agricultural populations in Central Europe - the LBK or Linear Pottery Culture. This is a subset of K&S's actual analysis, showing the expectations for Phase I to Phase II, and then from Phase II to Phase III.

Note that in both (actually all) cases, the observed data are far larger than a neutral or unbiased copying model, even in the presence of population growth, would predict.

The authors conclude that this provides no evidence of neutral transmission for decorative motifs on LBK pottery, and possibly weak evidence for anti-conformist or "pro-novelty" bias in cultural transmission.

Or is this inflation of class richness, due to time averaging effects in time-transgressive assemblages?

Effects of Time Averaging Upon Estimation of Θ

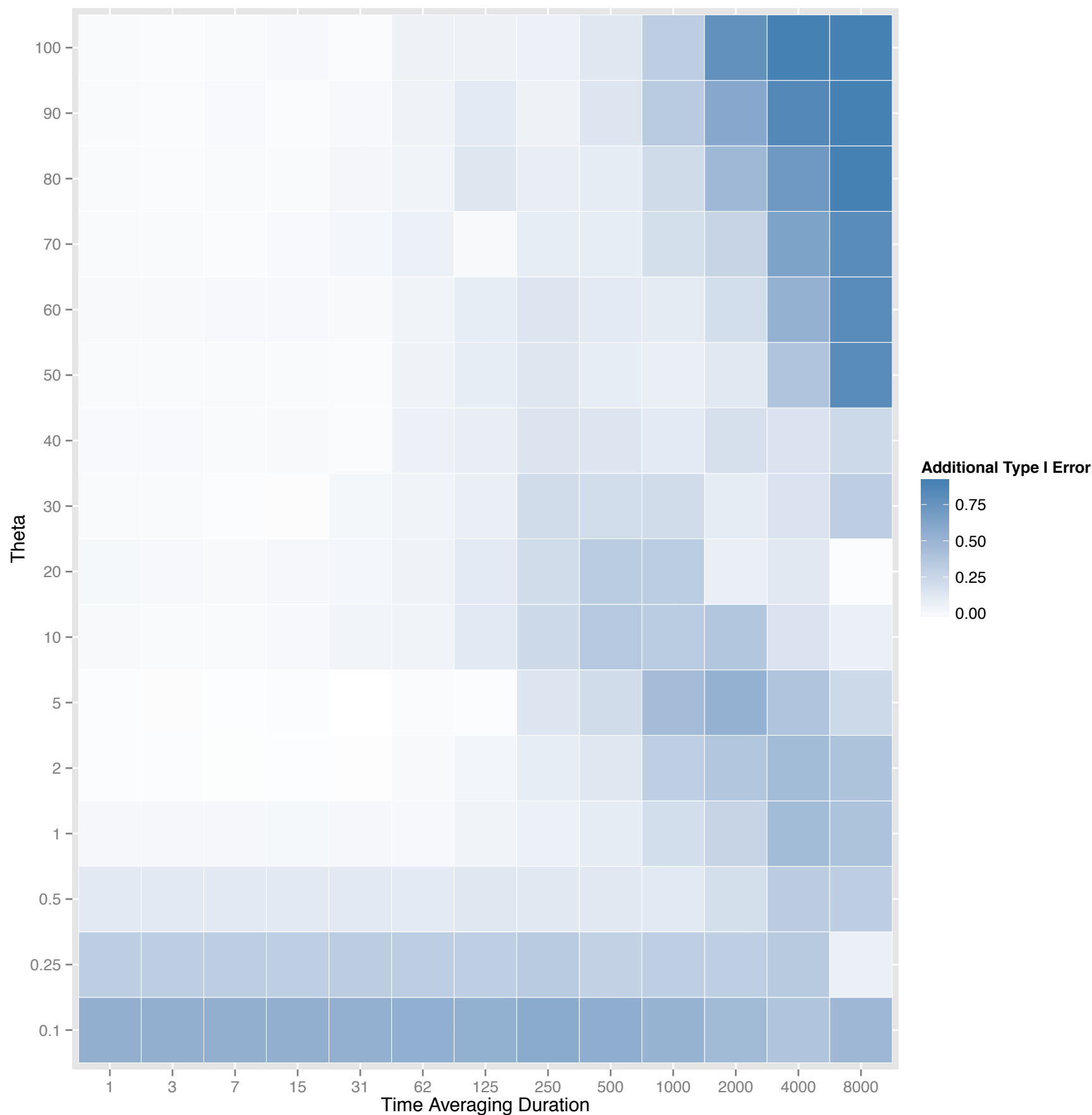


$$\Theta = 2 * \text{population size} * \text{innovation rate}$$

Wednesday, June 5, 13

Theta is estimated here via the Slatkin exact test, which tells us the most likely value of theta given the observed pattern of trait frequencies. Theta estimation in the infinite-alleles model is notoriously bad (better in infinite-sites given sequence data), but it's OK over a range of theta without TA. With TA, it's almost completely inaccurate except at theta = 1.0 (the dividing line between mutation- and drift-dominated behavior).

Important: Kohler et al. have tried to estimate innovation rates/population sizes, and other folks mention it -- Bentley and Shennan, etc.



Where do
neutrality tests
encounter
problems under
time averaging?

Wednesday, June 5, 13

1.1MM data points here, each one a Slatkin exact test at that level of population innovation and time averaging duration. That's about 5300 data points per cell.

Still, the theta levels below 1.0 are suspect, given extreme levels of variability and the difficulty of getting the process to settle down to a stationary distribution. So I don't yet trust these rows. Above that, we can see that neutrality tests are generally trustworthy at low durations, but there's a big region of higher innovation and higher duration where neutrality tests are basically useless in correctly detecting neutrality or unbiased copying.

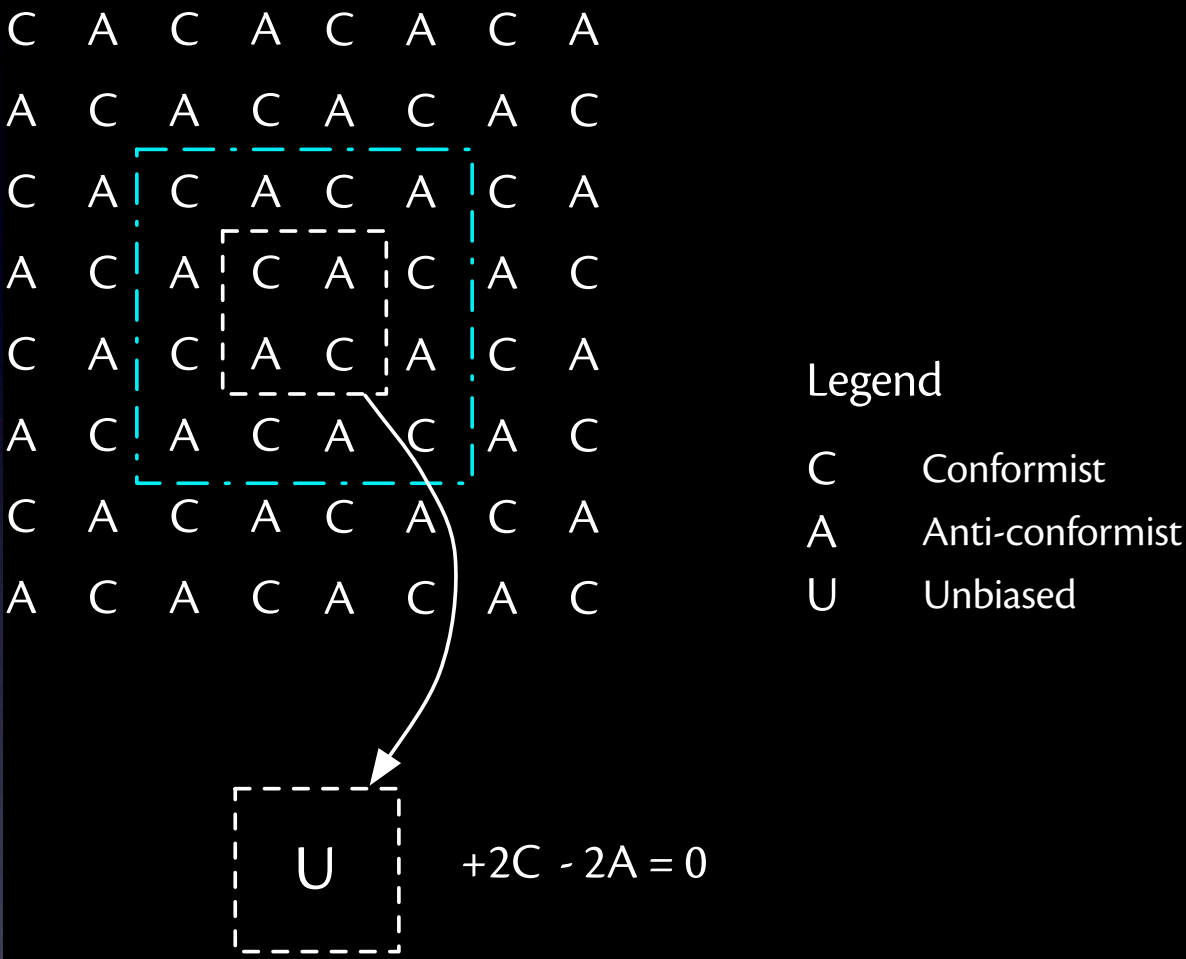
Use of neutrality tests is growing -- Scholnick, Premo, etc. So we have to be careful about how we use them, particularly when we don't know assemblage duration very well.

Real populations display diverse modes of social learning

Do all of the biases in CT cancel each other out, at the population level?

What tests are the most powerful discriminators?

How do the answers change with time averaging?



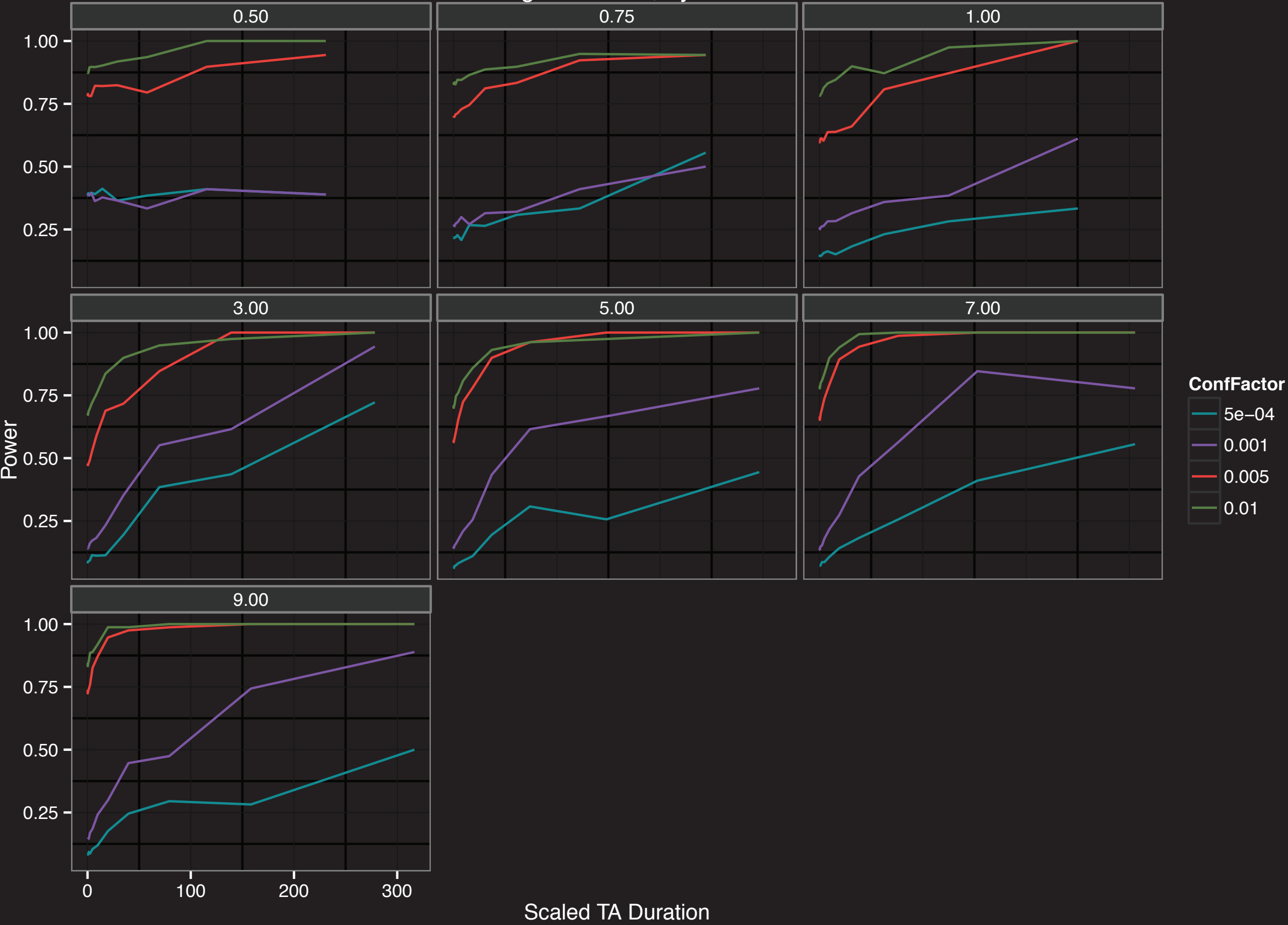
Wednesday, June 5, 13

It's easy to construct thought experiments or real simulations where the right mixtures of strategies in the population appear to “cancel out” biases like conformity and novelty-seeking, when you look at bulk samples or average/median values.

I'm doing numerical experiments to determine three things...

Some of these are in progress because the models are complicated with big parameter spaces, but I have some results today looking at standard tests of unbiased copying or neutrality, borrowed from population genetics....

Power and Scaled Assemblage Duration, by Conformism Factor and Theta



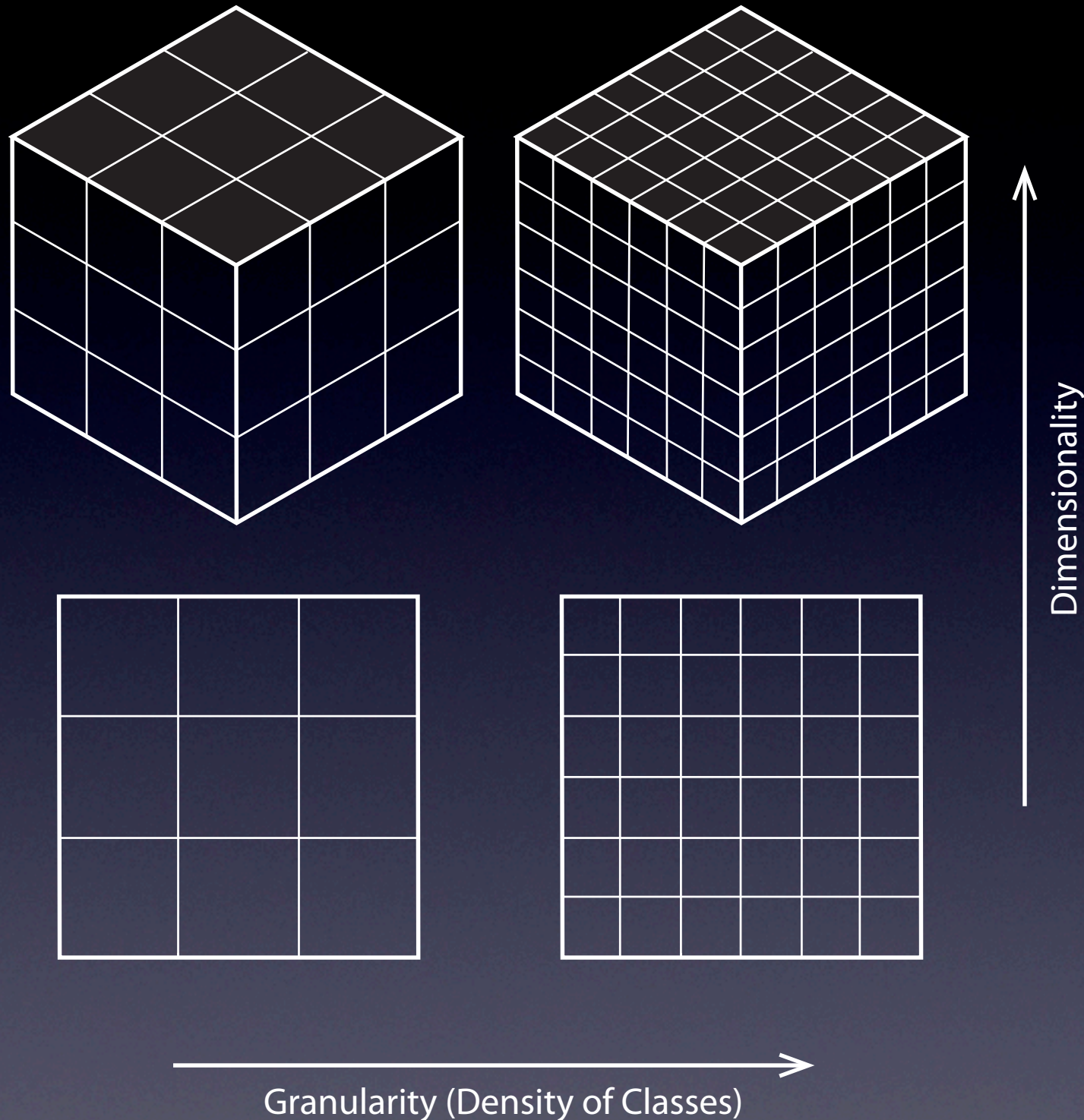
Wednesday, June 5, 13

Statistical power -- what is the chance that if our sample was NOT generated by the null hypothesis, that we'll conclude that it belongs in the Ho region of the test statistic? In other words, how often will we mis-identify a significant effect as insignificant?

Note that our ability to detect conformist transmission from trait frequency profiles depends *strongly* upon the strength of conformity in unaveraged samples.

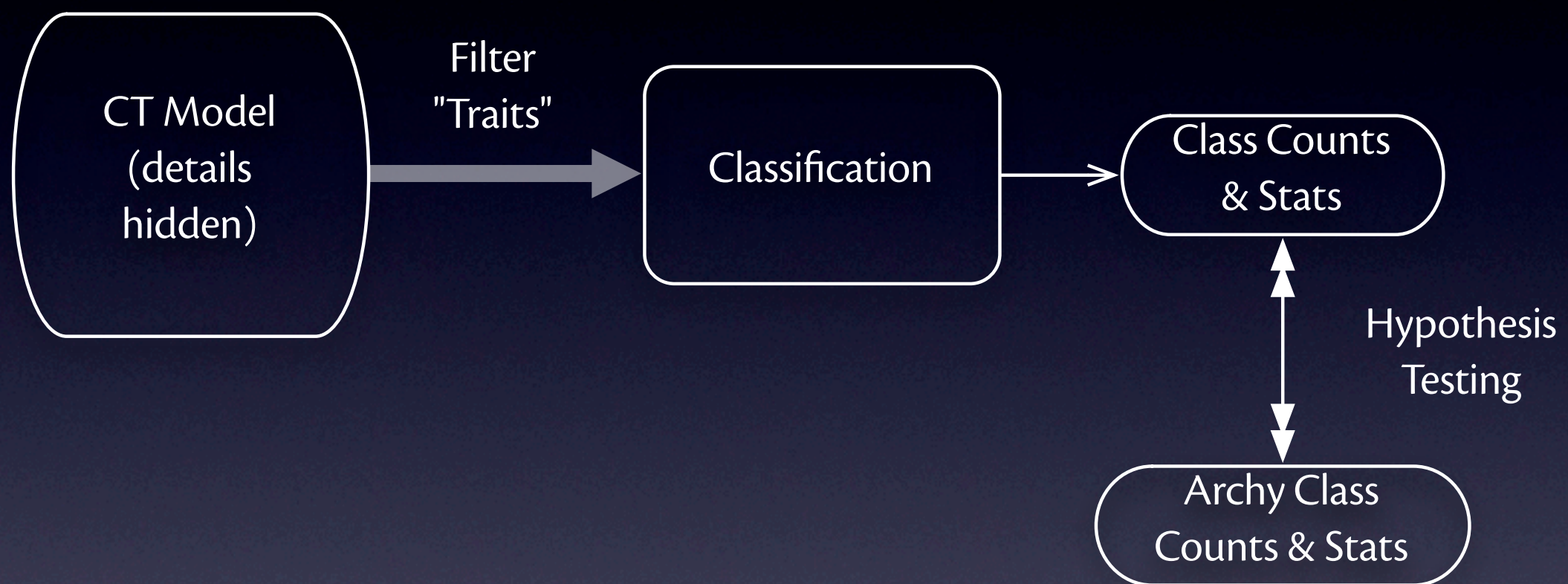
Time averaging actually enhances statistical power for our ability to detect conformism (even at weaker levels) -- but perhaps not with richer mixtures of social learning processes. That's an open question.

Classificatory Coarse Graining

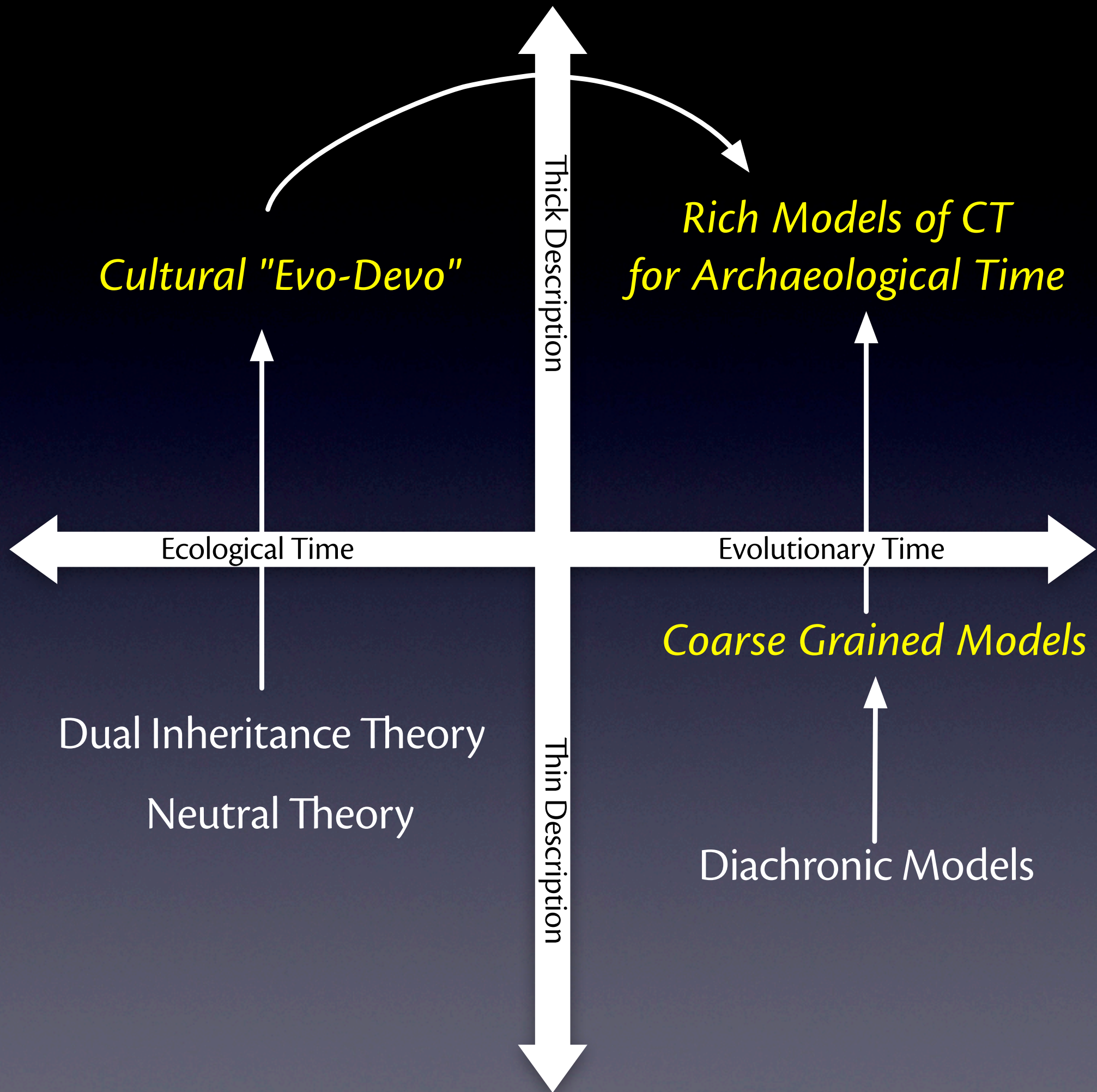


How do observables
from CT models change
when we measure
variation using types
with:

Different dimensionality
Different granularity



Schematic of CT models coarse grained through classification(s)



For more information:

mark@madsenlab.org

<http://notebook.madsenlab.org>



(c) 2013 Mark E. Madsen. This content is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License

Appendix

Objection: cultural behavior and information is too different from genetics to be modeled the same way

Cultural traits are **NOT**

- Alleles at fixed loci
- “Coded” replicators

Cultural traits **ARE**

- Networks of actions, facts, and strategies, linked to contexts
- Heritable given social, environmental, and developmental “scaffolding”
- Reconstructed, not transcribed, into products

Wednesday, June 5, 13

Note that these differences are often cause for people to claim that culture and cultural transmission processes must lie outside the realm of Darwinian theory, and natural selection. There are several folks who beat this drum regularly.

The usual reason for this criticism is that genes have the properties of replicators, whereas these kinds of differences are taken to mean that cultural information doesn't act in the right way to be “inheritance” from a Darwinian perspective

But when you look below the surface of the simplest textbook examples, genetics isn't simple either, and in fact is at least as complex as we think culture is....

BUT...Genetics is in the same boat...

- Genes require lots of “scaffolding”
- Gene effects are always reconstructed (“evo-devo”)
- Eventual result depends upon epigenetic variation
- The line between “acquired” and “inherited” is blurrier than critics think...



Marmosets: *chimerism*

Born as twins, each sharing sibling's genotype and parental genotypes

Oak: *mosaicism*

Each branch can develop a divergent genotype



Wednesday, June 5, 13

Marmosets: the chorion membranes of the twins merges, and embryonic stem cells are shared. Each marmoset has tissues that are genetically distinct, and are true chimeras. Minor example of how the germ and somatic cells are not always terribly well segregated, and this is in complex animals with specialized reproductive cells! It's much murkier in smaller, less differentiated organisms.

And even weirder in bacteria and plants. Without getting exotic, let's take the common oak tree....a mature oak's branches might be a hundred years old, and since plant cells don't move, any differentiation or mutation in the cells of each branch, particularly in the apical meristem, is then heritable by the reproductive cells constructed on that branch. An oak tree isn't an individual genome, reproducing, it's a family of closely related genomes, slowly diverging through acquired variation.