Theories and Models of the Evolution of Altruism

Unification vs. Unique Explanations

Jeffrey A. Fletcher

Systems Science Graduate Program

Portland State University

Outline

- Intro
- Example of unification effort
  - o Hamilton's rule applied to Reciprocal Altruism (including mutualisms)
- Example of framework emphasizing role of assortment
  - o Interaction Environments
- Examine claim that only Inclusive Fitness explains "true" altruism
  - o Implications for doing Science with Models

The Problem

- How can natural selection favor individuals that carry helping traits, over those that carry selfish ones?

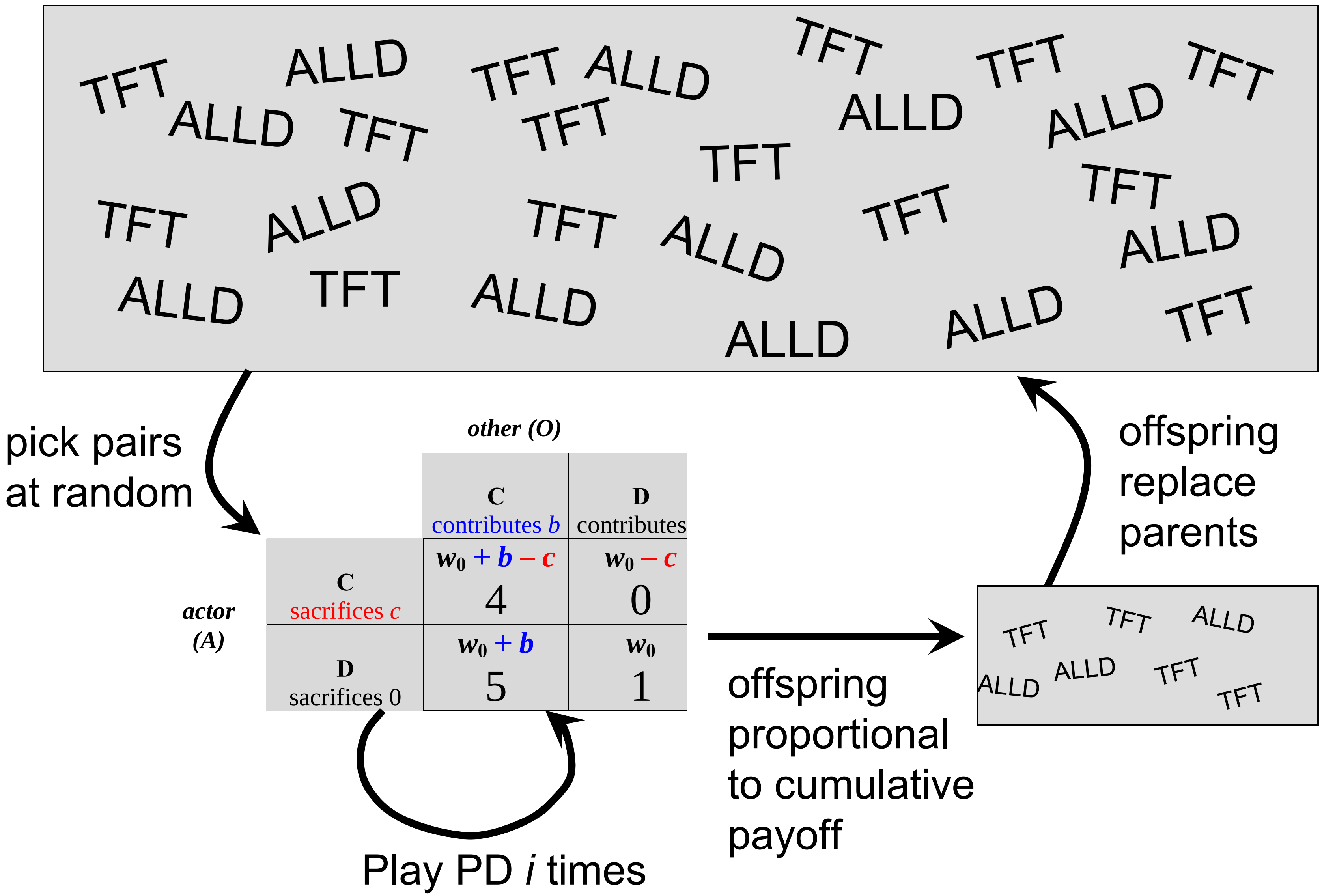
Main Theories for the Evolution of Altruism

- Multilevel (Group) Selection
  - Altruist dominated groups do better; altruists within groups do worse
  - $\Delta Q = \Delta Q_B + \Delta Q_W$
- Inclusive Fitness/Kin Selection
  - Gene self interest, Hamilton's rule ($\Delta Q > 0$ if $rb > c$)
  - $w_{inclusive} = w_{direct} + w_{indirect}$
- Reciprocal Altruism
  - Conditional behaviour, Iterated Prisoner's Dilemma (PD), emphasis on non-relatives, mutualism
  - Indirect reciprocity, strong reciprocity, reciprocity on graphs
- Others
  - By-product mutualism, conflict mediators, policing, social markets

Reciprocal Altruism Model

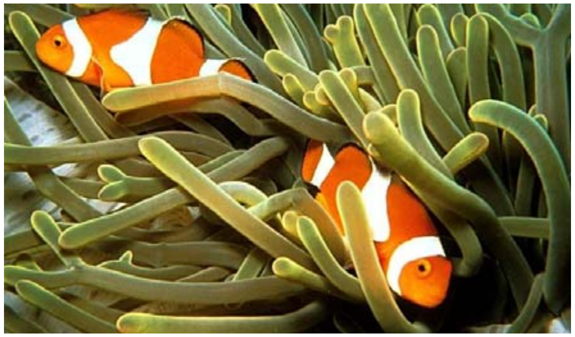
- Interactions modeled as a Prisoner's Dilemma Game (PD)
- Iterated conditional behaviours
  - o Genotype (G) no longer determines Phenotype (P)
- Axelrod's Tournaments (late 1970s on)
  - o Tit-For-Tat (TFT)
    - Anatol Rapoport
- Evolutionary experiments
  - o Random interactions
  - o offspring proportional to cumulative payoffs

Simple Iterated PD Model



Axelrod and Hamilton (1981)

- Distinguished two mechanisms
  - Inclusive Fitness for relatives
  - Reciprocal Altruism for non-relatives
- Why didn't Hamilton apply Hamilton's rule?
- Two obstacles for unification
  1. Phenotype/Genotype differences
  2. PD used is non-additive



Additive PD

		other (O)	
	actor's fitness (utility)	**C** contributes b	**D** contributes 0
actor (A)	**C** sacrifices c	$w_0 + b - c$ 4	$w_0 - c$ 0
	D sacrifices 0	$w_0 + b$ 5	w_0 1

- $w_0 = 1$; $b = 4$; $c = 1$

Non-additive PD

		other (O)	
	actor's fitness (utility)	**C** contributes b	**D** contributes 0
actor (A)	**C** sacrifices c	$w_0 + b - c + d$ 3	$w_0 - c$ 0
	D sacrifices 0	$w_0 + b$ 5	w_0 1

- $w_0 = 1$; $b = 4$; $c = 1$; $d = -1$ (diminishing returns)

Queller's Generalization

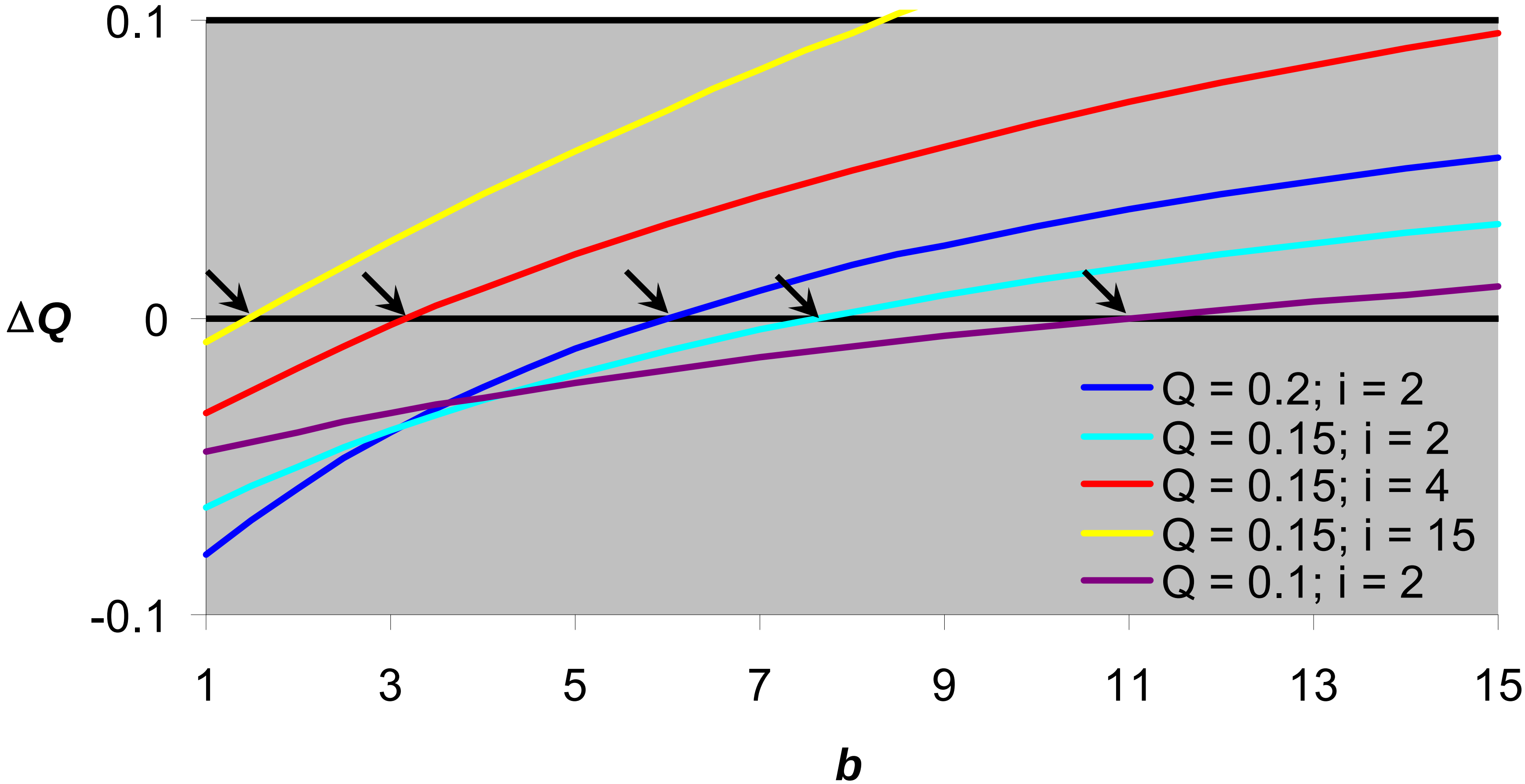
- To solve problem 1 (*G*/*P* difference)
  - o Use *phenotypes* (behaviours), not just *genotypes*, in Hamilton's rule
  - o Hamilton (1975) Queller (1985)

$$r = \frac{\mathrm{cov}(G_A, G_O)}{\mathrm{var}_t(G_A)} \qquad r = \frac{\mathrm{cov}(G_A, P_O)}{\mathrm{cov}(G_A, P_A)}$$

- To solve problem 2 (non-additivity)
  - – Use an additional term to account for deviations from additivity (Queller 1985)

$$\frac{\mathrm{cov}(G_A, P_O)}{\mathrm{cov}(G_A, P_A)} b + \frac{\mathrm{cov}(G_A, P_A P_O)}{\mathrm{cov}(G_A, P_A)} d > c$$

Numerical Simulations of Iterated PD varying *Q*, *i*, and *b* (*c* = 1)

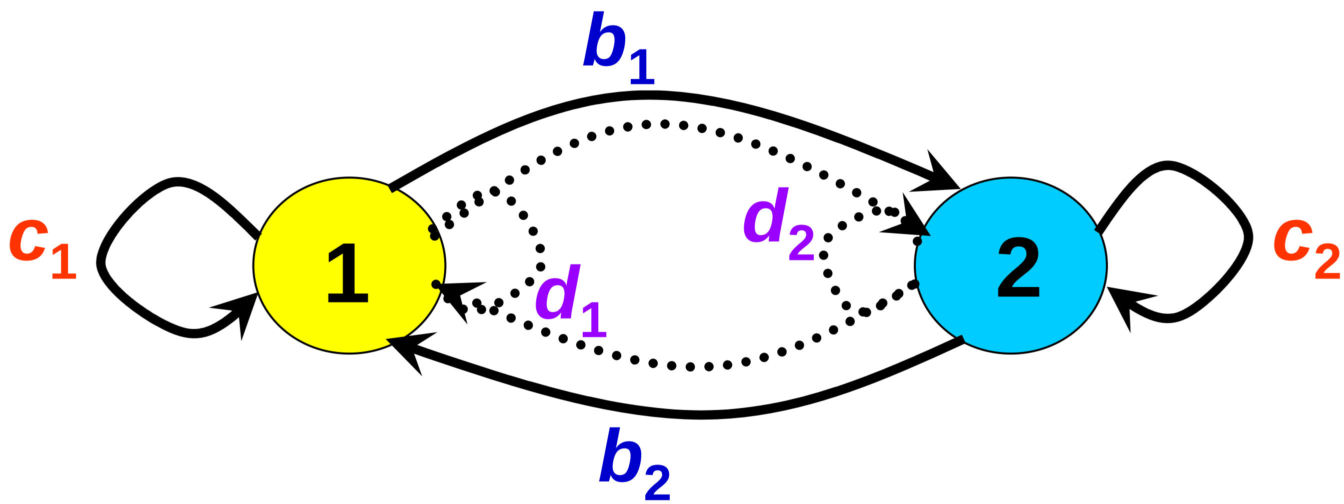


- Fletcher & Zwick, 2006. *The American Naturalist*

A Simple Mutualism Model

- Interactions are heterospecific and pair-wise
- Each species has two types
  - o ALLD type
  - o a cooperative type (e.g. TFT)
- *b, c, d,* and the cooperative strategy can all vary between species

A Simple Mutualism Model



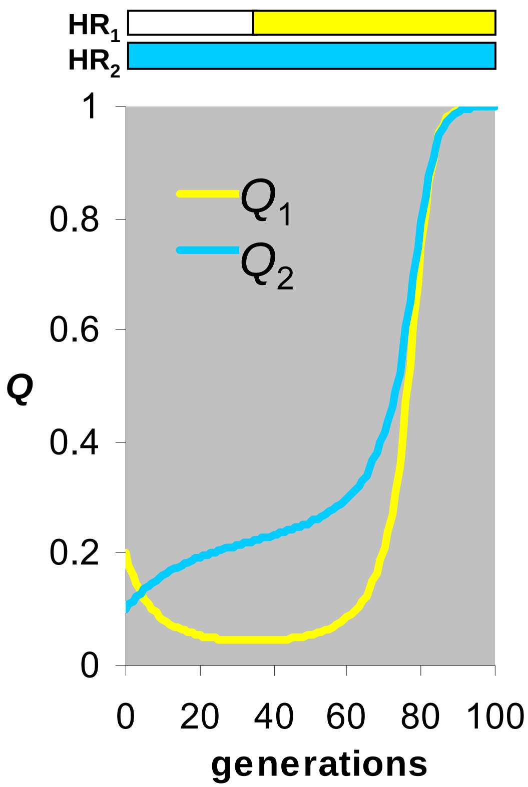
$$r_1 = \frac{\mathrm{cov}(G_1, P_2)}{\mathrm{cov}(G_1, P_1)}$$

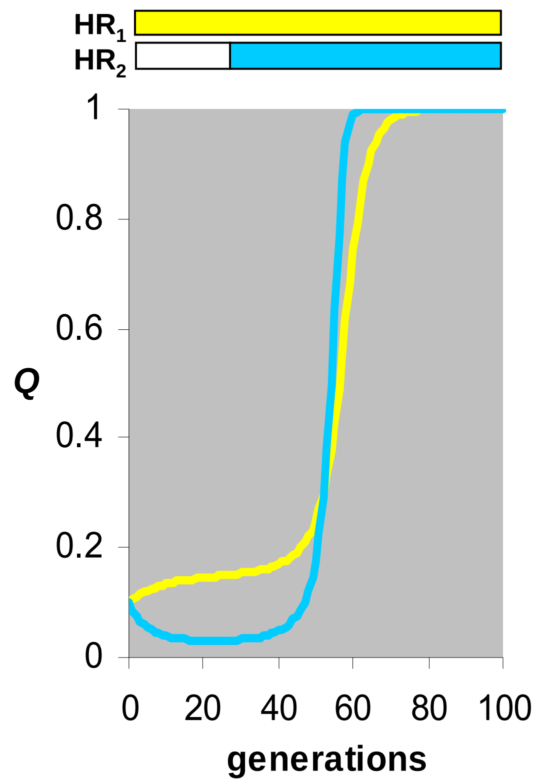
$$r_2 = \frac{\mathrm{cov}(G_2, P_1)}{\mathrm{cov}(G_2, P_2)}$$

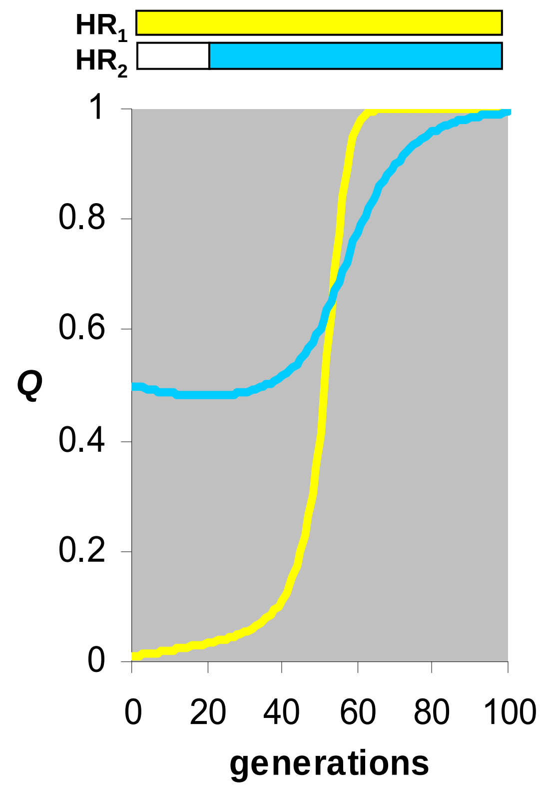
HR$_1$: $r_1 b_2 > c_1$

HR$_2$: $r_2 b_1 > c_2$

– Fletcher & Zwick, 2006. *The American Naturalist*







	1	**2**
b	1.5	5
c	2.0	0.1
w_0	2	2
d	0	0
Str	TFT	TFT
i	4	

	1	**2**
b	4	4
c	1	1
w_0	1	1
d	0	0
Str	TF2T	Pavlov
i	80	

	1	**2**
b	2	2.2
c	1	0.1
w_0	1	1
d	0	1.3
Str	ALLC	TFT
i	100	

– Fletcher & Zwick, 2006. *The American Naturalist*

There is no general theory of mutualism that approaches the explanatory power that 'Hamilton's Rule' appears to hold for the understanding of within-species interactions.

- Herre et al. 1999, *TREE* 14:49-53

Back to Basics of Selection

- Queller's version emphasizes direct fitness; no G_O term—genotype of *Other* irrelevant!

$$\frac{\operatorname{cov}(G_A, P_O)}{\operatorname{cov}(G_A, P_A)} b + \frac{\operatorname{cov}(G_A, P_A P_O)}{\operatorname{cov}(G_A, P_A)} d > c$$

- More intuitive form

$$\operatorname{cov}(G_A, P_O)\, b + \operatorname{cov}(G_A, P_A P_O)\, d > \operatorname{cov}(G_A, P_A)\, c$$

- An even simpler form

$$\operatorname{cov}(G_A,\ P_O\, b + P_A P_O\, d - P_A\, c) > 0$$

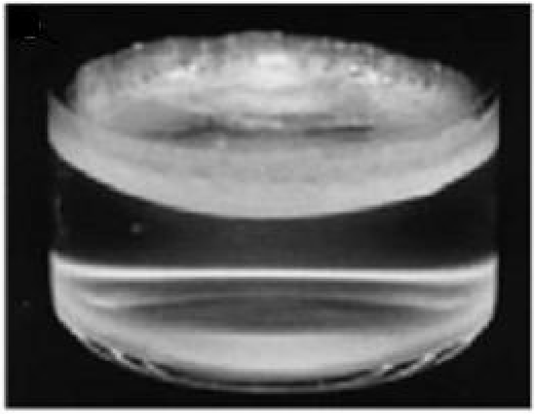
$$\operatorname{cov}(G_A,\ net\ fitness\ benefits\ to\ A) > 0$$

Outline

- Intro
- Example of unification effort
  - Hamilton's rule applied to Reciprocal Altruism (including mutualisms)
- Example of framework emphasizing role of assortment
  - Interaction Environments
- Examine claim that only Inclusive Fitness explains "true" altruism
  - Implications for doing Science with Models

Simple Public Goods Game

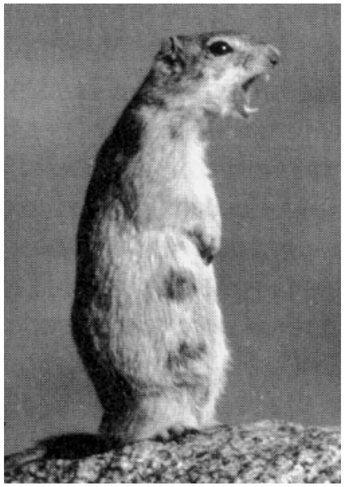
- Two types of behaviors
  - o Cooperate (C) and Defect (D)
- C and D behaviors have simple genetic basis
- Interaction environments of *N* individuals; split benefits evenly
- C behavior contributes *b*, at cost *c*
- $b > c$ (non-zero-sum-ness)
- D behavior contributes nothing and imposes no cost



Partition Single Interaction Environment

Phenotype	Payoff received from own behavior	Payoff received from the behavior of others in interaction environment (excluding self)	Total direct payoff (within group)
Cooperate (C)	$\frac{b}{N} - c$	$\frac{(k-1)b}{N}$ [k-1 cooperators, N-k defectors]	$\frac{kb}{N} - c$
Defect (D)	0	$\frac{kb}{N}$ [k cooperators, N-k-1 defectors]	$\frac{kb}{N}$

- Within any interaction environment, defectors (D) do better than cooperators (C)
- But C can be selected for when we consider a whole system of interaction environments
- This is the basic dilemma of altruism



– Fletcher & Doebeli, 2009. *Proceedings B*

Average Interaction Environment

- A "mean field" approach to social interactions
- Let e_C and e_D be average interaction environments of C and D individuals, respectively
- Measure e_C and e_D as the number of C behaviors among interaction partners (here N-1)
- Compare e_C with e_D

– Fletcher & Doebeli, 2009. *Proceedings B*

Partition Average Interaction Environment

Phenotype	**Average payoff received from own behavior**	**Average payoff received from others' behaviors in average interaction environment (excluding self)**	**Average total payoff**
Cooperate (C)	$\frac{b}{N} - c$	$\frac{e_C b}{N}$	$\frac{e_C b}{N} + \left(\frac{b}{N} - c\right)$
Defect (D)	0	$\frac{e_D b}{N}$	$\frac{e_D b}{N}$

- The condition for C genotype to increase: average net payoff to C is greater than average net payoff to D

$$\frac{e_C b}{N} + \left(\frac{b}{N} - c\right) > \frac{e_D b}{N}$$

- This is true of any trait!

– Fletcher & Doebeli, 2009. *Proceedings B*

Interaction Structures

$$\frac{e_C b}{N} + \left(\frac{b}{N} - c\right) > \frac{e_D b}{N} \qquad e_C - e_D > \frac{cN}{b} - 1$$

- Random Binomial Distribution: $e_C = e_D$
  - o Dividing line between weak ($b/N > c$) and strong altruism ($b/N < c$)
- Over Dispersion: every environment has one C
  - o $e_C = 0$; $e_D = 1$ (C decreases even if weak: $b/N > c$)
- Extreme Assortment: only C with C; D with D
  - o $e_C = N$-1; $e_D = 0$ (C increase if $b>c>0$)

– Fletcher & Doebeli, 2009. *Proceedings B*

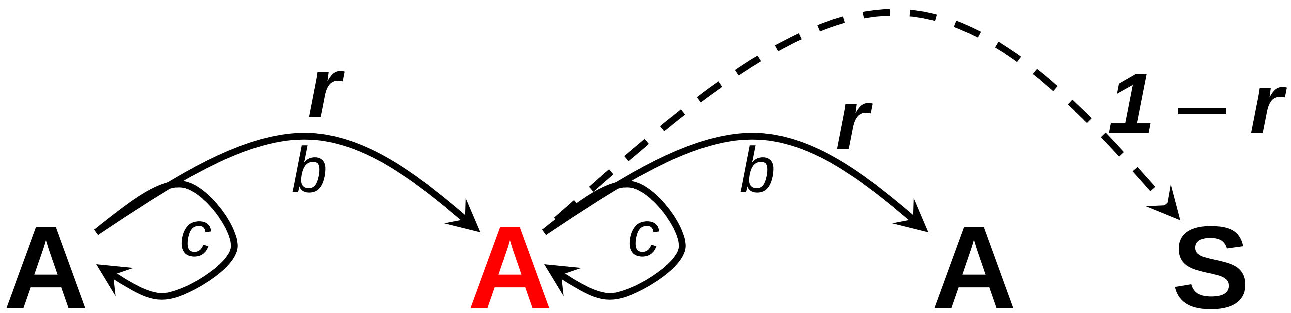
Outline

- Intro
- Example of unification effort
  - o Hamilton's rule applied to Reciprocal Altruism (including mutualisms)
- Example of framework emphasizing role of assortment
  - o Interaction Environments
- Examine claim that only Inclusive Fitness explains "true" altruism
  - o Implications for doing Science with Models

Claim (Hypothesis)

- “True” altruism only evolves via inclusive fitness (kin selection)
- "Direct benefits explain mutually beneficial cooperation, whereas indirect benefits explain altruistic cooperation”
  - o West et al. 2007, *JEB*.

Inclusive Fitness Concept
Direct Fitness Concept



- $w_{direct}(A) = w_{baseline} - c + rb$
- ~~$w_{indirect}(A) = rb$~~
- ~~$w_{inclusive}(A) = w_{direct}(A) + w_{indirect}(A)$~~
- ~~$w_{inclusive}(A) = w_{baseline} - c + rb$~~
- Hamilton's rule: $rb - c > 0$

Defining Altruism

- “Altruism: a behaviour which is costly to the actor and beneficial to the recipient...”
- “A general point here is that altruism is defined: (i) with respect to the lifetime consequences of a behaviour; (ii) on absolute fitness effects (i.e. does it increase or decrease the actor’s fitness, and not relative to just some subset of the population).”
- “For example, if a cooperative behaviour was costly in the short term, but provided some long-term (future) benefit, which outweighed that, it would be mutually beneficial and not altruistic.”
    - West et al. JEB 2007

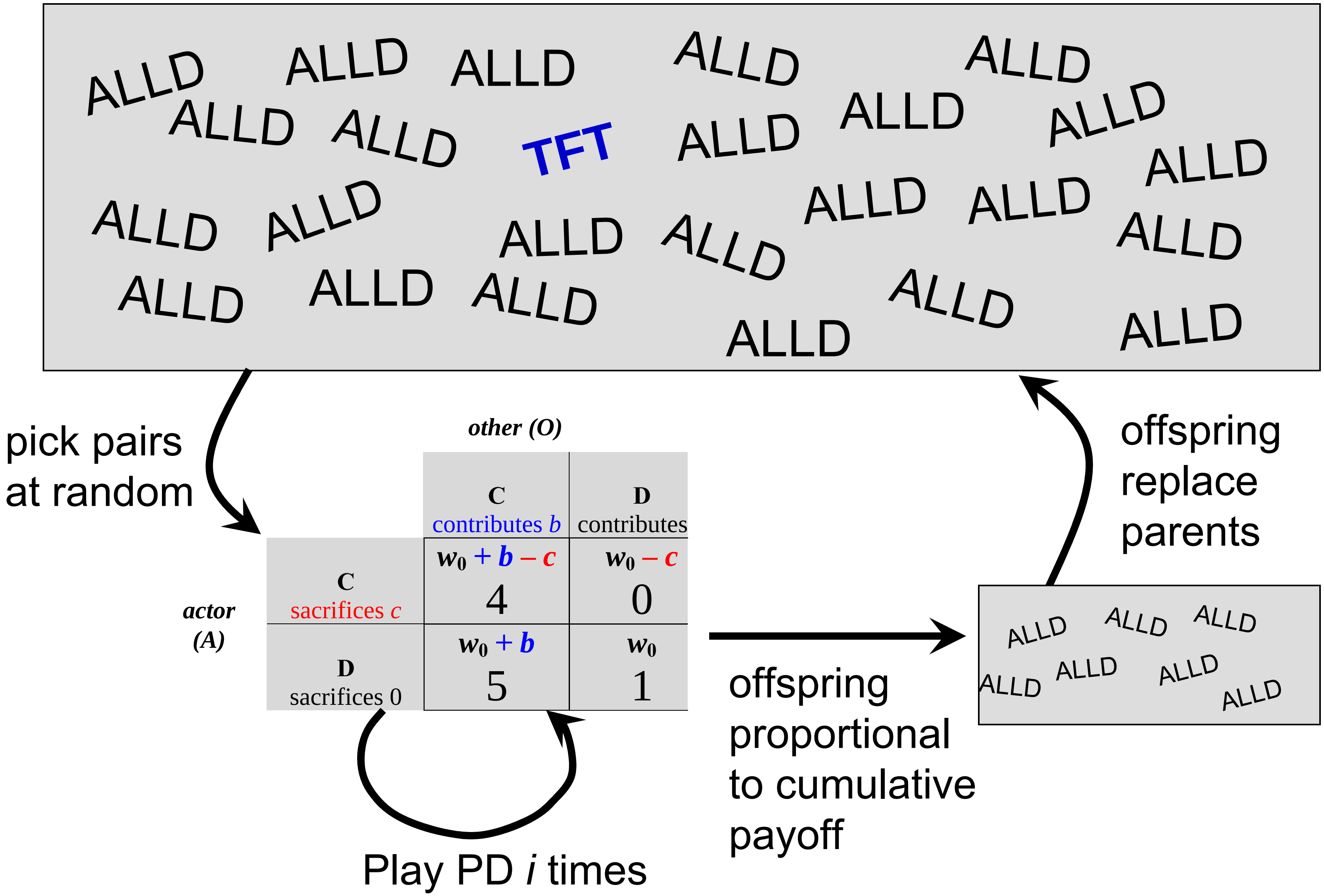
Issues in Defining Altruism

- Distribution of behaviours across population
  - o All individuals both givers and receivers
    - lifetime cost means altruism cannot evolve
  - o Strict separation of givers and receivers
    - (e.g. suicidal aid, sterility)
    - phenotype defines altruism; but analysis in terms of genotype frequency (which is the same)
  - o What qualifies as an altruistic genotype?

Where to Draw the Hierarchal Line?

- What constitutes a “lifetime cost” of a behaviour?
- What are the assumptions about individual influence on interaction structure?
- Is it OK that individuals become true altruists or not depending on their context (which they do not perceive or control)?

Simple Iterated PD Model



The Role of Models in Science?

- Is the claim a falsifiable hypothesis?
- What model could test this hypothesis?
- Models are simplifications
    - o We choose what is in and out
    - o Want to capture just what is essential
- Empiricists are more advanced in guarding against biases
- Need to learn and use each others' models

Conclusion

- Various theories of the evolution of altruism rely on the same underlying requirement for sufficient assortment between the genotype in question and help from others
- This is captured in Queller's version of Hamilton's rule and the notion of interaction environments
- Inclusive fitness is an accounting method, not a fundamental mechanism
- Testable Hypothesis: true altruism can evolve without interactions among kin (or genetically similar individuals)

Acknowledgements

- Coauthors
  - Michael Doebeli, David Wilson, Martin Zwick
- For Helpful Discussions
  - Ingi Agnarsson, Leticia Aviles, Alistair Blachford, Sam Bowles, Felix Breden, Kevin Foster, Fred Guillaume, Guy Hoelzer, Benn Kerr, Laurent Lehmann, Martin Nowak, Len Nunney, Sally Otto, John Pepper, David Queller, Patricio Salazar, Peter Taylor, Tom Wenseleers, Stuart West, Geoff Wild, Jon Wilkins, Ron Ydenburg
  - Avilés lab, Doebeli lab, and SOWD discussion group at UBC
- For Funding
  - NSF International Research Fellowship

Thanks!