

广义适合度理论之争的 概念研究

(申请清华大学理学硕士学位论文)

培 养 单 位 人文学院

学 科 科学技术史

研 究 生 王 康 桥

指 导 教 师 王 巍 教 授

二〇二四年九月

广义适合度理论之争的概念研究

王
康
桥

A Conceptual Study of the Inclusive Fitness Theory Controversy

Thesis submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Science

in

History of Science and Technology

by

Wang Kangqiao

Thesis Supervisor: Professor Wang Wei

September, 2024

学位论文公开评阅人和答辩委员会名单

公开评阅人名单

陈勃杭	研究员	浙江大学
陆俏颖	助理教授	北京大学
蒋 澈	副教授	清华大学

答辩委员会名单

主席	杨 舰	教授	清华大学
委员	孙承晟	教授	清华大学
	蒋 澈	副教授	清华大学
	陆俏颖	助理教授	北京大学
	王 巍	教授	清华大学
秘书	张晓天	助理研究员	清华大学

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；（3）按照上级教育主管部门督导、抽查等要求，报送相应的学位论文。

本人保证遵守上述规定。

作者签名： 王康桥

日 期： 2024年9月10日

导师签名： 王康

日 期： 2024.9.10

摘 要

广义适合度理论作为社会进化研究的重要理论在近年来饱受争议,争议内容既包括广义适合度理论的核心内容汉密尔顿规则的解释价值问题,也包括广义适合度理论、亲缘选择与真社会性进化的关系等问题。本文对广义适合度理论之争展开概念研究。就广义适合度理论自身的问题,我将论证支持者公认的一般汉密尔顿规则作为社会进化研究的组织框架的解释价值是成问题的,并提供一种新的方案来理解多种汉密尔顿规则的一般形式的解释价值。就广义适合度理论、亲缘选择与真社会性进化的关系问题,我将通过不同的亲缘系数概念和因果概念的区分将争议双方的争议点定位在系谱亲缘系数在真社会性的进化中的因果效力上,并指出批评者更倾向谈论净效果而支持者更倾向谈论组分效果。就广义适合度理论与能动者思维的关系问题,我将澄清奥卡沙对不同类型能动者思维的一些混淆,论证广义适合度最大化和能动者思维之间并无紧密联系。

关键词: 广义适合度; 汉密尔顿规则; 亲缘选择; 真社会性; 能动者思维

Abstract

There is an ongoing debate over inclusive fitness theory as an important theory of social evolution in recent years. The issues under controversy includes the explanatory value of Hamilton's rule as core content of inclusive fitness theory and the relationship among inclusive fitness theory, kin selection, and the evolution of eusociality. This paper is a conceptual study of the inclusive fitness controversy. Concerning the issues within inclusive fitness theory, I will argue that the proponents' consensus that the general form of Hamilton's rule can serve as the organizing framework of social evolution research is problematic, and I will propose a new way to understand the explanatory value of different general forms of Hamilton's rule. Concerning the relationship among inclusive fitness theory, kin selection, and the evolution of eusociality, I will differentiate different concepts of relatedness and different concepts of causation to locate the focus of the controversy on the causal role of pedigree relatedness. I will point out that critics care more about net effect of relatedness while proponents care more about component effect. Concerning the relationship between inclusive fitness theory and agential thinking, I will clarify some confusion on different types of agential thinking in Okasha's treatment of the issue and argue that there is no tight connection between inclusive fitness maximization and agential thinking.

Keywords: inclusive fitness; Hamilton's rule; kin selection; eusociality; agential thinking

目 录

摘 要	I
Abstract	II
目 录	III
插图和附表清单	V
符号和缩略语说明	VI
第 1 章 引言	1
1.1 选题背景	1
1.2 文献综述	4
1.2.1 广义适合度理论之争：背景、批评与辩护	4
1.2.2 广义适合度最大化和最优化	6
1.2.3 中文文献状况	10
1.3 研究方法	12
1.4 论文结构安排	12
第 2 章 广义适合度理论中汉密尔顿规则的解释价值	13
2.1 引言	13
2.2 汉密尔顿规则的不同版本	14
2.3 一般汉密尔顿规则的解释价值	17
2.4 对伯奇的反驳	19
2.4.1 分类的不稳定性	19
2.4.2 与其它分类方式的不兼容性	21
2.5 汉密尔顿规则的多种一般形式及其解释价值	24
2.6 本章小结	26
第 3 章 广义适合度理论、亲缘选择与真社会性的进化	27
3.1 引言：真社会性的进化解释问题	27
3.2 IBD 亲缘系数、IBS 亲缘系数与统计亲缘系数	29
3.3 真社会性的进化解释争议中的亲缘系数	31
3.4 亲缘系数与关于自然选择的统计主义之争	34
3.5 真社会性的进化中亲缘系数的因果效力	36
3.5.1 因果相关性与类型因果	36
3.5.2 净效果与组分效果	38

3.6 本章小结	40
第 4 章 广义适合度最大化、最优化与能动者思维	41
4.1 引言	41
4.2 支付可加情形下的广义适合度最大化	42
4.3 支付不可加情形下的广义适合度最大化	45
4.4 本章小结	49
第 5 章 结论	50
参考文献	51
致 谢	58
声 明	59
个人简历、在学期间完成的相关学术成果	60
指导教师评语	61
答辩委员会决议书	62

插图和附表清单

图 2.1 伯奇通过 HRG 对社会进化的不同解释的归类.....	18
图 2.2 诺瓦克等人用以指出 HRG 中参数不稳定性的案例.....	20
图 2.3 扩展的 HRG 提供的分类图示.....	22
图 2.4 一个社会进化的案例	23
图 2.5 基于 HRG 和扩展的 HRG 的分类比较	24
表 1.1 格拉芬“形式达尔文主义项目”早期成果概要	7
表 1.2 11 篇评论对形式达尔文主义项目的态度	8
表 2.1 可加支付条件下的支付矩阵	15
表 4.1 可加情形的个体支付	43
表 4.2 可加情形的广义适合度或效用	43
表 4.3 效用取广义适合度时的进化-理性链接	43
表 4.4 不可加情形的个体支付	45
表 4.5 不可加情形将广义适合度作为效用	47
表 4.6 不可加情形将格拉芬 1979 支付作为效用	47
表 4.7 效用取格拉芬 1979 支付时的进化-理性链接	47

符号和缩略语说明

B 或 b	行为为搭档带来的收益
C 或 c	行为主体付出的代价
R 或 r	亲缘系数
D 或 d	行为的协同效应
s	协同系数
w	适合度
g	育种值
g'	行为搭档的（平均）育种值
$\Delta\bar{g}$	群体平均育种值的代际变化
p	基因频率
cov	协方差
var	方差
E	数学期望
HRA	近似汉密尔顿规则 (Hamilton's rule—approximate)
HRG	一般汉密尔顿规则 (Hamilton's rule—general)
HRS	特殊汉密尔顿规则 (Hamilton's rule—special)
IBD	血缘同源 (identical-by-descent)
IBS	状态同源 (identical-by-state)

第1章 引言

1.1 选题背景

达尔文(Charles Darwin, 1859)的进化论通过自然选择解释了有机体为什么看似是设计好的。一个简单的说明是,自然选择会偏好那些具有更大适合度且可遗传的性状,由此这些性状能够在进化中保留下来。但是,自然界中普遍存在的利他主义在采用上述解释策略时会遇到困难,即利他行为乍看起来会损害行为者的利益而降低其适合度,使得利他主义在进化中被淘汰。汉密尔顿(William D. Hamilton, 1936–2000)通过引入广义适合度(inclusive fitness)解决了这一困难,即指出由于利他行为的接受者和行动者可以具有亲缘关系,我们可以将接受者从利他行为获取的利益部分地算作行为者的利益,进而利他行为可能使行为者具有更大的广义适合度(而非个人适合度(personal fitness)),从而在进化中占优势(Hamilton, 1964)。广义适合度理论(inclusive fitness theory)提出后得到了广泛的研究和应用,并产出了大量研究成果。

然而,广义适合度理论的价值在近二十年来一直饱受争议,最有代表性的是当时哈佛大学的三位生物学家诺瓦克(Martin A. Nowak)、塔尔尼策(Corina E. Tarnita)和威尔逊(Edward O. Wilson, 1929–2021)2010年发表在《自然》(*Nature*)杂志上的文章,该文对广义适合度理论进行了全面批评,并提出传统的自然选择进路^①是对广义适合度理论的更好替代(Nowak, Tarnita et al., 2010)。该文吸引了一组为广义适合度理论辩护的回应文章在《自然》杂志上发表,其中一篇文章有多达137人联名(Abbot et al., 2011),可见这场论战影响之深远。当时在牛津大学的加德纳(Andy Gardner)、韦斯特(Stuart A. West)和西安大略大学的怀尔德(Geoff Wild)合作完成了另一篇有代表性的为广义适合度理论辩护的论文(Gardner, West & Wild, 2011)。论战双方都是国际知名学者,这也从侧面反映了双方分歧之深刻。正反双方后来也持续发表论文进一步从各个角度辩护或批评(也包括澄清)广义适合度理论(Allen et al., 2013; Liao et al., 2015; Nowak & Allen, 2015; Allen & Nowak, 2016; Marshall, 2016; Queller, 2016; Nowak, McAvoy et al., 2017)。我将这场论战称

^① 诺瓦克等人的这一进路有时被称为自然选择进路、有时被称为数学进路、有时被称为种群遗传学进路、有时被称为进化动力学进路,不一而足,但所指的内容是大致相当的。我不试图在此给出统一的称呼。当然,为了明确所指,我将本文中所说的自然选择进路(以及数学进路、种群遗传学进路、进化动力学进路等)限定在诺瓦克等人在2010的论文中明确提出(Nowak, Tarnita et al., 2010)并在相关工作中推进的这一进路,而不多考虑该进路是否可能(尤其是在诺瓦克及其合作者以外的其他学者那里)有更丰富的内容。

为广义适合度理论之争^①。哲学家为澄清论战双方的概念使用和理论观点做出了重要贡献，进而也对调和论战双方的分歧做出了一定的努力（Birch, 2014b; Birch & Okasha, 2015; Birch, 2017a; 2017b; Rubin, 2018），但这场论战似乎还没有终结的趋势。本研究将立足于以上背景，对广义适合度理论之争展开哲学研究。下面我将简单介绍广义适合度理论之争所产生的问题以及现有的成果，并从中引出本文将要处理的三个相对独立的具体研究问题。

伦敦政治经济学院哲学、逻辑学与科学方法系教授伯奇（Jonathan Birch）和布里斯托大学哲学系教授奥卡沙（Samir Okasha）（Birch & Okasha, 2015）曾针对广义适合度理论之争提出三个开放问题：（1）如何对一般汉密尔顿规则^②中的回归系数进行因果解读？（2）进化将使个体尽力最大化广义适合度的观念在多大的范围内可用？（3）亲缘选择（kin selection）和多层次选择（multilevel selection）在什么条件下因果等价而不仅仅形式等价？东英吉利大学生物科学学院教授伯克（A. F. G. Bourke, 2019）一篇较近的论文将论战的主要成果概括为以下四点：（1）广义适合度理论和多层次选择理论（作为一种竞争理论）形式上是等价的，进而在许多应用中也是可互换的；（2）广义适合度理论和进化动力学（作为一种竞争理论）的关系大致是进化动力学在数学上更精确但更难应用于经验研究；（3）由普莱斯方程（Price equation）导出的一般汉密尔顿规则能为社会进化提供组织和整合框架；（4）近似汉密尔顿规则能够突破特殊汉密尔顿规则要求的严苛限制，使得广义适合度理论成为非常有用的建模工具。

本文处理的第一个问题是汉密尔顿规则，特别是一般汉密尔顿规则的理论地位。该问题定位于伯克总结的成果（3）。伯克所说的成果（3）的主要依据实际就是伯奇（Birch, 2017b）的观点，但有证据表明以伯奇为代表的同情一般汉密尔顿规则的价值学者忽视了一些重要的方面，特别是复杂群体结构带来的影响（Tartina & Taylor, 2014; van Veelen et al., 2017）。不过，加德纳等人（Gardner, West & Wild, 2011）的论文也提供了丰富的支持汉密尔顿规则的资源供参考。伯奇和奥卡沙提出的问题（1）也可看作回答该问题的一种努力方向，值得注意的是奥卡沙后来自己提供的解答也仅仅考虑了极度简化的情形（Okasha & Martens, 2016）。为

① 有些学者将这场论战称为“广义适合度之争（the inclusive fitness controversy）”（see Birch, 2017a）。我之所以增加“理论”一词称“广义适合度理论之争（the inclusive fitness theory controversy）”，是希望强调经受争议的是包含汉密尔顿规则、亲缘系数与亲缘选择、广义适合度及其最优化等诸多方面的广义适合度理论整体，而不仅仅是广义适合度概念本身。

② 汉密尔顿规则（Hamilton's rule）可分为不同版本，包括一般汉密尔顿规则（Hamilton's rule-general）、特殊汉密尔顿规则（Hamilton's rule-special）和近似汉密尔顿规则（Hamilton's rule-approximate）。参见第2.2节的相关讨论。

此，有必要通过对汉密尔顿规则更细致的分析，重新审视一般汉密尔顿规则的理论地位。

本文处理的第二个问题是广义适合度理论及相关的亲缘选择理论与真社会性（eusociality）的进化这一经验案例之间的解释或确证关系。该问题与伯奇和奥卡沙提出的问题（1）、问题（3）以及伯克总结的成果（2）都有一定关系。在广义适合度理论之争中，争议双方更多关注的是广义适合度理论自身的问题，对广义适合度理论和真社会性的进化这一经验案例之间的关系问题关注较少（see Liao et al., 2015; Nowak & Allen, 2015），哲学家在讨论中也往往回避了这方面的争议（e.g. Birch, 2017a; Rubin, 2018）。这一问题上的一个核心争议点是亲缘系数在真社会性进化中的因果角色，支持者认为亲缘系数是真社会性进化的重要原因，批评者则否认这一点。我将从这一核心争议点出发，试图通过亲缘系数和因果的相关概念澄清，评析争议双方就亲缘选择理论与真社会性的进化之间的解释或确证关系的不同观点。

本文处理的第三个问题是广义适合度最大化、最优化及其与能动者思维（agential thinking）的关系。该问题定位于伯奇和奥卡沙提出的问题（2）。适合度最大化一直是生物学家和哲学家关心的话题，此话题的一个主要代表就是格拉芬（Alan Grafen）的“形式达尔文主义项目（formal Darwinism project）”（Grafen, 1999; 2014），其中也包含下对广义适合度最优化的分析（Grafen, 2006b）。不过，格拉芬的模型设置了行为者控制支付（payoff）、支付可加（additive）、简单群体结构等严苛的条件，而奥卡沙后来自己提供的对问题（2）的解答将最大化推广到了不可加支付的情形，尽管被最大化的量不再是广义适合度（Okasha & Martens, 2016）。该话题近年的一个重要成果是奥卡沙（Okasha, 2018）的专著《进化中的主体和目标》（*Agents and Goals in Evolution*），其中对能动者思维的不同类型提供了较为细致的刻画，也讨论了能动者思维与适合度最大化的一般关联。我将从奥卡沙的工作出发，对广义适合度最大化和能动者思维之间的关系进行评估。

在更宽泛的意义上，对第一个问题的研究试图回答广义适合度理论自身是否是完善的科学理论；对第二个问题的研究试图回答广义适合度理论是否能得到经验案例的支持；对第三个问题的研究试图回答广义适合度理论有何哲学意义。这样看来，对这三个问题的研究合起来将能构成对广义适合度理论之争的整体性研究。

1.2 文献综述

1.2.1 广义适合度理论之争：背景、批评与辩护

对广义适合度的经典论述来自汉密尔顿，他对广义适合度给出了如下阐释：

广义适合度可以被想像为个人适合度在首先被去除所有可以被认为由于某个体社会环境造成的组分而保留其将在不暴露于该环境中的任何伤害或好处时表现的适合度，再加上其引起的其邻居适合度的伤害和好处的特定比例。这里的比例就是该个体与其影响的邻居间合适的亲缘系数。(Hamilton, 1964)⁸

基于广义适合度的概念，我们可以将有机体的社会环境纳入其适合度的考量，由此试图保留广义适合度在进化过程中的最大化。但以上定义的一个问题在于，我们在面对具体的案例时很可能难以区分有机体的适合度中哪些组分是其自身直接造成的，哪些组分又是由社会行为造成的。后人也常常使用另一种广义适合度的定义方式，即不考虑有机体适合度的组分拆分问题，而直接将某个有机体自己的适合度加上其邻居适合度的特定比例的和作为该有机体的广义适合度。这一定义在建模计算广义适合度时更为方便，但在概念层面则会带来困扰，如格拉芬（Grafen, 1982）就指出这一定义将导致不恰当的重复计算。

根据奥卡沙（Okasha, 2016）的解读，汉密尔顿的广义适合度理论可分为两部分。一部分是汉密尔顿规则作为判定一种社会行为能否在进化中占优势的条件。假定一种社会行为对行动者造成的损失是 c ，而对接受者造成的好处是 b ，两者的亲缘系数是 r ，那么汉密尔顿规则就是 $rb - c > 0$ ，即满足该条件的社会行为在进化中占优势，其中“ $rb - c$ ”即可大致看作该社会行为对行动者广义适合度的贡献。另一部分是将广义适合度（而非适合度）作为有机体行为的目标，由此达成“有机体作为最大化主体”的比喻。过往的研究更关注的是汉密尔顿规则这一部分，而对广义适合度最大化问题关注相对较少。

广义适合度理论作为一种研究社会行为进化的理论基础受到了一些学者的质疑。最有代表性的是诺瓦克等人 2010 年发表在《自然》杂志上的论文（Nowak, Tarnita et al., 2010）。该论文包含长达 40 页的补充材料，详尽论述了对广义适合度理论的批评意见，主张传统的自然选择进路完全优于基于广义适合度理论的亲缘选择进路。《自然》杂志之后还开辟了讨论该论文的栏目，刊登了 5 篇为广义适合

度理论辩护的评论，其中一篇论文有多达 137 人联名 (Abbot et al., 2011)，诺瓦克等人也对这些评论进行了回应。

诺瓦克等人 (Nowak, Tarnita et al., 2010) 对广义适合度理论的批评主要有如下几点：(1) 广义适合度进路只是较传统的自然选择进路而言另一种计数方法；(2) 广义适合度进路不如基于自然选择的博弈论进路具有一般性，而要求弱选择、支付可加性和简单群体结构等限制性条件；(3) 广义适合度经常被错误地定义；(4) 汉密尔顿规则几乎从不成立，即使限制性条件成立；(5) 在缺乏群体动力学模型时亲缘系数的测定是无定论的；(6) 广义适合度理论一般而言无法判定某种利他行为能否在进化中占优势。

针对以上批评，广义适合度理论的支持者 (Abbot et al., 2011) 给出了如下回应：(1) 广义适合度理论和自然选择理论是一致的，自然选择理论解释了生物表面上的设计，而广义适合度理论进一步解释了这种设计的用处；(2) 广义适合度理论成立一般而言并不需要限制性的条件，只不过在处理具体案例时研究者会采用限制性假设；(3) 广义适合度理论在 40 年来的行为和进化生态学研究产生了大量成果，而诺瓦克等人的进路却没能提供新奇的预测和对生物现象的解释。

然而，广义适合度理论的支持者的回应未能说服诺瓦克等人放弃他们的观点——论战仍在继续，以诺瓦克为代表的广义适合度理论的批评者持续发表论文批评广义适合度理论。阿伦 (Benjamin Allen) 等人 (Allen et al., 2013) 进一步讨论了广义适合度理论的限制。塔尔尼策和泰勒 (Peter D. Taylor) (Tarnita & Taylor, 2014) 指出在有穷种群且具有复杂互动结构时，广义适合度理论即使适用也将错误地给出进化方向的预测。诺瓦克和阿伦 (Nowak & Allen, 2015) 指出广义适合度理论关涉的现象与真社会性的进化无关。阿伦和诺瓦克 (Allen & Nowak, 2016) 指出不存在个体层面的广义适合度。诺瓦克等人 (Nowak, McAvoy et al., 2017) 指出汉密尔顿规则的一般形式不能用于预测也不能被经验检验。

针对广义适合度理论受到的批评，加德纳等人 (Gardner, West & Wild, 2011) 给出了一般性的辩护。他们列举了批评者认为汉密尔顿规则面临的 12 种困难——基因相互作用、强选择、频率依赖选择、纯策略、非亲属间的利他主义、条件策略、绿胡子、进化二叉分支 (evolutionary bifurcation)、间接基因效应、多代效应、群体选择和多于两个相互作用个体——并逐一列举文献反驳。其中一篇重要文献是加德纳等人 2007 年发表的论文，该论文证明了多基因座选择与亲缘选择之间数学模型上的对应关系，并论证了汉密尔顿规则的一般性和精确性 (Gardner, West & Barton, 2007)。加德纳等人 (Gardner, West & Wild, 2011) 认为批评者认为广义适合度理论面临困难主要是因为误解了汉密尔顿规则中的参数 c 、 b 和亲缘系数 r 。另

外，加德纳等人强调将一般性的亲缘选择理论和具体应用中可能包含简化的亲缘选择方法区分开来。最近，罗德里格斯（António M. M. Rodrigues）和加德纳（Rodrigues & Gardner, 2023）更进一步声称广义适合度理论构成了对“现代综合（the Modern Synthesis）”的科学革命。

另一位值得关注的为广义适合度理论提供辩护的生物学家是奎莱尔（David C. Queller）。奎莱尔（Queller, 1992）早在1992年就利用线性回归模型提出了一般的汉密尔顿规则，其中 c 、 b 和 r 都是某个线性回归系数，因而其生物学意义不算清楚。在广义适合度理论出现争议后，奎莱尔也有一篇一般性的回应，他认为广义适合度理论的批评者往往依赖于广义适合度理论历史发展的偶然特质，而没有认真对待真正的逻辑和证据（Queller, 2016）。

在批评广义适合度理论的同时，诺瓦克等人（Nowak, Tarnita et al., 2010）也提出了自然选择进路作为替代性的理论，并通过自然选择进路的数学建模为真社会性的起源提供了进化解释。廖（Xiaoyun Liao）^①等人（Liao et al., 2015）将诺瓦克等人的自然选择进路在经验断言上与广义适合度进路的差异概括为三点：（1）诺瓦克等人认为亲缘系数不是真社会性进化的本质要素；（2）广义适合度理论认为合作是在群体内不同成员直接的利益冲突基础上出现的，而诺瓦克等人认为工蜂不是独立的主体；（3）诺瓦克等人认为真社会性比人们通常设想的更难进化出现。廖等人逐一反驳了诺瓦克等人的以上观点，并指出诺瓦克等人的错误根源在于过度一般化了通过具体建模得到的结果。针对廖等人以上批评，诺瓦克和阿伦（Nowak & Allen, 2015）专门撰文反驳，指出了廖等人许多具体的误解和错误，认为自己对广义适合度理论的批评以及提供的新模型远比广义适合度理论支持者的工作深刻。

近年，凯（Tomas Kay）等人（Kay et al., 2020）考察了200篇研究利他行为进化且不将其归因于亲缘选择或否认亲缘系数地位的论文，从这些论文中概括出43种模型，主张这些新模型只是以不同的方式展示了亲缘选择和亲缘系数在利他行为进化中的重要地位。

1.2.2 广义适合度最大化和最优化

汉密尔顿（Hamilton, 1964）提出广义适合度时就明确提出了广义适合度最大化的思想。奥卡沙（Okasha, 2018）指出，广义适合度最大化的问题在后来社会进化的研究者那里往往没有受到足够的重视。伯奇和奥卡沙（Birch & Okasha, 2015）对广义适合度理论之争进行评述时，将有机体是否会在进化中尽力最大化其广义适合度作为该争论中一个未解决的开放问题。

^① 该中文姓为音译。

在适合度最大化问题上，格拉芬和他的“形式达尔文主义项目”提供了重要的理论工作（见表 1.1），广义适合度最优化是该项目的重要组成部分。其中一篇论文《广义适合度最优化》（“Optimization of inclusive fitness”）（Grafen, 2006b）直接处理广义适合度最优化问题，在收益由行为主体控制、支付可加和简单群体结构的限制条件下从数学上严格证明了社会行为进化模型与广义适合度最优化程序之间的形式链接（formal link）。

表 1.1 格拉芬“形式达尔文主义项目”早期成果概要（see Grafen, 2007c）

论文	成果概要
Grafen, 1999	提出了“形式达尔文主义”项目及其目标，讨论了两头下注（bet-hedging）。
Grafen, 2000	展示了多代的进化可以由普莱斯方程得到有效的解读。
Grafen, 2002	首次在一般性的条件下证明了群体遗传学模型和最优化程序之间的形式链接，但未考虑社会行为和群体结构。
Grafen, 2006a	讨论了一定条件下群体存在层级时的繁殖值，但尚未和最优化建立链接。
Grafen, 2006b	在一定条件下证明了社会行为的群体进化模型和广义适合度最优化程序之间的形式链接。
Grafen, 2007a	讨论了群体具有循环网络结构时的广义适合度。
Grafen, 2007b	应用广义适合度理论讨论了亲缘选择。
Grafen, 2008	为适合度最优化提供了最简单的形式化论证。

《生物学与哲学》（*Biology and Philosophy*）期刊 2014 年第 2 期为形式达尔文主义项目开辟了特殊议题专栏，刊登了编辑导言（Okasha & Paternotte, 2014a）、格拉芬对形式达尔文主义项目的纲领性介绍（Grafen, 2014a）、多达 11 篇其他学者（包括生物学家和哲学家）的评论及讨论（Birch, 2014c; Lehmann & Rousset, 2014; Ewens, 2014; Gardner, 2014; Shelton & Michod, 2014; Okasha & Paternotte, 2014b; Haig, 2014; Bourke, 2014; Sarkar, 2014; Orzack, 2014; Huneman, 2014）（这些学者对形式达尔文主义项目的态度见表 1.2）以及格拉芬对评论的回应（Grafen, 2014b）。根据格拉芬（Grafen, 2014a）的介绍，他所说的最优化程序包含三个元素：一个由最优化主体控制的变量，即工具（instrument）；该变量的取值范围，即限制集（constraint set）或可行集（feasible set）；以及一个表明每种变量取值对应的成功程度的函数，即最大量（maximand）。进而，在可行集范围内对工具值的选取如实现了最大量的最大化，则被称为解决（solve）了相应的最优化程序。在生物进化的

语境下，工具就是有机体的表型，可行集由群体遗传学情况决定，而最优化程序中的关键问题是建构一种恰当的最大量作为有机体试图最优化的目标，以寻求最优化程序的解决和演化结果的对应。格拉芬将他所希望达到的演化和最优化之间的形式链接应满足的条件表述为以下四点：

1. 如果一个群体中所有个体都解决了最优化程序，那么每种基因频率变化的期望都等于零，且没有任何可能的由稀有、显性的突变产生的表型能入侵该群体。
2. 如果所有个体达到了最大量的相同值但并未解决最优化程序，那么每种基因频率变化的期望都等于零，但存在可能的由稀有、显性的突变产生的表型能入侵该群体。
3. 如果个体达到的最大量的值不同，那么每种基因频率的变化量都等于其与个体所达到的最大量的值的协方差。
4. 如果每种基因频率变化的期望都等于零，且没有任何可能的由稀有、显性的突变产生的表型能入侵该群体，那么该群体中所有个体都解决了最优化程序。（Grafen, 2014a）¹⁶⁴

表 1.2 11 篇评论对形式达尔文主义项目的态度（see Okasha & Paternotte, 2014a）

态度	高度批评	部分批评	大致同情	强烈支持
论文作者	Ewens Orzack Shelton & Michod	Lehmann & Rousset Sarkar	Bourke Haig Okasha & Paternotte Birch	Gardner Huneman

格拉芬指出，基于以上条件的限制，符合要求的最大量就是生物学家想要定义的“适合度”概念。奥卡沙和帕泰诺特（Cédric Paternotte）（Okasha & Paternotte, 2014b）将其称为适合度的公理化理论。

格拉芬在有机体没有社会行为时为适合度最优化构建的形式链接（Grafen, 2002）相对容易被接受，莱曼和鲁塞甚至认为这一链接过于平凡而缺乏重要意义（Lehmann & Rousset, 2014）。对此主要的担心来自于频率依赖选择（Sarkar, 2014），频率依赖选择中适合度完全可能不最大化，一些学者也由此直接主张不可能找到某种在进化中趋于最大化的量（Rice, 2004）。至于格拉芬对广义适合度最优化的处

理 (Grafen, 2006b) 则更为可疑——他所建立的形式链接依赖于可加的支付, 而在支付不可加的情况下无法找到符合格拉芬要求的形式链接 (Lehmann & Rousset, 2014)。一种更一般的批评是, 格拉芬提供的形式链接在进化上不具有动力充分性, 其中不可侵入性 (non-invadability) 等概念较我们通常的理解有偏差 (Lehmann & Rousset, 2014); 即使上述四个条件由于前件为假而成立, 也将损害形式链接的生物学重要性 (Sarkar, 2014)。此外, 许多学者在评论中都提及最优化的主体和目标问题, 最有代表性的就是埃文斯 (Ewens, 2014) 明确拒绝格拉芬的有机体视角, 认为其包含着内在的混乱。

值得注意的是, 莱曼和鲁塞将格拉芬刻画的最优化看作是“严格的”最优化, 认为其最大量是一元函数, 因而排除了“最佳反应 (best-response)”概念——后者是借用自博弈论的概念 (Mas-Colell et al., 1995)²⁴²⁻²⁴³, 允许最优化主体将包括他人策略等在内的多种参数纳入决策的考量, 这也是在他们看来进化博弈论中实际使用的概念 (Lehmann & Rousset, 2014)。格拉芬似乎也认为行为生态学中对于广义适合度最优化采用的是最佳反应概念 (Grafen, 2014b)。最佳反应意义上的最优化值得我们作为一种可能的对严格最优化的替代性概念认真对待。

奥卡沙和马滕斯 (Okasha & Martens, 2016) 在一个简单的模型下讨论了支付不可加时的社会行为进化, 清晰地展示了此时有机体不会最大化广义适合度, 但找到了一个其它量在进化中趋于最大化, 尽管这个量不是任何通常定义下的广义适合度, 由此他们认为“有机体作为最大化主体”的比喻仍有可能得到挽留。莱曼和鲁塞 (Lehmann & Rousset, 2020) 基于他们自己推崇的最佳反应意义上的最优化概念详细分析了有机体最大化广义适合度的条件。

伯奇 (Birch, 2016) 区分了四种适合度最大化并对它们做出了系统的哲学分析, 澄清了费希尔的自然选择的基本定理和格拉芬的形式达尔文主义项目所指向的适合度最大化的不同含义, 但主张两者与适合度最大化的关系都非常可疑而缺乏重要生物学意义, 同时指出相较于寻找普遍的最优化原则, 不如致力于寻找自然选择导向适合度最大化的具体生物学条件。不过, 在分析广义适合度理论之争时, 伯奇 (Birch, 2019) 指出, 广义适合度作为表现型改善和最优化的标准很有价值, 尽管这完全不预设最优化在真实案例中普遍存在。

莱文 (Samuel R. Levin) 和格拉芬近年合作的论文对广义适合度理论做了推进。在 2019 年的论文中, 他们采纳了伯奇对广义适合度理论之争的分析, 承认广义适合度的最大化在足够一般的模型中无法达到, 但希望说明为何广义适合度理论对经验研究而言仍然是有帮助的框架 (Levin & Grafen, 2019)。在 2021 年的论文中,

他们试图将广义适合度的最大化推广到非可加支付的情形，其中的关键在于对广义适合度的恰当定义（Levin & Grafen, 2021）。

另外，奥卡沙（Okasha, 2018）在其专著中详细讨论了进化生物学中的能动者思维，这对我们理解“有机体作为最大化主体”的比喻很有帮助。奥卡沙将能动者思维分为两类，第一类能动者思维是将有机体看作试图实现某个特定目标的主体，第二类能动者思维是将大自然看作将有机体在进化过程中导向最优化的主体。奥卡沙肯定了第一类能动者思维的价值，并且认为其成立与否是一个经验问题。

1.2.3 中文文献状况

在中国知网上搜索主题为“广义适合度”的文献有 18 条结果，和生物学哲学有关的大致有 9 篇^①，下面简要介绍这 9 篇文献。刘奇（2006）的硕士论文详细讨论了亲缘选择和利他行为的进化机制，但论文主要是综述。刘奇的相关成果也与他人合作发表为期刊论文（刘鹤玲 等, 2007）。肖凤良和陈晓平（2012）分析了亲缘选择理论面临的问题，主张其是更一般的群体选择理论的补充。赵斌（2012）在讨论模态的功能理论时涉及到了广义适合度概念，但并未讨论广义适合度理论本身。罗力群（2016）介绍了汉密尔顿及其广义适合度理论，关注到了广义适合度理论之争，但并未实质性地介入问题讨论。罗力群（2018）关于进化生物学家特里弗斯（Robert Trivers）的人物研究平实地介绍了特里弗斯的理论观点，但没有批判性的分析。于书（2019）的硕士论文简要讨论了威尔逊对广义适合度理论的激烈批判，其中也涉及对格拉芬的批评，相关部分也以期刊论文的形式发表（于书, 2020）。但于书似乎并未正面给出自己对威尔逊相关理论观点的评估，也没能认识到威尔逊对广义适合度理论的批判并没有被主流学界所接受。罗力群（2019; 2020）后来的两篇论文主要讨论社会生物学的相关问题，其中涉及了广义适合度概念，但没有正面涉及对广义适合度理论本身的讨论。

在中国知网上搜索主题为“广义适应度”的文献有 359 条结果，和生物学哲学有关的大致有 2 篇^②。王健（2009）的论文讨论了亲缘选择理论在经济学中的应用；徐亮（2022）的博士论文从生物个体性的视角论及了广义适合度最大化问题。

在中国知网上搜索主题为“亲缘选择”的文献有 162 条结果，其中生物学学科下有 59 篇，哲学学科下有 9 篇^③，下面简要介绍其中相对重要且与我的研究主题更相关的几篇文献。刘鹤玲（2008）在汉密尔顿亲缘选择学说的基础上解释了膜翅

① 检索时间为 2024 年 6 月 19 日。

② 检索时间为 2024 年 6 月 19 日。

③ 检索时间为 2024 年 6 月 19 日。部分检索结果与之前的检索结果有重复，计数时未删去重复结果，但不再重复介绍相关文献内容。

目昆虫亲缘利他行为进化的基因机制。李建会和项晓乐（2009）和李建会（2009）两篇文章内容相近，简单回顾了进化理论从个体选择、群体选择、基因选择（亲缘选择理论和互惠利他理论）到多元选择的发展历程，认为埃利奥特·索伯（Elliot Sober）主张的自然选择单位的多元主义已经解决“利他难题”。邹笑笑（2012）详细分析了群体选择理论的兴衰史，认为群体选择理论曾被亲缘选择理论和互惠利他理论取代，但现已复兴并获得了全面的胜利——邹笑笑的这一判断与当今学界主流意见有较大差距。肖凤良和陈晓平（2013）批评了道金斯的基因选择观点，他们一个有趣的观点是主张基因选择理论是亲缘选择理论的进一步发展。吴林林等人（2013）讨论了亲缘选择理论在解释非人灵长类动物社会行为时的作用和局限。金萍和颜忠诚（2013）简单介绍了亲缘选择理论及其局限。李洁等人（2014）对植物亲缘选择研究的最新进展进行了综述。杨月等人（2015）讨论了微生物模型的合作行为进化，并将亲缘选择理论作为一种解释理论。赵斌和郭虹（2019）在讨论利他行为的进化时，将亲缘选择理论作为早期理论而抛弃，认为索伯的群体选择解决了利他问题，并讨论了在此基础上的模型的多元发展。杨阳（2020）的硕士论文讨论了利他行为的进化基础及其对人类道德的意义，对亲缘选择、群体选择和多层次选择都给出了相对中立（而非像前述部分学者那样厚此薄彼）的评析。贾栗（2013a; 2013b）的两篇论文简要介绍了亲缘选择理论、互惠理论和群体选择理论等对利他行为的解释。

在中国知网上搜索主题为“亲属选择”的文献有 28 条结果，和生物学哲学有关的大致有 2 篇^①，即贾栗（2013a; 2013b）的两篇论文，之前已经介绍。

除了以上文献，刘华杰（2022）的论文关注到了广义适合度理论之争，并对批评者的立场给出了进一步的辩护。^②

通过以上检索和分析，可以看出国内对社会进化的相关研究特别是哲学研究的几个特点：（1）中文文献对社会进化的研究常常出于对利他行为进化的关切；（2）专门介绍亲缘选择或广义适合度理论的中文文献往往对广义适合度理论面临的批评意见关注较少；（3）有些中文文献认为亲缘选择理论已经过时，而群体选择或多层次选择是更优的替代理论——这一态度似乎略显武断，尤其是对支持亲缘选择理论的学者在回应相关批评、发展改进亲缘选择理论上的大量努力有所忽视；（4）仅有极个别中文文献（罗力群, 2016; 刘华杰, 2022）关注到了 2010 年以来的广义适合度理论之争。

① 检索时间为 2024 年 6 月 19 日。

② 这篇论文尽管全文中包含“广义适合度”和“亲缘选择”字眼，但无法在之前限定主题的搜索中搜到。

1.3 研究方法

本文作为一篇科史哲论文，主要采取概念分析的方法展开工作。当然，在处理具体生物学理论和案例时，也将涉及简单的回归分析、博弈论等数学建模方法。

1.4 论文结构安排

在第2章中，我将仔细考察一般汉密尔顿规则和特殊汉密尔顿规则之间的区别及它们解释价值的可疑之处，论证伯奇对一般汉密尔顿规则的辩护依赖于一般汉密尔顿规则中的相关参数具有明确的生物学意义，而这一点并未得到充分辩护。通过对具体案例的分析，我将展示一般汉密尔顿规则作为组织框架面临的两点疑难，并提出一种新的方案来理解多种一般形式的汉密尔顿规则的解释价值。

在第3章中，我将仔细考察真社会性进化的案例，评估广义适合度理论和真社会性的进化这一经验案例之间是否具有可靠的解释或确证关系。我将从关于亲缘系数在真社会性进化中的因果角色的争议出发，通过对三种不同意义的亲缘系数概念的区分以及两组因果概念的区分，尽量清晰地展现争议双方在这一问题上所具有的不同观点，并初步展现这一问题下可能被忽视的困难之处。

在第4章中，我将考察奥卡沙就广义适合度最大化和能动者思维的相关论述，通过对奥卡沙概念体系的细致分析，重新评估广义适合度最大化与能动者思维之间的关系。

最后，我将在第5章总结以上成果，对广义适合度理论之争给出整体性的评估，反思本文工作的局限性并展望未来可能的研究方向。

第2章 广义适合度理论中汉密尔顿规则的解释价值

2.1 引言

广义适合度理论是当代社会进化^①研究的重要理论基础，但近年来一直饱受争议。2010年，当时哈佛大学的三位生物学家诺瓦克、塔尔尼策和威尔逊在《自然》杂志上发表文章全面批评广义适合度理论（Nowak, Tarnita et al., 2010）。该文吸引了一组为广义适合度理论辩护的回应文章在《自然》杂志上发表，其中一篇文章有多达137人联名（Abbot et al., 2011），可见这场论战影响之深远。正反双方后来也持续发表论文进一步从各个角度辩护或批评广义适合度理论（Gardner, West & Wild, 2011; Allen et al., 2013; Queller, 2016; Nowak, McAvoy et al., 2017）。国内也有一些学者注意到这场论战（罗力群, 2016; 刘华杰, 2022）。

广义适合度理论的一项核心内容就是汉密尔顿规则，这也是论战的一个焦点。汉密尔顿是英国博物学家、进化生物学家，也是广义适合度理论的提出者。以他名字命名的汉密尔顿规则提供了判定一种社会行为在进化中占优势的条件，由此回答了利他主义的进化这一难题（Hamilton, 1964）。乍看起来，施行利他行为会降低行为主体自身的适合度^②，使之在进化中处于不利地位。然而，自然界中却普遍存在着利他行为。那么，利他主义的进化是如何可能的呢？根据汉密尔顿规则，由于利他行为的接受者和行动者可以有一定的基因相似性，接受者受益也就部分相当于行动者自身受益，这远远胜过利他行为所需付出的代价，进而综合来看行动者的适合度可能其实是增大了。更一般地，假定一种社会行为对行动者造成的代价是 c ，而对接受者造成的好处是 b ^③，两者的亲缘系数（relatedness）^④是 r ，那么汉密尔顿规则就是 $rb - c > 0$ ，即满足该条件的社会行为在进化中占优势。直观上看，对利他行为而言 c 、 b 和 r 都是正数，进而如果 b 足够大、 c 足够小、 r 足够大，那么 $rb - c > 0$ 就能成立，利他行为也就可以在进化中流行开来。

① 本文中社会进化（social evolution）指社会行为的进化（evolution of social behaviour），即进化生物学中研究社会行为在进化中如何产生并演变的分支，其应用并不限于纯粹生物学的场景，而可延伸到其它领域如社会科学中。

② 适合度（fitness）的概念直观来说大致是有机体对环境的适应程度，在实际生物学研究中通常简单地把成年子女的数量作为适合度大小的衡量标准。本文也遵从这一通常的理解，而不涉及更多有关适合度概念的争论。

③ c 和 b 均可正可负。如果一种社会行为对行动者自身有益，那么 c 就取负数；如果一种社会行为对接受者有害，那么 b 就取负数。

④ 亲缘系数可以大致理解为一个衡量行动者和接受者基因相似度的指标，其准确含义将在第2.2节给出。

对于简洁的汉密尔顿规则，论战双方有显著的分歧。诺瓦克等人一开始严厉批评汉密尔顿规则“几乎从不成立”（Nowak, Tarnita et al., 2010），而广义适合度理论的支持者却说汉密尔顿规则普遍成立（Gardner, West & Wild, 2011）。后来，经过一些学者的澄清，人们认识到汉密尔顿规则有不同的版本，其中特殊汉密尔顿规则更符合人们的直观理解，但确如批评者所言需要严格的限制条件才能成立。而另一种版本的汉密尔顿规则，即一般汉密尔顿规则，则可以用纯粹数学的方式严格证明，几乎不需要任何限制条件，如支持者所言是普遍成立的，而且它在支持者看来可以作为支持广义适合度理论的坚实基础（Birch, 2014b）。由此可见，论战双方对汉密尔顿规则的上述相反意见其实是基于不同版本的汉密尔顿规则给出的，这并不构成实质的分歧。

不过，广义适合度理论的批评者对一般汉密尔顿规则也有批评意见，认为其尽管普遍成立，但没有解释价值，特别是无法直接应用于具体的经验案例（Nowak, McAvoy et al., 2017）。对此，伯奇给出了具有代表性的回应，即一般汉密尔顿规则可以作为社会进化研究的组织框架（organizing framework），由此具有特殊的解释价值（Birch, 2017a; 2017b）。尽管上述观点受到了一些学者的认可（Bourke, 2019; Levin & Grafen, 2019; Ågren, 2021），但我认为其并不成立。本章将给出对伯奇辩护方案的反驳，并尝试在伯奇的基础上提出一种理解汉密尔顿规则解释价值的替代方案。

本章结构如下。第2.2节将介绍两种不同版本的汉密尔顿规则，随后第2.3节将介绍一般汉密尔顿规则面临的解释价值问题以及伯奇对一般汉密尔顿规则解释价值的辩护。在第2.4节，我将提出对伯奇辩护方案的两点反驳。在第2.5节，我将反思汉密尔顿规则的多种一般形式对其解释价值的影响，并尝试提出一种替代性的辩护方案。第2.6节是本章小结。

2.2 汉密尔顿规则的不同版本

伯奇（Birch, 2014b）首先区分了一般汉密尔顿规则（Hamilton's rule-general, 简称 HRG）和特殊汉密尔顿规则（Hamilton's rule-special, 简称 HRS），后来又与奥卡沙合作区分出了第三种版本，即近似汉密尔顿规则（Hamilton's rule-approximate, 简称 HRA）（Birch & Okasha, 2015）。HRA 在社会进化研究实践中是使用最多的版本，但反而不是论战的重点。下面我们重点介绍 HRS 和 HRG。

我们首先使用伯奇（Birch, 2014b）采用的简单模型给出 HRS 的推导示例。考虑一个由两类个体构成的种群，其中一类基因型为“A”，另一类基因型为“S”。该种群中的个体会两两结对进行社会行为，其中 A 类个体将付出代价 C 而使搭档

获益 B ， S 类个体则不牺牲自己而只接受搭档可能赋予的利益。由此产生的支付矩阵如表1所示，其中的支付（payoff）可看作对行为主体适合度的影响。其中， A 和 A 互动产生的支付 $(B - C)$ 刚好是行为主体的牺牲 $(-C)$ 加上搭档行为赋予的收益 B ，这时我们说支付是可加的。

表 2.1 可加支付条件下的支付矩阵

支付矩阵		行为搭档类型	
		A	S
行为主体类型	A	$B - C$	$-C$
	S	B	0

在该模型中，亲缘系数 r 可理解为一个行为主体固定与同类型搭档发生社会行为的概率；而该主体另有 $(1 - r)$ 的概率随机选择搭档，即此时该主体的搭档是两种类型的概率分别等同于两种类型个体在种群中的频率^①。进而，两种基因型的期望支付分别是

$$\begin{aligned} W_A &= (r + (1 - r)p)(B - C) - (1 - r)(1 - p)C \\ &= (r + (1 - r)p)B - C, \end{aligned} \quad (2.1)$$

$$W_S = (1 - r)pB, \quad (2.2)$$

其中 p 指 A 类个体在种群中的频率。由此可得

$$W_A > W_S \text{ 当且仅当 } rB - C > 0. \quad (2.3)$$

这就是 A 类基因型及其相应社会行为在进化中占优势的条件，即HRS。

容易发现，HRS中的系数 B 和 C 有非常自然的因果解读。支付的可加性保证了，无论一个行为主体的搭档是什么类型，对该主体而言施行某社会行为所产生的收益 B 和代价 C 都不会改变，进而我们可以将 B 和 C 自然地看作该社会行为的因果贡献。由此，HRS可以直接用于因果解释，具有明确的解释价值。不过，HRS的成立依赖于支付的可加性、简单的种群结构等严格的限制条件，因而其应用范围极其受限。

HRG则指奎莱尔利用线性回归模型给出的汉密尔顿规则（Queller, 1992）。下面我们简要地给出推导过程。根据简化的普莱斯方程（Price equation），有

$$\Delta \bar{g} = \frac{1}{\bar{w}} \text{cov}(w, g). \quad (2.4)$$

① 在这一简单模型中，此种理解下的亲缘系数与后文推导HRG时通过回归方法定义的亲缘系数 $\beta_{g,g}$ 是一致的，我在此略去具体的推导。注意到， $r = 0$ 的直观意义是行为主体根据不同类型个体在种群中的频率随机选取搭档，而没有任何偏好。有时我们会将 r 取为负值，这意味着行为主体更倾向于和与自身类型相反的个体发生社会行为，即行为主体有 r 的概率选取相反类型的搭档而另有 $(1 - r)$ 的概率随机选择搭档。

其中 g_i 指父代中个体 i 某性状对应的育种值 (breeding value)^①, $\Delta\bar{g}$ 指种群平均育种值从父代到子代的变化量, w_i 指个体 i 的适合度, \bar{w} 指种群平均适合度, cov 表示协方差。该方程通过有机体适合度和育种值之间的协方差来描述育种值的代际变化量, 也就是进化的方向。

由于我们考虑的是社会行为的进化, 有机体的适合度与其自身育种值和其社会行为搭档的育种值都有关。我们建立有机体适合度 w 对其自身育种值 g 和其社会行为搭档育种值 g' 的线性回归模型:

$$w = \alpha + \beta_{wg.g'}g + \beta_{wg'.g}g' + \varepsilon. \quad (2.5)$$

注意到, 线性回归模型的存在并不要求 g 和 g' 这两个参数真的以严格线性组合的方式对有机体适合度 w 做出贡献——如果 g 和 g' 这两个参数的真实贡献不是线性的, 那无非是让该回归模型的误差项 ε 大一些, 但不会导致模型不成立。

我们将式 (2.5) 代入式 (2.4) 可得

$$\begin{aligned} \Delta\bar{g} &= \frac{1}{\bar{w}} (\beta_{wg.g'}\text{var}(g) + \beta_{wg'.g}\text{cov}(g', g)) \\ &= \frac{1}{\bar{w}} (\beta_{wg.g'} + \beta_{wg'.g}\beta_{g'g})\text{var}(g), \end{aligned} \quad (2.6)$$

其中 var 表示方差。由此, 我们有

$$\Delta\bar{g} > 0 \text{ 当且仅当 } \beta_{wg.g'} + \beta_{wg'.g}\beta_{g'g} > 0. \quad (2.7)$$

如果把 $\beta_{wg.g'}$ 看作该行为的代价 ($-c$), 把 $\beta_{wg'.g}$ 看作该行为的收益 b ,^② 把 $\beta_{g'g}$ 看作亲缘系数 r , 那么式 (7) 中所给出的该行为在进化中占优势的条件就有了 $rb - c > 0$ 的形式, 这也就是 HRG。

HRG 除了简化的普莱斯方程 (式 (2.4)) 所需假设外不预设任何限制条件, 是普遍成立的。然而, 由于 HRG 中的参数, 特别是 c 和 b , 都由统计意义上的回归系数给出, 其生物学意义并不清楚, 由此也就产生了关于 HRG 解释价值的一系列争议。

① 育种值即可加遗传值 (additive genetic value)。对于通常由许多不同位点上的基因共同控制的复杂社会行为而言, 我们可以方便地仅用育种值这一个可遗传的量, 来表达由所有这些位点上的基因效应线性组合得到的该社会行为的性状值。如果某行为或性状仅由一种基因的有或无决定, 育种值就退化为个体的基因频率, 即育种值为 1 代表某个体具有某基因及相应行为, 育种值为 0 代表某个体不具有某基因及相应行为。就本文内容所涉范围内, 不熟悉育种值的读者完全可以将其理解为更简单直观的个体基因频率。

② 我使用大小写字母区分 HRS 和 HRG 中的收益 (benefit) 项和代价 (cost) 项。大写字母 B 和 C 分别表示 HRS 中的收益和代价, 小写字母 b 和 c 分别表示 HRG 中的收益和代价。

2.3 一般汉密尔顿规则的解释价值

对 HRG 解释价值最清晰明确的质疑可见诺瓦克等人的论文《汉密尔顿规则的一般形式不能做出预测也不能被经验检验》(Nowak, McAvoy et al., 2017)。暂且不论汉密尔顿规则的版本,我们直观上会认为,将汉密尔顿规则应用于经验案例,也就是基于经验数据找出汉密尔顿规则中三个参数 b 、 c 和 r 的值,比较 $(rb - c)$ 和0的大小关系,进而得出相应社会行为在进化中是否占优势的结论。这一套操作程序对 HRS 而言是非常合理的,但很可能并不适合 HRG——诺瓦克等人所反对的主要就是基于上述程序应用 HRG 的价值。

对 HRG 而言,上述操作程序的问题在于,分别计算三个参数 b 、 c 和 r 的值反倒比直接计算 $(rb - c)$ 这个量的值要更复杂,甚至需要用到更多的经验数据。回看上一节对 HRG 的推导,三个参数 c 、 b 和 r 的值分别是 w 对于 g 、 w 对于 g' 和 g' 对于 g 的回归系数。想要算出这三个回归系数的值,假定种群大小是 n ,自然就需要知道 $w = (w_1, w_2, \dots, w_n)$ 、 $g = (g_1, g_2, \dots, g_n)$ 和 $g' = (g'_1, g'_2, \dots, g'_n)$ 这三组数据的值,其中 w_i 、 g_i 和 g'_i 分别表示种群中第 i 个个体的适合度、育种值和其行为搭档的育种值。然而,既然我们已经知道 $w = (w_1, w_2, \dots, w_n)$ 和 $g = (g_1, g_2, \dots, g_n)$ 这两组数据,那我们何不直接计算两者的协方差 $\text{cov}(w, g)$ 呢?根据简化的普莱斯方程(式(2.4)), $\text{cov}(w, g)$ 已经足以表明进化的方向。如果我们愿意,可以写出如下判定规则:

$$\Delta \bar{g} > 0 \text{ 当且仅当 } \text{cov}(w, g) > 0. \quad (2.8)$$

这一规则和 HRG 是完全等价的,但较 HRG 少要求一组数据 $g' = (g'_1, g'_2, \dots, g'_n)$,可见以判定进化方向为目标时逐一计算 HRG 中三个参数的值显然是绕了远路。因此,HRG 尽管在一般性和精确性上无可挑剔,但在经验应用上的价值却并不明朗。

既然以判定进化方向为目标时计算 b 、 c 和 r 这三个参数的值既麻烦又多余,那么我们计算出这些参数的值是否可能有其它价值呢?伯奇就刻画了这样一种辩护 HRG 解释价值的方案,即 HRG 作为社会进化研究的组织框架仍有解释价值。伯奇将组织框架刻画为能对众多具体模型进行分类的一个一般的模型。对 HRG 而言, rb (作为一个乘积)和 c 这两个参数的符号及大小关系就可以用于将各种具体的社会进化模型划分为不同的类型,因而说 HRG 能作为社会进化研究的组织框架。所谓作为组织框架而具有的解释价值,具体而言是指 HRG 作为组织框架所提

供的分类有助于我们比较和整合不同的具体社会进化模型，从而加深对社会进化的理解（Birch, 2017b）。^①

图1展示了伯奇提供的分类图示。对任何一种行为，我们都可以根据前述的回归模型算出其对应的参数值。假定该行为在进化中频率上升，那么（1）当 $rb > c \geq 0$ 时，该行为的进化需要间接适合度解释，因为这类行为单看对行动者自身的影响是不利的，但若将对接受者的影响纳入考量则会发现其进化优势所在；（2）当 $c < rb \leq 0$ 时，该行为的进化需要直接适合度解释，因为这类行为单看对行动者自身的影响就非常有利，而对接受者的影响反而是不利的；（3）当 $rb > 0$ 且 $c < 0$ 时，该行为的进化需要混合解释，即该行为既对自身有利，其对接受者的影响也有利；（4）当 $rb \leq c$ 时，该行为的进化需要（完全或部分的）非选择解释，因为单看汉密尔顿规则这类行为是不应当在进化中占优势的，肯定是某些特殊的、偶然的因素左右了进化结果。

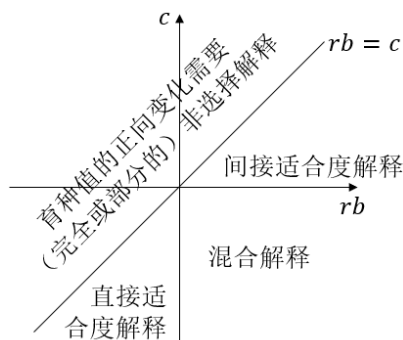


图 2.1 伯奇通过 HRG 对社会进化的不同解释的归类（Birch, 2017b）⁴⁹

下面，我们通过一些例子对以上分类功能给出一些进一步的说明。间接适合度解释要求 $rb > c \geq 0$ ，显然 r 、 b 和 c 均大于0的利他行为^②符合这一特征。其实，怨恨（spite）或损人不利己的行为也属于这一类。例如，我们常说“杀敌一千，自损八百”，不妨将其解读为一行为对接受者造成伤害 $b = -1000$ ，而行为者自身付出的代价 $c = 800$ 。假如行为者在辨别敌人上做得足够好，使得 $r = -0.9$ ，就有 $rb - c = 100 > 0$ 成立，这种行为就得以在进化中占优势。基于HRG的分类告诉我们，怨恨和利他主义的进化其实共享相同的解释类型（Birch, 2017b）。

① 这里介绍的是伯奇较晚时候的观点（Birch, 2017b）。伯奇2014年的论文就主张HRG能够作为社会进化研究的组织框架，但对组织框架具体涵义的解读与后来不同（Birch, 2014b）。本文中仅考察伯奇较晚的观点。

② 需要留意的是，HRG中的 r 、 b 和 c 都是回归系数，而我们通常说利他主义行为有 r 、 b 和 c 均大于0是在类似于HRS中可作因果解读的参数意义上理解的，两者未必一致。广义适合度理论的批评者在这一点上着墨很多，认为 b 和 c 这两个参数根本不能与我们通常理解的社会行为收益和代价等同（Allen et al., 2013; Nowak, McAvoy et al., 2017）。

直接适合度解释要求 $c < rb \leq 0$ 。这一类解释的目标对象包含一些常见的互惠合作。例如，一位猎人和他人合作猎鹿，尽管打猎时付出了一定代价，但打猎成功带来的收益是巨大的，远远胜过参与打猎的付出。此时我们应当认为合作猎鹿的行为对猎人自身而言就是有正收益的，也就是 $c < 0$ 。如果合作猎鹿的猎人间并无亲缘关系，即 $r = 0$ ，该情形就会被归入直接适合度解释。这一案例中的互惠合作可以不需要合作者具有亲缘关系，历史上常与亲属间的利他行为相混淆。而基于 HRG 的分类告诉我们， $c < 0$ 的互惠合作和 r 、 b 和 c 均大于 0 的利他行为其实属于不同的解释类型（Birch, 2017b）。

2.4 对伯奇的反驳

单纯实现分类并不困难——只要分类标准清晰明确，我们当然能够准确地开展分类工作。问题是，HRG 对具体社会进化模型的分类有用吗？为什么偏偏把社会进化模型分为前述四类呢？问题的关键在于基于 HRG 的分类的生物学意义。如伯奇所说：“只有当它（指 HRG）不能为我们对社会进化的理解增添价值时，它才会是非解释性的；而我主张，通过将对具体现象的因果解释区分为**有生物学意义的类别**，它确实有助于理解。”（Birch, 2017b）⁶⁹而且，伯奇实际上认为 HRG 在大部分语境下都是最优的组织框架（Birch, 2017b）⁸³。所以，基于 HRG 的分类的生物学意义还不能太弱，否则不足以有力地辩护 HRG（相较其它潜在候选项）作为组织框架的解释价值。

如上一节末尾所示，伯奇确实提供了一些例子来从直观上说明基于 HRG 的分类有何生物学意义。然而，在我看来，基于 HRG 的分类可能经不起更严格的推敲。接下来，我将通过具体案例展示基于 HRG 的分类的可疑之处。^①

2.4.1 分类的不稳定性

我们使用诺瓦克等人提供的一个案例来说明这一点（Nowak, McAvoy et al., 2017）。如图 2.2，（A）中给出了一个种群中的四种个体 1、2、3、4，其中深色代表具有某基因，白色代表不具有某基因。四种个体之间箭头上的数字表示四种行为搭档之间社会行为的权重，特别地， ε 代表一个微小的扰动。（B）中给出了（A）中四种个体的 $w = (w_1, w_2, w_3, w_4)$ 、 $g = (g_1, g_2, g_3, g_4)$ 和 $g' = (g'_1, g'_2, g'_3, g'_4)$ 这

① 一篇新近论文同意 HRG 能作为组织框架，也同意 HRG 有解释价值，但反对 HRG 有解释价值是由于能作为组织框架（Kolofotis & Verreault-Julien, 2022）。在我看来，这种关于科学解释的哲学讨论恰恰回避了问题的关键，即 HRG 中的参数有何生物学意义。我也不认为上述论文能够有效应对我接下来要提出的两点不利于 HRG 解释价值的意见。

三组数据。当 $w_1 = 1.6$, $w_2 = w_3 = w_4 = 0.6$, $x = 0.5$ 时, 我们容易计算出基因频率的变化

$$\Delta \bar{g} = \frac{1}{\bar{w}} \text{cov}(w, g) = \frac{1}{10 + \frac{10}{3}\varepsilon} \quad (2.9)$$

由于 ε 是一个微小的扰动, 我们可以取 $\varepsilon \rightarrow 0$ 时 $\Delta \bar{g}$ 的极限值 0.1, 认为 ε 非常接近 0 时育种值的变化量非常稳定地接近 0.1。

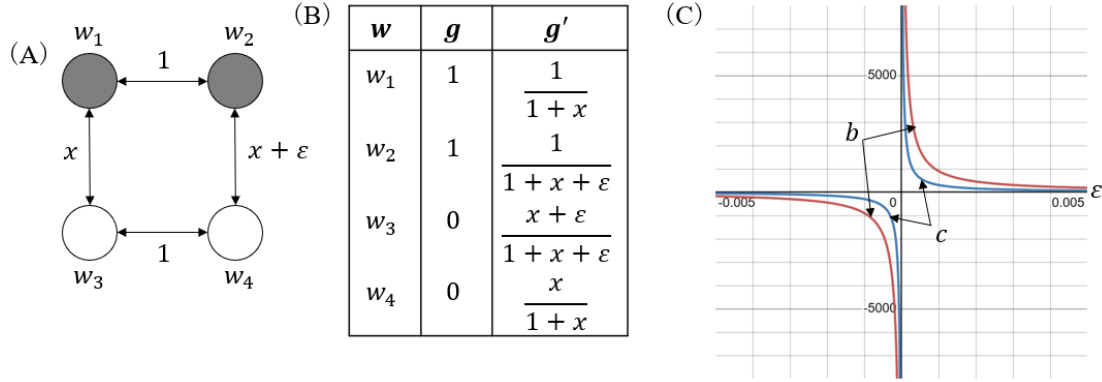


图 2.2 诺瓦克等人用以指出 HRG 中参数不稳定性的案例 (Nowak, McAvoy et al., 2017)

Fig. 4

然而, HRG 中的参数 b 和 c 却并不稳定。图 2.2 的 (C) 展示了 ε 非常接近 0 时参数 b 和 c 的值。可以看到, 当 ε 取非常接近 0 的负数时, b 和 c 都将是绝对值非常大的负数, 该模型会被伯奇归入直接适合度解释类型; 而当 ε 取非常接近 0 的正数时, b 和 c 都将是非常大的正数, 该模型会被伯奇归入间接适合度解释类型。

直觉上, 在 ε 非常接近 0 的前提下, ε 为正还是为负似乎没有什么特别重大的区别, 这一点也反映在育种值变化结果的稳定性上 (式 (2.9)); 换句话说, ε 并不仅仅是一个任意的微小扰动, 而是一个对进化结果几无影响的微小扰动。但是, 伯奇基于 HRG 的分类却将会这两种情形分为截然不同的类型! 注意到, 在伯奇的分类下, 直接适合度解释和间接适合度解释的区域除原点外甚至没有交界 (见图 1) ——而且考虑到 ε 非常接近 0 时 b 和 c 的绝对值非常大, 可以说 ε 的正负变化使得该模型以一种跳跃性的方式跨过了不同的解释类型。

基于 HRG 的分类的不稳定性自然会让我们怀疑其实际意义。打比方说, 中医认为感冒有风热和风寒两种类型。假若两位感冒患者症状差不多, 体质差不多, 吹的风也差不多, 但却分别被中医归入风热和风寒两种不同类型, 我们就会怀疑风热和风寒是否真的是有医学意义的区分。当然, 我并不预设这里基于 HRG 给出的分类一定是错误的, 但伯奇分类方案的支持者有责任说明为何一个会被微小的扰动

剧烈影响的参数可以作为标准用于有意义的分类，这在伯奇现有的论述中似乎是缺失的。

另一点补充说明是，参数的不稳定性问题主要影响的是间接适合度解释、混合解释和直接适合度解释之间的界限。至于非选择解释和其它解释类型（不妨统称为选择解释）之间的界限（即直线 $rb = c$ ）倒不会有问题，这是因为 $(rb - c)$ 这一量的符号本身就表示基于自然选择的进化的方向，有非常明确的生物学意义，其为正和为负确实有本质上的区别。不过，这一界限其实并不需要 HRG 才能给出，基于普莱斯方程的规则式（2.8）已经能实现将非选择解释和选择解释区分开的目的，因此单有这一界限的有效性尚不足以辩护 HRG 异于更简单的规则式（2.8）的解释价值。

2.4.2 与其它分类方式的不兼容性

HRG 并不是唯一能够实现分类功能的规则。我们刚刚提到规则式（2.8）也能实现非选择解释和选择解释的二分，只是过于粗糙。奎莱尔（Queller, 2011）还提出过一种考虑协同效应^①的汉密尔顿规则，它同样能实现对具体社会进化模型的分类。我们对其简要介绍如下。建立有机体适合度 w 对其自身育种值 g 、其社会行为搭档育种值 g' 以及两者协同效应 gg' 的线性回归模型：

$$w = \alpha + \beta_{wg}g + \beta_{wg'}g' + \beta_{w,gg'}gg' + \varepsilon. \quad (2.10)$$

其中的三个回归系数都是偏回归系数，为方便起见不再在角标中详细注明。我们将式（2.10）代入简化的普莱斯方程（式（2.4））可得

$$\begin{aligned} \Delta \bar{g} &= \frac{1}{\bar{w}} \left(\beta_{wg} \text{var}(g) + \beta_{wg'} \text{cov}(g', g) + \beta_{w,gg'} \text{cov}(gg', g) \right) \\ &= \frac{1}{\bar{w}} (\beta_{wg} + \beta_{wg'} \beta_{g'g} + \beta_{w,gg'} \beta_{gg',g}) \text{var}(g). \end{aligned} \quad (2.11)$$

我们把 β_{wg} 看作该行为的代价 $(-c)$ ，把 $\beta_{wg'}$ 看作该行为的收益 $b^{\text{②}}$ ，把 $\beta_{w,gg'}$ 记作协同效应参数 d ，把 $\beta_{g'g}$ 看作亲缘系数 r ，把 $\beta_{gg',g}$ 记作协同系数 s 。由此，我们有

$$\Delta \bar{g} > 0 \text{ 当且仅当 } rb - c + sd > 0. \quad (2.12)$$

我们称规则式（2.12）是扩展的 HRG，它和 HRG 一样是数学上精确和普遍成立的。

① 协同效应（synergy effect）可以大致理解为多个行为主体交互作用时产生的效果较这些个体各自行为效果的简单加和的偏离程度。更通俗地讲，正的协同效应就好比团队协作时“1 + 1 > 2”的效果。

② 请读者注意：由于采用了不同的回归模型，这里对 c 和 b 的定义已经与 HRG 中原本的定义不同。

伯奇也考虑了扩展的 HRG 作为社会进化研究的组织框架的可能性，但认为其不如 HRG 更适合作为组织框架，理由是扩展的 HRG 提供的分类过于精细，以至于在实际应用时显得笨重和不必要（Birch, 2017b）⁷⁸。换言之，HRG 既（较扩展的 HRG）足够简单，（较规则（2.8））又有足够的信息量，因而作为组织框架的精细度是最佳的。这里的一个前提是 HRG 和扩展的 HRG 所提供的分类之间确实有一粗糙—精细的关系，但这实际上是伯奇的误解。

一方面，伯奇根据扩展的 HRG 中 rb 、 c 和 sd 这三个量的符号正负将具体的社会进化模型划分为了八类（见图 2.3）：（1） $rb > 0, c \geq 0, sd > 0$ ；（2） $rb > 0, c \geq 0, sd \leq 0$ ；（3） $rb > 0, c < 0, sd > 0$ ；（4） $rb > 0, c < 0, sd \leq 0$ ；（5） $rb \leq 0, c < 0, sd > 0$ ；（6） $rb \leq 0, c < 0, sd \leq 0$ ；（7） $rb \leq 0, c \geq 0, sd > 0$ ；（8） $rb \leq 0, c \geq 0, sd \leq 0$ 。^①另一方面，伯奇说扩展的 HRG 提供的划分可以由如下方式得到：将 HRG 提供的划分中选择区域的三类（即直接适合度解释、混合解释和间接适合度解释）各分为协同效应（即 sd ）为正和协同效应为负两个子区域，再额外补充一个仅有协同效应为正而 b 为负、 c 为正的区城，得到八个区域（Birch, 2017b）⁷⁸。但是，后一种说法和前一种说法是对应不上的，甚至后一种说法根本无法实际操作，原因在于 HRG 和扩展的 HRG 采用的回归模型不同，因而其中的参数 b 和 c 并不一致。下面我们举例说明这一点及其后果。

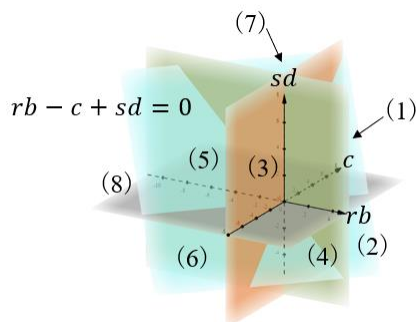


图 2.3 扩展的 HRG 提供的分类图示

如图 2.4，（A）展示了一个种群中所有个体之间的相互作用，其中深色的个体持有某基因，白色的个体不持有该基因；（B）根据（A）列出了该种群的三组数据 g 、 g' 和 gg' 。根据（B），我们容易算出该种群的亲缘系数 $r = \frac{7}{15}$ ，协同系数 $s = \frac{2}{3}$ 。

① 伯奇自己表述的分类似乎没有考虑这三个量为 0 的情形，我参考伯奇之前基于 HRG 的分类对伯奇的表述做了修正。

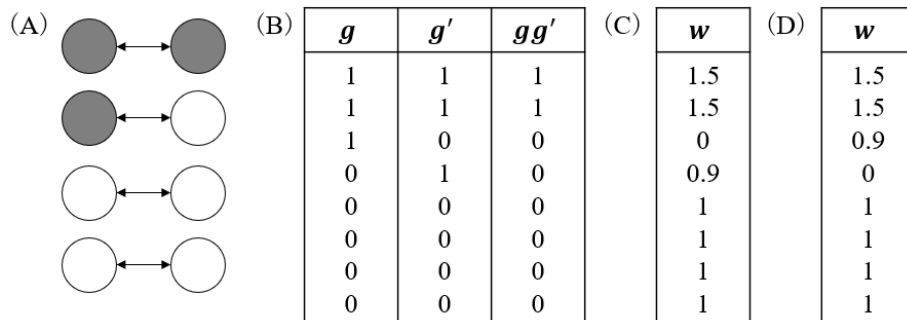


图 2.4 一个社会进化的案例

我们不妨把该种群看作一个乐团，其中的乐手两两组队排练并表演。深色的乐手会认真排练，白色的乐手则偷懒不排练。我们假定，一组乐手只有两人都认真排练，演出效果才能提升，获得更高收益；而独自一人排练不仅白费力气，还会打扰对方。

现在考虑两种不同的情形，两种情形下的适合度或收益数据分别如图 2.4(C)、(D) 所示。情形 (C) 中乐团表演的是古典乐，精致的古典乐需要深色乐手耗费很大精力认真排练，使自身收益大为减少（第 3 行深色乐手适合度为 0）；而古典乐较为柔和，因此深色乐手独自排练时对偷懒的白色乐手打扰并不明显（第 4 行白色乐手适合度为 0.9）。情形 (D) 中乐团表演的是摇滚乐，随性的摇滚乐对认真排练的深色乐手而言较为轻松（第 3 行深色乐手适合度为 0.9），但排练时吵闹的噪音会让偷懒的白色乐手收益大减（第 4 行白色乐手适合度为 0）。无论哪种情形，深色乐手组队认真排练时都会产生明显的收益（两种情形下前 2 行深色乐手的适合度都是最高的，为 1.5）。

我们考虑扩展的 HRG 对以上两种情形的分类。容易发现两种情形下都有 $b < 0$ 、 $c > 0$ 和 $d > 0$ ，且 $rb - c + sd > 0$ 成立，由此它们均会被扩展的 HRG 归入类别 (7)，直观意义是这两种情形下协同效应都是唯一使得种群平均育种值正向变化的因素。这非常符合我们的直觉，因为根据我们的设定，一组乐手中双方都排练（即 $gg' = 1$ ）时产生的协同效应才是为排练这一行为带来收益的唯一因素。

然而，HRG 对 (C) 和 (D) 两种情形给出的分类却有所不同。对情形 (C)，我们容易算出 b 和 c 均大于 0，由此将被 HRG 归入间接适合度解释；而对情形 (D)，我们容易算出 b 和 c 均小于 0，由此将被 HRG 归入直接适合度解释。我们发现，对于同一个社会进化情形，HRG 中的参数 b 和 c 对比扩展的 HRG 中的参数 b 和 c 不仅数值可能不同，连符号也未必一致。情形 (C) 和情形 (D) 根据在伯奇看来更精细的扩展的 HRG 会被分到同一类，但根据在伯奇看来更粗糙的 HRG 反而会被分为截然不同的类型！

这一差别的根源在于，HRG 强迫我们将本来由协同效应贡献的适合度组分强行分解到自身育种值和行为搭档育种值的贡献中去。^①也就是说，HRG 强迫我们回答：乐手排练到底是损人利己，还是损己利人？都不是！排练的收益就是双方都排练时协同效应贡献的收益！可见，对我们的案例而言，扩展的 HRG 才能提供有意义的分类，而基于 HRG 的分类反而是误导性的。

以上案例表明，我们不当把 HRG 和扩展的 HRG 提供的分类看作一粗糙一精细的关系，它们是根本上不同的。更一般地，按照图 2.4 (A) 所示的种群结构，HRG 和扩展的 HRG 提供的两种不同的分类在选择解释区域的比较如图 2.5 所示，其中 (A) 展示了扩展的 HRG 对选择解释的分类，(B) 将其与基于 HRG 的分类进行比较，显然两者截然不同。因此，我们不能简单地通过在不同组织框架的简单性和信息量之间进行权衡来谋求最优的组织框架。

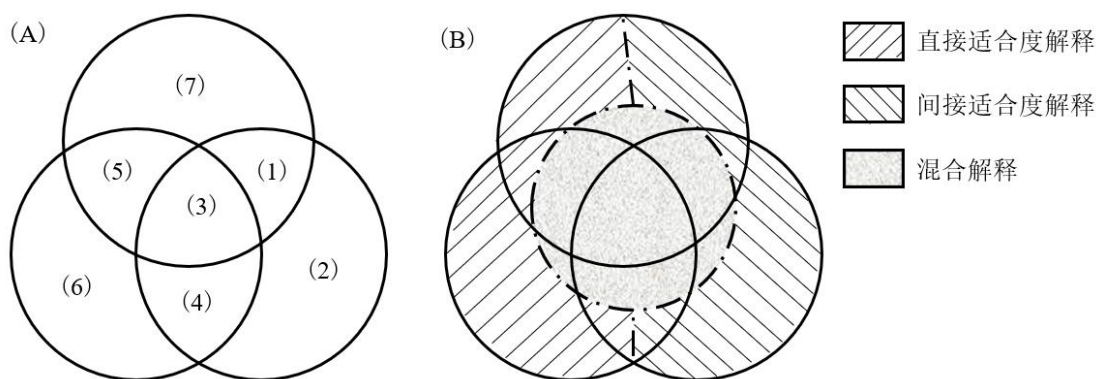


图 2.5 基于 HRG 和扩展的 HRG 的分类比较

2.5 汉密尔顿规则的多般形式及其解释价值

上一节中介绍的扩展的 HRG 并不是唯一可能的 HRG 的扩展版本。实际上，单从数学上看，我们完全可以用更复杂的回归模型来扩展 HRG。范·伟伦 (Mathijs van Veelen) 等人 (van Veelen et al., 2017) 严格证明了我们可以在回归模型中引入任意多个预测变量和相应的协同效应项，并且由此产生的汉密尔顿规则的一般形式和 HRG 一样精确和普遍成立。简而言之，基于不同的回归模型可以导出汉密尔顿规则的不同一般形式。显然，汉密尔顿规则的各种一般形式都是组织框架的有效候选者，因为它们中的任一个都可以根据其构成项的符号提供对社会进化情形的分类。但是，汉密尔顿规则的不同一般形式所提供的不同分类是不相容的。那么我们可能会问：哪种一般形式的汉密尔顿规则是最优的？回想一下，我在上一节中反

^① 伯奇其实注意到了这一点 (Birch, 2017b) 72, 82, footnote 18, 但似乎没有充分认识其后果。

驳了汉密尔顿规则的不同一般形式间具有可良好比较的精细度关系。考虑到这一点，我不认为有任何简单的方法可以挑选出汉密尔顿规则的唯一最优形式作为大多数情况下的唯一最优组织框架。相反，我建议将我们的注意力转移到以下问题上：在某种给定情形下，哪种一般形式的汉密尔顿规则是最优的？需要说明的是，我并非意图简单地将作为最优组织框架的属性相对化到某些语境下，给出类似“汉密尔顿规则的第一种形式在语境 A 中是最优的，汉密尔顿规则的第 2 种形式在语境 B 中是最优的……”的答案。实际上，恰恰是判断哪种形式的汉密尔顿规则最适合给定情形的工作能够生成一个新的维度供我们比较和整合特定的社会进化情形，进而增进我们对社会进化理解。

我建议，可以根据以下两个维度比较和整合特定的社会进化情形：

(D1) 汉密尔顿规则的哪种一般形式最适合表征给定情形？

(D2) 根据汉密尔顿规则的某一一般形式（最好是最适合表征给定情形的那种形式）提供的分类，给定情形属于哪一类型？

再次以图 2.4 中的情形 (C) 和 (D) 为例。在维度 (D1) 上比较，(C) 和 (D) 属于同一类别，即扩展的 HRG 是最适合表征这两种情形的汉密尔顿规则。其理由是，这两种情形都包含显著的协同效应。在维度 (D2) 上比较，(C) 和 (D) 仍然属于同一类别，即类别 (7)，其中 $rb \leq 0$ ， $c \geq 0$ 且 $sd > 0$ 。其理由是，在这两种情况下，协同效应是育种值正向变化的唯一贡献因素。以上比较能够令人信服地揭示 (C)、(D) 两种情形在两个不同维度上的共同本质。然而，如果我们只根据维度 (D2) 且仅将 HRG 作为唯一的分类标准对它们进行比较，我们将无法捕捉到它们的共同本质，而只能得出一些似是而非的差异。因此，如果我们想更好地理解社会进化，就必须考虑维度 (D1)，进而维度 (D1) 清楚地展示了汉密尔顿规则的多种一般形式（而非仅 HRG 一种形式）作为比较和整合社会进化情形的标准的解释价值。

现在考虑对我的上述方案的一点可能的反对意见。有人可能会认为，只要我们在回归模型中添加更多的预测变量，它就可以更准确地表征真实情形。那么，只要我们的能力允许，最好用更精细的汉密尔顿规则对社会进化情形进行分类。出于这个考虑，维度 (D1) 听起来极度平凡且无聊。

我对此有两点回应。首先，更精细的回归模型未必能更好地表征给定社会进化情形，这是因为实际情形较一个粗糙模型的偏差可能只是由某些随机误差导致的，而并不反映任何根本的机制。因此，引入过多的预测变量可能会导致过拟合的问题，这将使相应的精细分类不仅冗余，而且具有误导性。其次，我们可以在经验上检验回归模型中的某些预测变量是否对进化过程有显著贡献。例如，史密斯等人

(Smith et al., 2010) 研究了某种微生物间合作的进化, 发现模型中收益项和协同效应项的泰勒展开中的 10 至 15 阶项最为重要。他们的研究结果表明, 这种合作行为进化的最优分类所需的汉密尔顿规则的一般形式应包括这些 10 到 15 阶的项, 而不必包括更高阶的项。

然而, 退后一步, 我承认本文没有提供一个完全明确的对某种一般形式的汉密尔顿规则是否适于表征某一情形的判断依据——我已经给出的判断基本上是直觉的。我想强调, 本文的主要关注点在于辩护维度 (D1) 的重要性——即便这一点也经常被忽视甚至拒斥。至于如何精确刻画维度 (D1) 以使其更易应用, 仍需本文范围以外的进一步的工作。

2.6 本章小结

汉密尔顿规则有不同的版本, HRS 可以自然地用于因果解释, 但在 HRG 中社会行为的收益和代价都由回归系数定义, 其生物学意义并不清楚, HRG 的解释价值也备受争议。本文分析得出, HRG 不仅不适用于经验预测或经受经验检验, 而且作为社会进化研究的组织框架也没有明显的优势。一方面, 参数 b 和 c 的不稳定性使得 HRG 提供的相应分类也不稳定; 另一方面, HRG 中的参数 b 和 c 与其它规则如扩展的 HRG 中的参数并不一致, 使得 HRG 和可能作为组织框架的其它候选项之间没有简单的优劣关系。

一个值得注意的事实是, 无论是支持还是反对 HRG 解释价值的学者, 对 HRG 的理解往往都停留在第 2.2 节介绍的原始版本, 而很少考虑 HRG 的各类推广或变形。我建议在根据某种形式的汉密尔顿规则的参数符号对场景进行排序 (维度 (D2)) 之前, 考虑哪种汉密尔顿规则的一般形式最适合给定的场景 (维度 (D1))。我们可以在这个新的维度上比较和整合特定的社会进化情景, 这有助于揭示被单独坚持 HRG 所掩盖的共同本质或显著差异。我的建议更好地解释了一般汉密尔顿规则的解释价值——不仅仅是 HRG, 而且是汉密尔顿规则的各种一般形式。

第3章 广义适合度理论、亲缘选择与真社会性的进化

3.1 引言：真社会性的进化解释问题

在上一章中，我评估了广义适合度理论之争中论战双方就汉密尔顿规则的解价值产生的分歧，这是针对广义适合度理论本身的。在本章中，我将处理广义适合度理论之争中另一个方面的争议，即广义适合度理论（作为一种科学理论）和真社会性的进化（作为一种经验现象）之间的关系问题。自20世纪60年代汉密尔顿提出广义适合度理论并用以解释真社会性的进化以来，人们普遍承认广义适合度理论能够解释真社会性的进化，反过来真社会性的进化也确证了广义适合度理论^①。

真社会性在当前文献中通常指社会群体中出现的繁殖分工现象，即某些成员专门从事繁殖工作而另一些成员专门照料其它成员的子女，从而分化为不同的品级（caste）。^②最有代表性的真社会性出现在社会性昆虫中，包括蜜蜂、胡蜂、蚂蚁、白蚁等，本文对真社会性的讨论也限定在社会性昆虫的语境下。尽管学者们对真社会性在昆虫以外的推广较为谨慎，但许多人相信对真社会性的研究有助于我们理解更复杂的社会组织形式。许多学者都将多细胞有机体向真社会性社会的进化看作是进化史上的重大进化跃迁（major evolutionary transition）之一（Maynard Smith & Szathmáry, 1998; Bourke, 2011; Szathmáry, 2015），该跃迁形成的真社会性社会（如蜂群）还可被看作更高层级的生物个体（Queller & Strassmann, 2009; Bourke, 2011），可见其重要地位。

真社会性昆虫中出现的不繁殖的工蜂或工蚁（英文统称 worker，为方便起见后文仅谈论“工蜂”作为代表）品级是进化中较难解释的现象。从进化论的角度看，不繁殖而仅照料蜂后（或蚁后，英文统称 queen，为方便起见后文仅谈论“蜂后”作为代表）的子女是一种典型的利他行为，这将使工蜂在进化中处于不利地位。20世纪60年代，汉密尔顿提出广义适合度理论的同时，也将其应用于解释真社会性的进化这一案例（Hamilton, 1964；参见本文第2章的相关讨论）。简单来说，广义适合度理论要求我们将社会行为主体赋予他人的收益也部分地算作其自身的广义

① 本文无力仔细处理“科学理论甲能够解释经验现象乙”和“经验现象乙能够确证科学理论甲”这两个论断之间的区别，而简单认为它们大致是等价的。

② 真社会性这一概念在历史上有许多不同的含义和用法，且与超个体（superorganism）概念有密切联系。布姆斯玛（Jacobus J. Boomsma）和高恩（Richard Gawne）对这两个概念的含义演变给出了详细的辨析（Boomsma & Gawne, 2017）。本文仅在较为朴素的意义理解真社会性概念，而不能处理有关概念争议的细节。

适合度。对于利他行为而言，如果行为接受者的收益较行为主体的付出足够大，且两者的亲缘关系足够近，那么利他行为就可以获得进化上的优势。这种通过为近亲带来的利益而在进化中被选择的机制被称作亲缘选择（Maynard Smith, 1964）。在具有真社会性的蜂群中，姐妹工蜂的亲缘关系非常近，因此极有利于其间利他行为的进化。

尽管汉密尔顿的解释中的具体细节后来受到了挑战，但“亲缘系数在真社会性的进化中具有重要地位”这一核心洞见在广义适合度理论和亲缘选择的支持者那里被保留了下来，由此坚持广义适合度理论和亲缘选择在真社会性的进化案例中有良好的解释效力（Queller & Strassmann, 1998; Abbot et al., 2011; Bourke, 2011; Bourke, 2015; Queller, 2016）。然而，21 世纪以来，以威尔逊为代表的生物学家开始拒斥亲缘选择理论对真社会性进化的解释效力（Wilson, 2005; Wilson & Hölldobler, 2005; Wilson, 2008）。这一论断也是 2010 年以来的广义适合度理论之争中批评者的一个常被忽视的重要论点（Nowak, Tarnita et al., 2010; Wilson & Nowak, 2014; Nowak & Allen, 2015）。总地来看，批评者一方面否认亲缘系数在真社会性进化中的重要因果效力（Wilson, 2005; Wilson & Hölldobler, 2005; Wilson & Nowak, 2014），另一方面认为广义适合度理论自身的缺陷使其根本难以应用于包括真社会性进化在内的经验研究（Nowak, Tarnita et al., 2010; Nowak & Allen, 2015; Nowak, McAvoy et al., 2017）。由此可见，生物学家就亲缘系数在真社会性的进化中的因果角色有着严重分歧。许多哲学家已经就有关广义适合度理论和亲缘选择的争论给出了概念层面和方法论层面的分析（Birch & Okasha, 2015; Birch, 2017a; Rubin, 2018），但他们往往回避了上述广义适合度理论和亲缘选择在真社会性进化中的解释效力问题^①。

本文试图正面处理生物学家有关亲缘系数在真社会性的进化中的因果效力的分歧，指出争议双方对“亲缘系数”和“因果效力”的不同理解是产生争议的重要原因，进而试图对批评者有价值的意见做合理重构。我将在第 3.2 节中区分两种不同的亲缘系数概念，我分别称为 IBD 亲缘系数、IBS 亲缘系数和统计亲缘系数（statistical relatedness），并在第 3.3 节中澄清以上两种不同的亲缘系数概念在真社会性的进化解释争议中引发的混淆。在第 3.4 节中，我将辩护应将系谱亲缘系数的因果效力作为。在第 3.5 节中，我将引入“类型因果（type causation）和因果相关性（causal relevance）”以及“净效果（net effect）和组分效果（component effect）”两组区分，试图部分澄清争议双方就亲缘系数因果效力的不同关切。我将在第 3.6

^① 一个例外是佩尔努（Tuomas K. Pernu）和海兰泰雷（Heikki Helanterä）的论文（Pernu & Helanterä, 2019a; 2019b），但我认为他们的分析未能有效刻画出批评者真正有价值的意见。

节中总结本章,并对广义适合度理论、亲缘选择与真社会性的进化三者间的关系给出简要的评估。

3.2 IBD 亲缘系数、IBS 亲缘系数与统计亲缘系数

在汉密尔顿(Hamilton, 1964)提出广义适合度理论时,他将赖特(Sewall Wright, 1922)定义的“关系系数(coefficient of relationship)”用作亲缘系数^{①②}。这种定义下的亲缘系数的含义是两个个体某一基因座上具有血缘同源(identical-by-descent, 即 IBD)基因的比例或概率,我将其称为 IBD 亲缘系数。所谓 IBD 是指两个个体共同具有的某一基因遗传自某一共同祖先。IBD 亲缘系数通常以通径法计算。对二倍体生物,在不存在近交(inbreeding)时,有 IBD 亲缘系数 $R = \sum 0.5^L$,^③其中 L 表示一条经由某共同祖先连接两个体的进化路径上的代际数,求和符号针对所有这样的路径求和。以同父同母的姐妹为例,两者共有父亲和母亲两个共同祖先,且两条进化路径(姐姐→父亲→妹妹;姐姐→母亲→妹妹)的长度都是 2,因此有 $R = 0.5^2 + 0.5^2 = 0.5$ 。这表明,同父同母的姐妹之间有大半比例的基因是相同的。而对于同母异父的姐妹而言,仅有通过母亲的进化路径能够连接她们,因此有 $R = 0.5^2 = 0.25$ 。这表明,同母异父的姐妹间 IBD 基因的比例大致有四分之一。

对于广义适合度理论而言,亲缘系数实际上就可以作为衡量利他行为收益的一个标准。根据广义适合度理论中的核心内容汉密尔顿规则,假定一种利他行为使行为主体付出的代价是 C ,使行为接受者得到的收益是 B ,双方的亲缘系数是 R ,那么若 $RB - C > 0$,则该利他行为能够在进化中占优(Hamilton, 1964)。举例来说,假定同父同母的姐妹之间施行某种利他行为会使行为主体自身少繁殖 1 个后代,而将使行为接受者多繁殖 3 个后代。此时我们有 $RB - C = 0.5 \times 3 - 1 = 0.5 > 0$,说明施行该利他行为是有利可图的。直观上,这是因为行为主体若自己繁殖 1 个后代则仅有 $1/2$ 的基因得以传递(考虑有性生殖),而若使与自己有多达一半比例基因相同的姐妹繁殖 3 个后代,则将有 $(1/2) \times (1/2) \times 3 = 3/4$ 的基因得以传递,

① 通常而言,英文“coefficient of relationship”、“coefficient of kinship”、“kinship coefficient”、“coefficient of relatedness”和“relatedness”等大致对应于“亲缘系数”的不同短语在中文中被统一译为“亲缘系数”。更麻烦的是,这些不同的英文写法往往并没能反映出有实质意义的概念区别,很多时候只是单纯地混用。反过来,不同亲缘系数概念的差别也往往未能通过区分不同的写法而反映出来。本章中,除特殊说明外,我采用“亲缘系数”泛指以上所有英文短语,并通过“IBD 亲缘系数”、“IBS 亲缘系数”和“统计亲缘系数”三种概念区分出亲缘系数概念下有实质差异的三种含义。

② 需要说明的是,汉密尔顿(Hamilton, 1964; 1972)在较早的论文中已经意识到了“关系系数”只能作为亲缘系数的良好近似,但他也并未明确广义适合度理论中亲缘系数应有的严格定义方式。

③ 我使用大小写区分 IBD 或 IBS 亲缘系数与统计亲缘系数。大写 R 表示 IBD 或 IBS 亲缘系数,小写 r 表示统计亲缘系数。

这自然是更优的选择。由此可见，选择那些有利于近亲而非自己的性状及其基因的亲缘选择是一种独特的选择机制，对利他主义的进化有很好的解释力。

在此，我们需要注意血缘同源（IBD）和状态同源（identical-by-state，即 IBS）的区分^①。所谓 IBS 是指两个个体无论出于何种原因而具有相同基因，而不像 IBD 要求相同的基因是由于遗传自共同祖先导致的。在此种理解下，IBD 蕴涵 IBS，但 IBS 不蕴涵 IBD。一种典型的符合 IBS 但不符合 IBD 的情形是相似的基因突变导致两个个体偶然地具有相同基因。进一步地，IBS 亲缘系数即两个个体在某一基因座上无论出于何种原因具有相同基因的比例。考虑到 IBS 与 IBD 之间的单向蕴涵关系，IBS 亲缘系数的值将大于或等于 IBD 亲缘系数的值，这或许是两者最明显的差异。然而，如果我们假定两个个体中所有我们所关注的共同基因都是遗传自共同祖先而来的，是否就意味着 IBS 亲缘系数和 IBD 亲缘系数在数值上等同呢？严格根据定义看或许如此，但如果考虑生物学实践则未必。这是因为，实践上 IBS 基因的比例往往是直接通过比较两个体的基因信息得到准确数值，而 IBD 基因的比例往往是通过系谱信息预测的期望值。举例而言，我们前面由通径法算得同父同母的姐妹间 IBD 基因比例的期望是 $1/2$ ，但实际上该姐妹完全可能从父母那里遗传更多或更少的相同基因。进而，尽管 IBS 亲缘系数自然地会反应两个个体相同基因比例的实际值，IBD 亲缘系数就其算法而言很可能不考虑实际值较预测值的偏差，而只提供基于系谱信息得到的预测值。我们可以认为，在理想情况下，基于系谱信息算得的 IBD 亲缘系数能够反映两个个体实际 IBD 基因的比例；而在不考虑相似突变等特殊情形时，IBD 亲缘系数也将是 IBS 亲缘系数的良好近似。这样看来，两者的差异更多地是概念上的，在理想条件下不至于引致数值上的差异。

统计亲缘系数与 IBD 亲缘系数、IBS 亲缘系数非常不同。第2章中曾提到，汉密尔顿（Hamilton, 1964）提出广义适合度理论特别是汉密尔顿规则时，采用的是极其简化的模型，依赖于支付可加、群体结构简单等众多限制性条件才能成立，这使得汉密尔顿的理论在复杂的经验情形下难以普遍应用（Queller, 1985; Nowak, Tarnita et al., 2010; van Veelen et al., 2017）。为此，理论家们对汉密尔顿规则做了推广，得到了汉密尔顿规则的一般形式，即 HRG（Queller, 1992；参见本文第2章的相关讨论）。HRG 仍有“ $rb - c > 0$ ”的形式^②，而几乎无须任何限制条件而普遍成立，但其中的参数均由某种回归系数所取代。特别地，其中的亲缘系数（ r ）被定

① 由于 IBD、IBS 以及相关的亲缘系数概念在各类文献中已经被严重混淆，我在此不敢保证我所提供的定义和区分能够在事实上符合绝大多数人对相关概念的用法。我仅试图在最低限度上提供一种合理的定义和区分，并用于澄清相关的理论争议。

② 我在汉密尔顿规则的一般形式中采用小写字母，在汉密尔顿规则的经典形式中采用大写字母，以对两者做出区分。

义为群体中行为搭档基因值 (g') 关于行为主体基因值 (g) 的回归系数 ($\beta_{g'g}$)^①, 我称之为统计亲缘系数。举例而言, 我们考虑一个包含甲、乙、丙和丁共 4 个个体的种群, 其中甲和乙进行社会互动, 丙和丁进行社会互动, 四者有关社会行为的基因值分别为 1、1、1 和 0。由此, 我们共有四组基因值 (g, g'), 即 (1,1)、(1,1)、(1,0) 和 (0,1)。进而, 由线性模型拟合可得回归系数 $\beta_{g'g} = -1/12$, 这也就是该群体的统计亲缘系数。

尽管在极度简化的情形下 IBD 或 IBS 亲缘系数或许能够作为统计亲缘系数的良好近似 (Michod & Hamilton, 1980), 但统计亲缘系数与前两者有着巨大的实质差异。我参考伯奇 (Birch, 2014b) 的工作^②将其概括为三点。第一, 统计亲缘系数一定是针对某个特定性状或基因而言的, 而 IBD 或 IBS 亲缘系数尽管理应同样定义在具体的基因座上, 但更多的时候 (特别是由通径法计算的 IBD 亲缘系数) 给人的印象是对两者亲缘关系的一般性的衡量, 而这种印象若应用于广义适合度理论计算则是完全错误的。第二, 统计亲缘系数是一个群体整体的性质, 而 IBD 或 IBS 亲缘系数是群体中某一组社会行为搭档的性质。第三, 统计亲缘系数可以为负, 负值的含义为群体中的个体更倾向于与基因不相似的个体进行社会互动, 而 IBD 或 IBS 亲缘系数总是正的。这些重大差异表明, 当我们因情形复杂而需要应用汉密尔顿规则的一般形式时, 其中的亲缘系数必须严格采用统计亲缘系数, 而不能使用 IBD 或 IBS 亲缘系数随意取代。

3.3 真社会性的进化解释争议中的亲缘系数

有了对两种亲缘系数的区分, 下面我们考虑真社会性的进化解释中究竟关涉何种意义上的亲缘系数。首先考虑汉密尔顿 (Hamilton, 1964) 基于广义适合度理论对真社会性的进化提供的经典解释。其中的关键是汉密尔顿所提出的单倍二倍性假说 (haplodiploidy hypothesis), 即单倍二倍性是真社会性进化的主要驱动力。我们知道, 膜翅目 (Hymenoptera) 昆虫具有单倍二倍性, 雄性个体由未受精卵发育而成、是单倍体, 雌性个体由受精卵发育而成、是二倍体。考虑一个一妻一夫制的蜂群, 对其中一雌性工蜂而言, 它和其它姐妹工蜂 (或者说是蜂后的女儿) 的“亲属系数 (coefficient of kinship)”或亲缘系数是 0.75; 而如果和无亲缘关系的雄蜂交

① 基因值可简单理解为某一基因座上目标基因的数量或比例 (采用数量和比例并无实质区别)。当然, 考虑到社会行为因其复杂性一般由多个基因座上的基因共同决定, 更实际的做法是采用育种值 (breeding value) 即可加遗传效应 (additive genetic effect) 代替基因值 (Queller, 1992)。

② 伯奇 (Birch, 2014a) (以及通常) 所说的系谱亲缘系数 (pedigree relatedness) 大致对应于我所说的 IBD 亲缘系数; 伯奇所说的另一种统计意义上的亲缘系数 (他似乎并未起一个明确的名字) 就是我所说的统计亲缘系数。伯奇似乎没能有意识地区分开 IBD 亲缘系数和 IBS 亲缘系数。

配，它和自己女儿的亲缘系数只有 0.5。因此，在同等条件下，相较于自己离开蜂群独立生育，辅佐蜂后生育自己的姐妹是更有利于自身基因传递的选择。这就解释了为何工蜂能够聚集在蜂群中形成真社会性社会。很明显，汉密尔顿这里所说的亲缘系数就是我所称的 IBD 亲缘系数，是根据两个体间的遗传系谱计算得出的，与其所处的社会环境无关。此外，汉密尔顿并未计算工蜂社会行为的代价 C 和收益 B ，也没有计算其广义适合度。因而，可以说汉密尔顿的解释中唯一的关键就是 IBD 亲缘系数。

需要说明的是，汉密尔顿提供的上述解释仅仅是尝试性的假说，他所提供的理论推导和经验证据都非常有限。尽管如此，单倍二倍性假说在当时学界仍然非常流行。例如，著名的进化生物学家、昆虫学家威尔逊（Wilson, 1975）在综述社会性昆虫的相关研究时明确提出亲缘选择唯一解释了膜翅目昆虫的真社会性，并且认为单倍二倍性假说就是解释的关键所在^①。特里弗斯（Robert L. Trivers）和黑尔（Hope Hare）将汉密尔顿的理论与费希尔（R. A. Fisher）的性比理论（sex ratio theory）结合起来，主张亲缘选择理论对社会性昆虫中性比的新奇预测得到了良好的经验确证，进而社会性昆虫的独特性状也能由亲缘选择理论唯一解释（Trivers & Hare, 1977）。

许多生物学家很快认识到单倍二倍性假说的问题，并对其展开了严厉的批评。安德森（Malte Andersson, 1984）在其综述中指出了反对单倍二倍性假说的五点证据，这包括（1）大量具有单倍二倍性的膜翅目昆虫并未进化出真社会性，而真社会性在不具有单倍二倍性的其它物种中也广泛出现；（2）在许多社会性蚂蚁和蜜蜂中，蜂后会与多个雄性交配，使得蜂群中同母异父的姐妹工蜂间的亲缘系数将低于工蜂与其女儿间的亲缘系数；（3）一些真社会性的蜜蜂和胡蜂蜂群中包含多个蜂后，此时工蜂和生殖蜂后之间的亲缘系数将会更低；（4）单倍二倍性假说对性比的预测与对实际性比的经验观察之间并不像特里弗斯和黑尔声称的那样高度吻合，即使在符合预测的经验情形下相应性比也可能有单倍二倍性以外的其它解释；（5）细致的建模分析显示，姐妹工蜂之间的高亲缘系数未必有利于不生育的工蜂品级出现。

上述对单倍二倍性假说的质疑很快被主流生物学家所接受。然而，亲缘选择理论的支持者并未将亲缘选择理论与单倍二倍性假说一起抛弃。安德森（Andersson, 1984）自己就明确主张，无论单倍二倍性假说成立与否，亲缘选择在真社会性的进化中都起着十分重要的作用。奎莱尔和施特拉斯曼（Joan E. Strassmann）在承认单倍二倍性假说并不成功的前提下，坚持主张真社会性的进化为亲缘选择提供了坚

① 威尔逊后来放弃了此观点，转而主张亲缘选择并非驱动真社会性进化的主要动力（Wilson, 2005）。

实的支持 (Queller & Strassmann, 1998)。尽管奎莱尔自己是基于统计亲缘系数的 HRG 的提出者, 他和施特拉斯曼在讨论真社会性进化的案例时仍然大量采用基于通径法计算的 IBD 亲缘系数, 一方面将亲缘系数看作是两个相互作用的个体间的关系而非群体的性质, 另一方面基本将亲缘系数的高低等同于群体内亲属关系的远近。他们认为, 大量的研究证实了亲缘选择理论的预测, 即社会性昆虫的利他行为是在亲属间发生的, 社会性的蜂群也就是一个家族 (Queller & Strassmann, 1998)。

不过, 单倍二倍性假说的失败也引致了对亲缘选择理论的反对意见。以威尔逊为代表的反对者开始主张亲缘选择并非驱动真社会性进化的主要动力, 并把单倍二倍性假说这一原本支持亲缘选择的证据的失败看作是一项重要理由 (但不是唯一理由) (Wilson, 2005; Wilson & Hölldobler, 2005; Nowak, Tarnita et al., 2010)。值得注意的是, 威尔逊在其论文中特别强调, 对真社会性的起源而言重要的是共同具有真社会性基因, 而非亲缘系数。他还说, 理论上即便亲缘系数为 0, 也有可能产生真社会性 (Wilson & Hölldobler, 2005)。这些说法表明, 威尔逊将亲缘系数概念限定为了 IBD 亲缘系数——即便两个个体并非亲属, 只要它们共同具有真社会性基因, 就会导致两者间一定的正 IBS 亲缘系数。

针对威尔逊的反对意见, 广义适合度理论和亲缘选择的福斯特 (Kevin R. Foster) 等人专门撰文驳斥 (Foster, Wenseleers & Ratnieks, 2006; Foster, Wenseleers, Ratnieks & Queller, 2006), 坚称亲缘选择是解释真社会性进化的关键, 且广义适合度理论没有任何问题。他们认为, 威尔逊把广义适合度理论中的亲缘系数限定为共同祖先导致的亲缘系数是一种误解, 特别强调亲缘系数可以用于表示任何形式的基因相关性 (Foster, Wenseleers & Ratnieks, 2006)。也就是说, 他们认为不能将亲缘系数等同于我所说的 IBD 亲缘系数, 而至少要考虑 IBS 亲缘系数^①。由此, 他们驳斥了威尔逊关于真社会性的进化甚至允许亲缘系数为 0 的“荒谬”主张。这并不奇怪, 因为威尔逊所说的是 IBD 亲缘系数, 他也并未举出连 IBS 亲缘系数都为 0 却仍有真社会性的例子。另外, 值得注意的是, 福斯特等人的两篇论文更多地是理论性的, 并未深入讨论 IBS 亲缘系数或统计亲缘系数如何与经验研究文献嫁接起来。不过, 有一些其他学者明确指出, 近亲关系就是真社会性的进化中 IBS 亲缘系数或统计亲缘系数为正的最主要的贡献要素, 正面辩护了 IBD 亲缘系数在真社会性进化中的重要原因地位 (Helanterä & Bargum, 2007; Pernu & Helanterä, 2019a)。

① 他们理解的亲缘系数究竟是 IBS 亲缘系数还是统计亲缘系数似乎并不清楚。但考虑到他们引用了奎莱尔提出 HRG 的论文 (Queller, 1992), 至少他们应当相信统计亲缘系数是最一般的广义适合度理论中应当采用的亲缘系数。也许他们同意在理想条件下 IBS 亲缘系数可以作为统计亲缘系数的良好近似。

也就是说，在他们看来，威尔逊所设想的 IBD 亲缘系数为 0 但 IBS 亲缘系数为正的情形即便在理论上可能，也是不符合真社会性进化的实际情况的。

通过以上分析可以发现，真社会性的进化解释争议中大家主要关心的是 IBD 亲缘系数。尽管广义适合度理论的支持者清楚严格的、一般的广义适合度理论应当采用统计亲缘系数而非 IBD 亲缘系数，但却很少提供基于统计亲缘系数的对经验案例的分析。

3.4 亲缘系数与关于自然选择的统计主义之争

佩尔努和海兰泰雷的工作少有地从哲学角度分析了有关真社会性的进化中亲缘系数的因果角色争议 (Pernu & Helanterä, 2019a; 2019b)。他们主要有如下两个论点：(1) 亲缘系数是真社会性的进化的必要不充分条件；(2) 至于亲缘系数在真社会性进化中的效力是否是因果的，则取决于自然选择是否是因果过程——他们自己认为自然选择当然是因果过程。我认为以上两个论点都是似是而非的，其部分缘由就在于他们没能有意识地清晰区分不同的亲缘系数概念。本节开始，我将从他们的工作出发重新评估亲缘系数在真社会性的进化中的因果角色相关争议。我将在本节论证，如果将争议中的亲缘系数概念理解为 IBD 亲缘系数，那么其效力是否是因果的与关于自然选择的统计主义之争无关；我将在第 3.5 节引入“因果相关性与类型因果”和“净效果与组分效果”两组因果概念的区分，试图更精细地刻画生物学家围绕真社会性的进化中亲缘系数的因果角色可能持有的不同观点。

佩尔努和海兰泰雷 (Pernu & Helanterä, 2019a) 指出，争议双方都同意真社会性的进化是自然选择（而非漂变或突变等非选择因素）导致的。而亲缘系数又是真社会性状可遗传的必要条件，因为不生育的工蜂只有通过亲属才能使其性状得以间接地遗传。既然如此，如果承认自然选择是因果过程，那么亲缘系数在其中自然起到了实质性的因果贡献。问题在于，谁会反对上述论证，特别地，会反对自然选择是因果过程呢？威尔逊、诺瓦克等亲缘系数因果效力的批评者看起来并无直接针对上述论证的反对意见^①。佩尔努和海兰泰雷认为可能反对上述论证的是自然选择的统计主义者 (statisticalist)。他们指出，统计主义者认为自然选择是纯粹的统计过程，进而至多接受亲缘系数是真社会性进化的必要条件，而不承认其是必要的因果条件 (Pernu & Helanterä, 2019a)。我认为他们的诊断是误导性的。

^① 也正因如此，这个问题本来也并非有关真社会性的进化中亲缘系数的因果角色争议中的核心议题。不过，与其直接相关的关于自然选择的统计主义则是近 20 年来生物学哲学界热议的话题，因此我仍然严肃对待并试图澄清这一问题。

统计主义 (statisticalism) 是一种关于自然选择地位的哲学立场, 代表人物包括沃尔什 (Denis M. Walsh)、阿里乌 (André Ariew) 等 (Walsh, Lewens et al., 2002; Walsh, Ariew et al., 2017)。根据统计主义者最新的纲领性文章 (Walsh, Ariew et al., 2017), 他们首先将适合度分为了个体适合度 (vernacular fitness) 和性状适合度 (trait fitness), 前者是某一具体个体的因果性质, 后者指某一抽象性状的增长率。进一步地, 统计主义者将自然选择区分为达尔文式选择 (Darwinian selection, 简称“D 选择”) 和现代综合选择 (Modern Synthesis selection, 简称“MS 选择”), 前者指个体适合度差异导致的种群世系结构变化, 后者指性状适合度差异导致的种群性状分布变化。相应地, 表征前者的 D 模型会援引个体层面的生态原因来解释种群世系结构变化, 表征后者的 MS 模型则仅援引性状类型的统计性质来解释性状分布的变化。统计主义者特别强调, 所谓“统计主义”针对的仅是 MS 模型, 即主张 MS 模型仅通过援引性状的统计性质而对种群变化提供自足的统计解释, 不援引种群变化的原因。D 模型则不是统计主义直接关心的对象——这一点似乎没有得到统计主义的反对者的应有重视。

现在的问题是, 亲缘系数在真社会性进化中的地位问题究竟关涉哪一种模型? 我认为, 如果考虑统计亲缘系数, 那么其效力确实由 MS 模型表征。这是因为, MS 模型要求考虑抽象性状而非具体个体的适合度, 而 IBD 或 IBS 亲缘系数都只是对两个具体个体之间亲缘关系的衡量, 只有采用作为群体性质的统计亲缘系数才能一般地衡量我们所关注的某类性状在多大程度上与基因型相似的个体进行了相互作用。进而, 如果采用统计主义立场, 就会否认亲缘系数在真社会性的进化中有任何因果效力。这似乎也符合佩尔努和海兰泰雷的论点。^①

然而, 统计亲缘系数并非关于亲缘系数在真社会性中的因果角色争议的重点。回想第 3.3 节的分析, 尽管统计亲缘系数是一般的广义适合度理论的构成基础, 生物学家实际关心的主要还是 IBD 亲缘系数。而 IBD 亲缘系数恰好对应于统计主义所不关涉的 D 模型, 统计主义者未必反对 IBD 亲缘系数也即群体中的近亲关系能够作为真社会性进化的生态学原因, 即便反对也不会是出于统计主义的要旨。因此, 统计主义立场与真社会性的进化解释争议是基本无关的^②, 佩尔努和海兰泰雷高估了统计主义在此的重要性。

对于真社会性的进化解释争议的重点在于 IBD 亲缘系数, 除了符合论战的实际情况外, 还有一点额外的理由。考虑争议的大背景即广义适合度理论之争, 支持

① 至于佩尔努和海兰泰雷 (Pernu & Helanterä, 2019a) 提供的自然选择是因果过程的理由是否足以反驳统计主义立场, 本文不再展开讨论。

② 这里所说的无关主要不是指该争议中没有统计主义者参与, 而是指持有统计主义立场与否基本不会影响争议双方的观点。

亲缘系数在真社会性进化中的因果效力的往往同样支持用亲缘选择解释真社会性的进化，而反对者往往建议用群体选择取代亲缘选择。然而，如果采用 IBS 亲缘系数或统计亲缘系数，亲缘选择和群体选择至少在形式上是等价的。由于正的 IBS 亲缘系数或统计亲缘系数都可以有包括亲属关系、扩散限制（limited dispersal）等多种可能的贡献因素，而 IBD 刚好衡量了亲属关系对应的 IBS 亲缘系数或统计亲缘系数的贡献，这恰恰是狭义上我们所说的亲缘选择所要求的（see Birch, 2020）^①。从统计主义的视角看，正的 IBS 亲缘系数和统计亲缘系数这一统计性质可能有多种个体层面的生态原因，恰好说明了其作为统计性质没有统一的因果解读，这也就是援引统计性质的 MS 模型和强调个体因果的 D 模型之间的区别。因此，IBD 亲缘系数（而非 IBS 亲缘系数或统计亲缘系数）的因果效力与争议背后的大问题“亲缘选择能否解释真社会性的进化”息息相关，这是争议的重点，却与统计主义无关。

3.5 真社会性的进化中亲缘系数的因果效力

本节中，除非特别说明，我们将讨论限定在 IBD 亲缘系数上，而省去“IBD”这一标记。前面提到，佩尔努和海兰泰雷（Pernu & Helanterä, 2019a; 2019b）的一个论点是亲缘系数是真社会性进化的必要不充分条件。但这其实并不新奇，因为明显批评者至少不否认 IBS 亲缘系数是真社会性进化的必要条件^②，而支持者也并不否认亲缘系数不是真社会性进化的充分条件。问题在于，在我看来，批评者在这一问题上的一些重要意见未能通过必要条件和充分条件的简单区分得到精确刻画。为此，我引入“因果相关性与类型因果”和“净效果与组分效果”两组区分来细致区分支持者和批评者就亲缘系数的因果角色问题的不同关切。

3.5.1 因果相关性与类型因果

亲缘系数因果效力的支持者对他们所谈的因果给出过如下定义：“我们认为因果性的意思是亲缘系数的改变会导致真社会性进化可能性的改变。”（Liao et al., 2015）³ 这一因果定义非常符合原因作为差异制造者的常规理解。如果反映在具体模型中，即可体现为亲缘系数这一变量和真社会性基因频率变化这一变量之间的因果，这也符合因果模型文献中对变量之间因果的理解。

① 当然，部分学者如加德纳等人（Gardner et al., 2011）将亲缘选择理论等同于广义适合度理论，这样亲缘选择也就不一定指出于近亲利益的选择了。我将这种对亲缘选择的理解看作是广义的。注意这种理解似乎并不符合其他支持亲缘选择学者的意见（see Queller, 2016）。

② 当然，佩尔努和海兰泰雷（Pernu & Helanterä, 2019a）强调仅有正的 IBS 亲缘系数却没有正的 IBD 亲缘系数并不符合真社会性进化的实际，这无疑是对部分批评者意见的正面反驳。

需要注意的是，变量之间的因果和通常所说的类型因果并不等同：伍德沃德（James Woodward, 2003）有时将变量之间的因果称为“因果相关性（causal relevance）”^①以示和类型因果的区分；盖勒（J. Dmitri Gallow, 2022）将类型因果界定为常数之间的因果，而将变量之间的因果称为“影响（influence）”^②。我在此采用伍德沃德的术语。举例而言，“吸烟导致肺癌”是典型的类型因果表述。据此，对任一个体，如果他吸烟，那么他将得肺癌。然而，如果我们将吸烟与否和肺癌与否都看作取值0或1的二值变量，那么两者因果相关性的表述“吸烟变量与肺癌变量因果相关”至少可能对应两种类型因果关系：其一是吸烟导致肺癌（而不吸烟导致不肺癌）；其二是吸烟导致不得肺癌（而不吸烟导致肺癌）。尽管后者并不符合实际情况，但逻辑上仍然符合因果相关性的要求，简单来说即一个人吸烟与否的变化将导致其得肺癌与否的变化。可见，因果相关性在上述意义上是一个比类型因果更弱的概念。

至少按照梅纳德·史密斯（John Maynard Smith, 1964）的经典定义，亲缘选择所说的是行为主体和搭档之间的高亲缘系数将导致有利于搭档的行为被选择，而不仅仅是亲缘系数在选择中会起作用。对此，从因果角度的解读应为高亲缘系数是利他行为被选择的类型原因。批评者的许多意见都是针对这一点的。其中最有代表性的是威尔逊，他主张亲缘选择在真社会性的进化中起到的主要是消解性的作用（Wilson & Hölldobler, 2005）。注意到，这并未否认亲缘系数与真社会性的进化可能有因果相关性，而是否认其因果作用是积极的，也即否认了类型因果的存在。

我们以一个例子说明高亲缘系数可能具有的消解性作用。拉特尼克斯（Francis L. W. Ratnieks）与同事的一系列工作（Ratnieks, 1988; Ratnieks & Visscher, 1989）表明，蜂群得以维持真社会性的一个重要原因是工蜂的监督行为（worker policing），即若蜂群中偶有工蜂产出雄性卵，将会被其它工蜂破坏。直觉上，我们似乎觉得工蜂擅自产卵这一自私行为将在蜂群内亲缘系数过低的情形下发生，因此高亲缘系数的环境将有效抑制此类行为。然而，理论推导和实验证据则告诉我们相反的结论，即只有亲缘系数足够低时，工蜂才有动力破坏其它工蜂的卵；若亲缘系数过高，则工蜂的监督行为将失效，使得工蜂更多地擅自产卵而瓦解蜂群的真社会性。

类似的例子当然为亲缘系数重要性的支持者所承认（Foster et al., 2006; Queller, 2016）。然而，这对于高亲缘系数的类型原因地位是很大的打击。对此，支持者的一种回应策略是，强调亲缘系数作用的两面性（Foster et al., 2006）。^③在真社会性

① 为了方便表述，我在此约定将“因果相关性”理解为有方向的概念。当我说亲缘系数与真社会性的进化因果相关时，都是指亲缘系数变量是真社会性进化变量的原因，而非反之。

② 盖勒所说的“影响”仍是一种因果概念，且与刘易斯（David Lewis）等人所说的“影响”含义不同。

③ 但这不是唯一的回应策略。

进化的案例中，一方面，蜂群中的亲属关系使得亲缘系数大于0，因而能促进利他主义和合作行为的进化；另一方面，亲属毕竟不是自身的克隆，亲缘系数总是小于1，因而也将导致亲属间的竞争和矛盾。由此，真社会性中体现的任何合作和竞争要素都总是与亲缘系数大小相关。这样的解释本身并没有错，但相较于强调高亲缘系数的重要地位是一个明显更弱的主张。支持者如想辩护高亲缘系数是真社会性进化的类型原因，就必须诉诸对其因果角色更细致的刻画，其具体方案将在第3.5.2节讨论。

许多批评者甚至怀疑亲缘系数与真社会性的进化之间根本没有因果相关性（Wilson, 2008; Nowak, Tarnita et al., 2010）。为建立亲缘系数和真社会性的进化之间的因果关系，在主流干预主义因果理论（Woodward, 2003）的视角下，我们必须设法改变亲缘系数的大小，进而考察真社会性的进化条件是否发生了改变。在真社会性昆虫的场景下，对蜂群内亲缘关系大小最直接的影响源于蜂后的配偶数量。蜂后交配的不同雄蜂数量越多，蜂群内姐妹工蜂之间的亲缘关系就越远，同时真社会性进化的条件也会更严格。休斯等人（Hughes et al., 2008）据此认为高亲缘系数更有利于真社会性进化。问题在于，蜂后的配偶数量能否作为亲缘系数的合法的干预变量呢？这一点是可疑的，因为蜂后的配偶数量可能不经由亲缘系数而通过其它路径影响真社会性的进化。例如，诺瓦克等人（Nowak, Tarnita et al., 2010）提出，是否有可能蜂后多次外出交配会增加蜂群遭受捕猎者袭击的风险，从而对真社会性蜂群的进化造成负面影响？这一设想如果成立，那么高亲缘系数就只是真社会性进化中选择较少的蜂后配偶数量的副产品，也就意味着高亲缘系数和真社会性进化间仅有统计相关性而缺乏因果关系。

3.5.2 净效果与组分效果

如果我们尚且承认亲缘系数和真社会性的进化之间是因果相关的，那么我们或许会认为，亲缘系数在真社会性进化的某些方面可能起到消解性的作用，但对特定明确的利他行为进化而言其作用总是正面的。这一点也可以从汉密尔顿规则看出：利他行为进化所要求的条件是 $RB - C$ 足够大（至少大于0），那么只要 B 是正数，亲缘系数 R 就一定是越大越好。汉密尔顿自己在解释真社会性的进化时就说，由于姐妹工蜂之间的亲缘系数高于工蜂与其女儿的亲缘系数，“在其它情况均同时，它将偏好回到其母亲的巢穴供养一个姐妹工蜂而非独自供养其女儿。”

（Hamilton, 1964）²⁹ 这里的“其它情况均同”条件值得我们注意。从因果的角度看，“其它情况均同”条件要求我们固定亲缘系数和真社会性进化之间可能具有的其它因果路径不变，而只考察对一个特定行为的适合度效果而言亲缘系数的因果

贡献是否是正面的。为此，我将引入净效果和组分效果的区分，进一步澄清亲缘系数在真社会性进化中可能具有的因果角色。

考虑赫斯洛夫 (Hesslow, 1976) 提供的一个经典案例。一种避孕药具有引发血栓的副作用。然而，怀孕同样会引发血栓，服用避孕药则能避免怀孕从而避免血栓。这样，服用避孕药和得血栓之间就有两条效果相反的因果路径。通过构造适当的参数，可以使其中任一路径的效果占据主导地位，也可以使两条路径的效果相互抵消。希契科克 (Hitchcock, 2001) 由此将类型因果具有的效果分为两类：一类是通过所有因果路径共同导致的效果，即净效果；另一类是固定其它路径时通过某条单一路径造成的效果，即组分效果。后者也被称为路径特异效果 (path-specific effect) (Weinberger, 2019)。在上述案例中，避孕药对血栓在一条路径上有正面的组分效果，而在另一条路径上有负面的组分效果；避孕药的净效果则可正可负，可见其与组分效果没有必然联系^①。

上述案例启示我们，亲缘系数重要性的支持者所想要刻画的或许是亲缘系数的某种组分效果。许多支持者强调，高亲缘系数并非亲缘选择理论的全部，亲缘选择的理论家同样关心利他行为收益和代价的比值 (即 B/C) 对其进化的作用 (Foster et al., 2006; Queller, 2016)，而对于特定行为而言，亲缘系数的贡献总是积极的 (Queller, 2016)。问题在于，在具体的经验情形下，行为的收益和代价是非常难以计算的 (Bourke, 2011; Queller, 2016)，甚至在复杂情形下两者都会成为关于亲缘系数的函数 (即成为 $B(R)$ 、 $C(R)$) (Gardner, West & Barton, 2007; Nowak, Tarnita et al., 2010)。进而，我们可以设想这样的情况，尽管 R 对真社会性进化的影响是正向的，但由于 B/C 关于 R 的函数 (在某些区间上) 递减，使得 R 经由 B/C 对真社会性进化还有负面影响，由此 R 对真社会性的进化可能起到负面的净效果。不过，只要我们固定 B/C 不变，就可以看出 R 对真社会性进化的组分效果总是正向的。在这种刻画下，如果支持者主张亲缘系数总是具有正面的组分效果，而批评者主张亲缘系数可能具有负面的净效果，两方其实是完全兼容的。

不过，亲缘系数的组分效果究竟为何并不清晰，这是因为画出可供分析 R 的组分效果的因果图本身就面临着困难。我们知道，亲缘系数无非是两个个体基因型之间的某种关系。如果我们将两个个体的基因型和它们的亲缘系数全部置入因果图中，就很难设想对亲缘系数的干预的实际意义。如果我们严格采用固定两者基因型的硬干预来阻断亲缘系数到真社会性进化之间非直接的因果路径，那就意味着我们在没有改变两者基因型的条件下改变了通常由两者基因型直接决定的它们的关

^① 温伯格 (Weinberger, 2019) 在路径特异效果和总体效果 (total effect, 大致对应于希契科克所说的净效果) 之间建立了联系。

系，其实际意义并不清楚。对这一困难的一个侧面佐证是，奥卡沙（Okasha, 2016b）在画社会进化的因果图时，将亲缘系数置于了因果图外，而没有放进因果图中作为一个变量。这也许就暗示着画出包含亲缘系数的因果图面临的困难。这样看来，亲缘系数对真社会性进化具有正面的组分效果尽管看起来显然，但也仍需进一步的工作来刻画其确切含义。

3.6 本章小结

亲缘系数可能具有 IBD 亲缘系数、IBS 亲缘系数和统计亲缘系数等不同的含义。真社会性的进化解释争议的核心是 IBD 亲缘系数在真社会性的进化中是否具有因果效力，这一问题与关于自然选择的统计主义无关。针对真社会性的进化这一经验案例，争议双方对 IBD 亲缘系数的强调直接相关于亲缘选择在解释真社会性的进化中的理论地位。不过，广义适合度理论特别是其一般版本并不处在争议的中心，其反映是 HRG 中的基础概念统计亲缘系数并未得到与经验案例结合的充分讨论。考虑到从广义适合度理论的角度看 IBD 亲缘系数在广义适合度计算应用时极其受限，即便支持者辩护了 IBD 亲缘系数在真社会性进化中的重要因果地位，可能也只能辩护亲缘选择这一机制在真社会性进化中的重要作用，而并不能直接说明广义适合度理论和真社会性的进化之间有可靠的解释或确证关系。

就 IBD 亲缘系数在真社会性的进化中的因果角色这一具体问题而言，过往的研究只概括出争议双方的共识是 IBD 亲缘系数是真社会性进化的必要不充分条件。我进一步分析并展示了，支持者有时将对亲缘系数因果角色的辩护停留在因果相关性上，而批评者则明确关心类型因果；支持者更强调亲缘系数的组分效果，批评者则关心亲缘系数的净效果，至于如何画出足以明确刻画两者的包含亲缘系数变量的因果图还需进一步研究。

第4章 广义适合度最大化、最优化与能动者思维

4.1 引言

汉密尔顿在如下段落中阐释了它的广义适合度概念：

广义适合度可以被想像为个人适合度在首先被去除所有可以被认为由于某个体社会环境造成的组分而保留其将在不暴露于该环境中的任何伤害或好处时表现的适合度，再加上其引起的其邻居适合度的伤害和好处的特定比例。这里的比例就是该个体与其影响的邻居间合适的亲缘系数。(Hamilton, 1964)⁸

直观上，广义适合度概念将个体行为对其搭档适合度的影响纳入考量，进而利他行为可能以个人适合度为代价而提高行为主体的广义适合度，使得该行为在进化中占优势。在此，我想首先澄清汉密尔顿文本中出现的三个不同的概念。**广义适合度效果** (inclusive fitness effect) 指某社会行为对个体适合度的增量或减量。**广义适合度** 指经过广义适合度效果修正的个体适合度。通常我们仅单独考虑某一社会行为的进化，这时我们常常假定除该社会行为的进化外种群已处于稳定的群体遗传学平衡，即将每个个体的基准适合度设为1，进而每个个体的广义适合度就是1加上广义适合度效果。需要特别注意的是，广义适合度效果和广义适合度都是定义在个体层面的量，即我们通常只能谈某个体的广义适合度。在群体层面，我们可以谈论的是**平均广义适合度** (mean inclusive fitness)，即该种群中每一个体广义适合度的算术平均值。严格来说，根据汉密尔顿自己的说法，在进化过程中趋于最大化的量是种群的平均广义适合度，而不是某一个体的广义适合度。

尽管汉密尔顿自己的原始模型需要较严格的限制条件才能成立，许多学者相信广义适合度是理解社会进化的关键概念，并大力发展了相关理论。问题在于，在有关广义适合度理论的研究中，人们常常较为随意地谈论广义适合度最大化而没有提供对最大化的目标及含义的充分澄清。当然，有一部分学者意识到了这一问题，并致力于让对广义适合度最大化的讨论精致化、严格化 (Grafen, 2006b; Gardner, West & Wild, 2011; Okasha & Martens, 2016)。

奥卡沙是讨论广义适合度最大化议题的一位重要哲学家。奥卡沙和马滕斯 (Okasha & Martens, 2016) 论证了在支付可加的情形下广义适合度是最大化的，

但在支付不可加情形下广义适合度最大化不能由汉密尔顿规则的一般形式得到辩护，而他们所称的“格拉芬 1979 支付”^①至少可以在他们讨论的简单的支付不可加的模型中作为最大化的目标。奥卡沙（Okasha, 2018）的独著《进化中的主体和目标》（*Agents and Goals in Evolution*）收录了他和马滕斯合作的上述论文的修订版本，并将关于广义适合度最大化的讨论和能动者思维联系了起来。在本书中，奥卡沙区分了进化生物学中两种类型的能动者思维。根据他的定义，第一类能动者思维指将有机体看作好似在如此行为以尽可能达到某种目标的主体；第二类能动者思维指将自然选择看作驱使有机体在进化过程中最优化的主体。奥卡沙进而将广义适合度最大化与第一类能动者思维联系了起来，而认为其与第二类能动者思维无关。

我认为奥卡沙对广义适合度最大化的分析是成问题的，特别地，他在讨论广义适合度最大化问题时对能动者思维概念的应用与他对能动者思维概念的定义及在其它地方的应用相矛盾。本章将反驳奥卡沙对广义适合度最大化的分析并试图建立从能动者思维视角理解广义适合度最大化的合适方式。在第 4.2 节，我将讨论支付可加情形下的广义适合度最大化。尽管支付可加的情形较为简单且并不涉及很多争议，我认为有必要在此时澄清一些常见的混淆，以辅助后续讨论更复杂的情形。在第 4.3 节，我将评估奥卡沙对支付不可加情形下广义适合度最大化的相关论述。我将回应他对朴素意义上广义适合度理论最大化的两点反对意见，并尝试解释为何格拉芬 1979 支付在奥卡沙的工作中看起来很重要。第 4.4 节是本章小结。

4.2 支付可加情形下的广义适合度最大化

支付可加情形指每一个体受到的支付可以全部直接相加得到总体的支付值，其更准确的刻画将在后文的模型中给出。我在此径直采用奥卡沙（Okasha, 2018）书中的模型。考虑一个包含无穷多单倍体无性繁殖有机体的种群。这些有机体可以被分为两种由基因决定且完美遗传的类型：利他个体（A）和自私个体（S）。这些有机体的互动是成对进行的，其中利他个体将以自身收益为代价使其搭档受益，而自私个体仅仅接受其它个体可能赋予的收益而不付出任何代价。对处于某种关系中的一对个体 i 和 j ，令 $V(i, j)$ 表示 i 在这部分关系中的收益，即行动者的支付或个人支付。 i 和 j 的取值为 A 或 S，这表示它们的类型。表 4.1 展示了一组社会行为的支付矩阵，其中 B 和 C 为正。注意到，左上角的 $(B - C)$ 显示了收益 B 和代价 $(-C)$ 是可加的。由于 $B > B - C$ 且 $0 > -C$ ，如果将个体支付视为进化博弈的效用，则 A

① 详见第 46 页开始的讨论。

型个体永远无利可图。为了正确计算施行利他行为可能具有的效用，我们应该考虑搭档的利益。

表 4.1 可加情形的个体支付 (see Okasha, 2018) ¹²²

$V(i, j)$		搭档 (j)	
		A	S
行动者 (i)	A	$B - C$	$-C$
	S	B	0

现在我们引入广义适合度 $U(i, j)$ 并将其定义为 $V(i, j) + rV(j, i)$ ，其中 r 是两者的亲缘系数。直观上，由于搭档收益的效用低于行为者自身的直接利益，在将搭档收益算作行动者自身收益时应乘以 r 作为折扣。本章中我不再对亲缘系数做详细讨论，而只引入一个合理的假设，即我们的模型中 r 是进化过程中的常数。广义适合度 $U(i, j)$ 也即该模型中行为的效用。结果如表 4.2 所示。需要注意的是，该定义由汉密尔顿 (Hamilton, 1971) 提供，而又与汉密尔顿 (Hamilton, 1964) 的原始定义不同。之所以使用这个定义，主要是因为它与个人支付直接相关而容易直接计算。这两个定义之间的一个重要区别是，这里定义的广义适合度并不完全取决于行动者的类型。也就是说，为了最大化其效用，行动者应该考虑其搭档的类型。

表 4.2 可加情形的广义适合度或效用 (see Okasha, 2018) ¹²⁴

$U(i, j)$		搭档 (j)	
		A	S
行动者 (i)	A	$(B - C)(r + 1)$	$-C + rB$
	S	$B - rC$	0

表 4.3 效用取广义适合度时的进化-理性链接 (Okasha, 2018) ¹²⁶

$rB > C \Leftrightarrow$ A 进化	\Leftrightarrow	(A, A) 是唯一的纳什均衡
$rB < C \Leftrightarrow$ S 进化	\Leftrightarrow	(S, S) 是唯一的纳什均衡
$rB = C \Leftrightarrow$ 无进化	\Leftrightarrow	所有纯粹或混合策略都是纳什均衡

确定效用函数后，我们可以很容易地找到博弈的纳什均衡。我在此略去技术细节，直接给出从我们模型中得到的表 4.3。从进化的角度来看，我们可以得到著名的汉密尔顿规则，即 $rB > C$ ，它表达了利他行为在进化中占优势的条件。可以肯定

的是，它的证明需要支付可加假设。奥卡沙似乎认为进化与理性的联系很重要。他这样解释这个想法：

标准博弈论预测，如果玩家是理性的，他们最终将处于博弈的纳什均衡，在这个均衡中，每个玩家都以对方的行为为条件最大化自己的效用。然后，我们可以问，理性博弈的纳什均衡是否与上述进化过程的结果相对应。如果是这样，我们可以得出结论，进化将导致有机体表现得好像试图最大化相应的效用函数。（Okasha, 2018）¹²⁴

我认为奥卡沙在这里的表述是误导性的。纳什均衡所描述的只是博弈的结果，而不涉及向结果演变的过程。诚然，在纳什均衡下，每个参与者的效用最大化都取决于搭档的行为。然而，“搭档的行为”在这里具有特定的含义，即它只是“搭档在纳什均衡下的行为”。此外，考虑我们假定的在博弈中要求最大化的效用函数。它在其定义域内最多只包含三个不同的值，其自变量是参与者和合作伙伴的类型，其可能的取值也是有限的。这一观察部分是出于我们模型的简单性，但它至少表明效用函数是一个离散函数，这意味着我们无法找到一个持续最大化的过程，我们也不能在博弈不处于纳什均衡时谈论任何关于最大化的问题。一旦确定了行动者及其搭档，效用函数的值也就确定了，那么就没有最大化的余地了。此外，我想指出，从这个意义上说，最大化偏离了我们通常的理解。回想汉密尔顿（Hamilton, 1964）的表述，其中被最大化的是平均广义适合度。它是在种群而不是个体上定义的量，因此它可以随时间或代际而变化。在个体层面上，我们只能说它们的广义适应度在进化完成时通过自然选择最大化了。也许还可以补充的是，他们的基因型必须具有一定的优势才能实现最大化。总之，个体不能成为最大化或最大化行为的主体。

有机体不能成为最大化主体并不能排除将其视为主体的可能性，实际上，根据奥卡沙（Okasha, 2018）的定义，在这里应用第一类能动者思维是可行的。他为第一类能动者思维之成立提出了三个独立的理由：目标导向性、行为灵活性和拥有适应性特征。在我们的案例中，行为灵活性是不成立的，因为社会行为是由遗传完全决定的；拥有适应性特征，虽然与适应度最大化有关，但只在进化结束时成立（并且确实成立）——这似乎是奥卡沙的分析实际上所暗示的，也是奥卡沙所关心的；然而，我们似乎没有任何理由拒绝在整个进化过程中始终满足的目标导向性。奥卡沙（Okasha, 2018）²⁴ 承认：“这一‘目标’概念与一些进化讨论中的概念不同，在一些进化讨论中，生物体的‘目标’被假定为最大限度地提高适应性。”

当然，目标导向性和第一类能动者思维一般要求一个限制，即在生物体中没有任何遗传冲突或拮抗作用，但这种目标统一性的限定仍然不需要任何形式的最大化。作为证据，可以考虑奥卡沙（Okasha, 2018）对格拉芬形式达尔文主义项目的评判。格拉芬（Grafen, 2002; 2014）建立了一个选择-最优性链接，这与前面讨论的奥卡沙的进化-理性链接非常相似。只要格拉芬的链接成立，目标统一性就成立的原因很简单——格拉芬使用普莱斯方程进行分析：

$$\Delta p = \text{cov}\left(p_i, \frac{w_i}{w}\right) + E\left(\frac{w_i}{w} \Delta p_i\right). \quad (4.1)$$

根据奥卡沙的说法，只有当每个基因的普莱斯方程的期望项为零时，格拉芬的链接才成立，这意味着导致减数分裂驱动等遗传冲突的因素不存在，因此目标统一性得到满足。现在回到我们讨论广义适合度的案例。奥卡沙（Okasha, 2018）还使用如下简化形式的普莱斯方程来推导出确切的进化过程：

$$\Delta p = \frac{\text{cov}(w_i, p_i)}{w}. \quad (4.2)$$

其中也没有包括数学期望项。因而在任何进一步的联系建立之前，它应当能以同样的方式证明目标统一性是合理的。因此，在我们的案例中，在整个进化过程中应用目标导向意义上的第一类能动者思维是合理的。

4.3 支付不可加情形下的广义适合度最大化

表 4.4 不可加情形的个体支付（see Okasha, 2018）¹²⁷

$V(i, j)$		搭档 (j)	
		A	S
行动者 (i)	A	$B - C + D$	$-C$
	S	B	0

我遵循奥卡沙（Okasha, 2018）的方法，将以前的模型调整为支付不可加案例。表 4.4 列出了经修订后的个体支付。左上角的 D 项表示存在协同效应，两个 A 型个体相互作用时其支付不能通过直接叠加支付 B 和 $-C$ 得到。在这种情况下， $rB - C > 0$ 不再是 A 型较 S 型在进化中占优的条件，因而许多学者指出汉密尔顿定律在非加性情况下是失败的（e.g. Nowak, Tarnia et al., 2010）。加德纳等人（Gardner, West & Wild, 2011）认为对汉密尔顿规则的上述批评是不恰当的，并主张我们可以推广汉密尔顿规则来解决这个问题。根据加德纳等人（Gardner, West & Barton, 2007）的说法，我们可以引入两个新参数 b 和 c ，如下所示：

$$b = B + \frac{r + (1-r)p}{1+r} \cdot D; \quad (4.3)$$

$$c = C - \frac{r + (1-r)p}{1+r} \cdot D。 \quad (4.4)$$

其中 p 表示A型个体在群体中的频率。进而由新定义的 b 和 c 表示的 $rb - c > 0$ 成为一般汉密尔顿规则^①，作为A型个体频率 p 随进化过程增大的条件。我在此略去技术细节，而只提及奥卡沙也遵循了这种方法，即引入 b 和 c 并应用一般汉密尔顿规则，以推导出支付不可加情形下的确切进化过程。从他们对普莱斯方程的使用中，我们可以注意到，加德纳等人（Gardner, West & Barton, 2007）和奥卡沙（Okasha, 2018）都没有考虑传递偏差项，这应该立即意味着遗传冲突已被排除，并支持在我们的案例中，在目标导向的意义上应用第一类能动者思维。

现在回到奥卡沙（Okasha, 2018）的工作。他对此种情形下的最优化提出了两点意见以反对广义适合度最大化在某些“幼稚”的意义上成立。第一点反对意见针对使用一般汉密尔顿规则来推导效用函数，进而论证某种最大化的合理性。奥卡沙和马滕斯（Okasha & Martens, 2016）认为效用函数应是关于 B 、 C 、 D 和 r 的函数，因而包含基因频率 p 作为参数的 b 和 c 不应出现在效用函数中。但提出这一要求的缘由似乎并不清楚。实际上， B 、 C 、 D 和 r 都是我们模型中的常数，因此如果效用函数不包含任何其他变量，它将是一个常数。我们绝不能谈论常数的最大化——至少应将表型变量包含在效用函数中，以实现可能的最大化。回想我们在第4.2节中对效用函数最大化的澄清，我们会发现效用函数最大化只是为了实现最佳表型以获得相应的收益，而这种陈述仅限于进化的结果。如果我们只考虑进化完成的时间点，那么 p 也是一个常数。在支付不可加情形下，最佳表型和最大收益之间的联系与在支付可加情形下的形式相同。加德纳等人（Gardner, West & Wild, 2011）向前迈出了一步。他们认为行为的代价和收益完全可以与频率有关，并再次引用了加德纳等人（Gardner, West & Barton, 2007）的工作作为证据。这个建议听起来很合理，因为理性主体在做出选择时考虑一些与频率相关的因素是很正常的，如果是这样，那么即使 p 确实充当效用函数中的变量，也可以谈论效用最大化。无论如何，我并不理解为什么奥卡沙不允许被最大化的量中包含 p 。

奥卡沙对广义适合度最大化的第二点反对意见是相当平凡的。这一点意见只是说，如果我们以与支付可加情形相同的方式将效用函数定义为广义适合度（见表4.5），博弈论分析得到的均衡结果将与进化分析得到的进化结果不匹配。这种不匹配实际上不足为奇。有趣的是，奥卡沙转而将效用函数定义为所谓“格拉芬 1979

① 这一规则也就是第2章中所说的HRG在两两结对互动的社会进化情形下的特殊形式。

支付”，并发现进化-理性链接竟然成立了。格拉芬 1979 支付的定义为 $U(i, j) = rV(i, i) + (1 - r)V(i, j)$ 。采用该定义的结果如表 4.6 和表 4.7 所示。

表 4.5 不可加情形将广义适合度作为效用 (see Okasha, 2018) ¹³¹

$U(i, j)$		搭档 (j)	
		A	S
行动者 (i)	A	$(B - C + D)(r + 1)$	$-C + rB$
	S	$B - rC$	0

表 4.6 不可加情形将格拉芬 1979 支付作为效用 (see Okasha, 2018) ¹³¹

$U(i, j)$		搭档 (j)	
		A	S
行动者 (i)	A	$B - C + D$	$-C + rB + rD$
	S	$(1 - r)B$	0

表 4.7 效用取格拉芬 1979 支付时的进化-理性链接 (Okasha, 2018) ¹³²

(A, A)是唯一的纳什均衡	\Rightarrow	A 进化至固定
(S, S)是唯一的纳什均衡	\Rightarrow	S 进化至固定
(π_A, π_A) 是唯一的纳什均衡	\Leftrightarrow	在 $p = \pi_A$ 处出现稳定的多态
(π_A, π_A) 、(A, A)和(S, S)都是纳什均衡	\Leftrightarrow	在 $p = \pi_A$ 处出现不稳定的多态

$$\text{其中 } \pi_A = \frac{C - r(B + D)}{(1 - r)D}。$$

首先应该注意的是，格拉芬 1979 支付不是广义适合度的任何既定定义 (Allen & Nowak, 2016)，因此广义适合度最大化在此仍然是无关的。

下面考虑采用格拉芬 1979 支付的好处。一个明显的要点是这样会使效用函数不再包含 p ，因此效用的四个可能值在整个进化过程中是固定的。这一点重要吗？我对此持怀疑态度。奥卡沙可能认为格拉芬 1979 支付是一个固定的目标，这才使得以该固定目标为对象的最大化得以可能。然而，如果我们只考虑进化结果，这一点是平凡的，适应意义上的第一类能动者思维成立，但在这种情况下，“固定”听起来毫无意义，因为 p 仅是一个常数。

当然，奥卡沙可能意图把这一观察扩展到整个进化过程，但这样一来目标其实并不是固定的。这里的问题在于效用函数的不可控性。如果个体的目标是固定的，那么最大化效用所需的（理想）搭档类型也应该为参与者固定。为此，博弈的纳什均衡必须是唯一的，否则行动者将不确定它的搭档属于哪种类型，进而它并不会有一个固定的目标。然而，进化与理性的链接并不完美。前两条线的箭头是单向的，当 (π_A, π_A) 、 (A, A) 和 (S, S) 都是纳什均衡时， $p = \pi_A$ 并非唯一可能的进化终点——当 $p = \pi_A$ 是不稳定的多态时， A 和 S 仍然可能进化到固定状态，具体进化结果取决于 p 的取值。进而，我们有两种解释选择：一种是像奥卡沙那样拒绝将 p 作为被最大化的量中的参数，那么当 (π_A, π_A) 、 (A, A) 和 (S, S) 都是纳什均衡时，就不会有固定的目标；另一种是允许 p 在效用函数中出现，但即便如此进化-理性链接中的第4条链接仍然不足以确定一个固定的目标，我们还需要进一步比较 p 与 π_A 的大小，以确定究竟哪个目标实际起效。

接下来，我将从能动者思维的角度来评估以上内容。回想我在第4.2节中的分析，我们会发现，在支付不可加的情形下，第一类能动者思维的应用较支付可加情形下的应用几乎没有什么不同：目标导向性当然成立；行为灵活性当然不成立；适应成立且仅在进化完成时成立。在我看来，奥卡沙（Okasha, 2018）对支付不可加的情形下最优化的繁琐讨论实际上与第二类能动者思维（而非第一类能动者思维）有关，尽管他似乎没有正确意识到这一点。在支付可加的情形下，最大效用是完全固定的，因此自然选择可以被视为驱动群体向预定的最优状态进化，这一过程中群体的平均广义适合度趋于最大化。在支付不可加的情形下，采用格拉芬1979支付所提供的正是这类趋于预定目标的第二类能动者思维，尽管应用范围限于存在唯一纳什均衡的情形。

阿伦和诺瓦克（Allen & Nowak, 2016）¹²⁶在广义适合度理论之争中提出了两个共识，不妨在这里讨论：“对于一般的进化过程，（i）不存在个体层面的广义适合度，以及（ii）自然选择不会驱使个体表现得好像最大化其广义适合度。”在（i）中，他们主要指的是广义适合度更常用的简化版本（也就是本节采用的版本）。他们为（i）提供的理由主要是协同效应，这是支付不可加情形所独有的。在支付可加的情形下，至少在我们的模型中，广义适合度的简化版本等价于汉密尔顿（1964）的原始定义，因此我们可以使用完全在行为者控制下并且完全合理的原始广义适合度来导出相同的广义适合度最大化结果。这一点从第一类能动者思维的视角看也就是将个体视为适应的个体，并将其视为在进化结束时实现了广义适合度最大化。而从第二类能动者思维的角度，我们也可以合理地将自然选择视为驱动群体趋于平均广义适合度最大化。然而，在支付不可加的情形下，广义适合度的原始版本

和简化版本是不一致的。由于广义适合度的简化版本在这里并无意义，尽管我们仍然可以谈论进化结束时个体的适应，但适应所指的不再是最大化的广义适合度。尽管在格拉芬 1979 支付的帮助下，我们仍然可以将自然选择视为最大化的驱动者，但最大化的对象不再是广义适合度——这也是阿伦和诺瓦克所说的（ii）的主要含义。^①

4.4 本章小结

从第一类能动者思维的角度来看，只要没有遗传冲突，目标导向就成立；只要相关表型是由遗传决定的，行为灵活性就不成立。这两点与广义适合度或其最大化无关。个体仅在进化结束时才具有适应性状，在支付可加情形下，这些性状实现了最大的广义适合度，而在支付不可加情形下，这些性状与广义适合度无关。当存在单一纳什均衡时，它们反而实现了格拉芬 1979 支付。

从第二类能动者思维的角度来看，自然选择在支付可加情形下驱动整个种群达到最大化，其中最大化的量是平均广义适合度。在支付不可加情形下，自然选择至少在存在单一纳什均衡时会将整个种群推向最大化，最大化的量是格拉芬 1979 支付。

^① 阿伦和诺瓦克当然不熟悉奥卡沙定义的第一类能动者思维和第二类能动者思维这两个概念。是我采用了这两个概念解读他们的观点。

第5章 结论

本文从哲学角度研究了广义适合度理论之争的三个方面。第一，从广义适合度理论自身的角度看，汉密尔顿规则的解释价值是可疑的。伯奇（Birch, 2017b）对一般汉密尔顿规则作为社会进化研究的组织框架的辩护是成问题的，我通过考虑具有显著协同效应的社会进化案例进行了反驳。进而，我引入了一种新的方案来理解多种一般形式的汉密尔顿规则的解释价值，这种新的方案有助于我们揭示具体社会进化案例之间可能具有的不同维度的共同点和差异。这种新的方案如何精确刻画有待于进一步的研究。此外，本文对汉密尔顿规则的讨论没有深入分析 HRA，而 HRA 尽管哲学讨论不多，却是进化生物学研究实践中最常用形式的汉密尔顿规则，因而本文对汉密尔顿规则解释价值的讨论仍有一定局限性。

第二，争议双方在关于广义适合度理论或亲缘选择理论与真社会性的进化之间的关系的争论中的一些观点在很大程度上包含着对亲缘系数和因果概念的混淆，极大地阻碍了我们对相关争议的充分理解。通过对 IBD 亲缘系数、IBS 亲缘系数和统计亲缘系数的区分，我们可以发现争议中最重要的争议点在于 IBD 亲缘系数在真社会性进化中的地位。本文较现有的工作（Pernu & Helanterä, 2019a; 2019b）更细致地刻画出了批评者的意见。至于争议双方的观点在多大程度上具有合理性，仍有待进一步的研究。

第三，广义适合度最大化在支付不可加的情形下难以成立，但广义适合度最大化成立与否与能动者思维成立与否没有直接联系。奥卡沙（Okasha, 2018）在分析相关问题时混淆了他自己定义的第一类能动者思维和第二类能动者思维。我的重新评估发现，第一类能动者思维关涉的目标导向性和行为灵活性与广义适合度最大化无关，适应仅在支付可加情形下能得到广义适合度最大化的支持。第二类能动者思维在支付可加情形下能够成立，而在支付不可加情形下需要一定限制条件。整体上看，广义适合度最大化和能动者思维并无紧密联系。其哲学意义如与适应主义、有机体设计等议题的关联有待进一步研究。

参考文献

- 贾栗, 2013. 达尔文难题及其解决路径[J]. 河北学刊, 33(02): 222-225.
- 贾栗, 2013. 浅议利他行为的解释理论及其分类[J]. 广东医学院学报, 31(06): 657-660.
- 金萍, 颜忠诚, 2013. 动物的亲缘选择[J]. 生物学通报, 48(12): 5-8.
- 李建会, 2009. 自然选择的单位: 个体、群体还是基因? [J]. 科学文化评论, 6(06): 19-29.
- 李建会, 项晓乐, 2009. 超越自我利益: 达尔文的“利他难题”及其解决[J]. 自然辩证法研究, 25(09): 1-7.
- 李洁, 孙庚, 胡霞等, 2014. 植物的亲缘选择[J]. 生态学报, 34(14): 3827-3838.
- 刘鹤玲, 2008. 亲缘选择理论: 生物有机体的亲缘利他行为及其基因机制[J]. 华中师范大学学报(自然科学版), 125(01): 114-118.
- 刘鹤玲, 蒋湘岳, 刘奇, 2007. 广义适合度与亲缘选择学说: 亲缘利他行为及其进化机制[J]. 科学技术与辩证法, 134(05): 26-29.
- 刘华杰, 2022. E.O.威尔逊的“博物学家”头衔[J]. 生物多样性, 30(01): 19-24.
- 刘奇, 2006. 生物有机体的亲缘利他行为及其进化机制研究[D]. 华中师范大学.
- 罗力群, 2016. 汉密尔顿及其广义适合度理论[J]. 自然辩证法研究, 32(02): 94-97.
- 罗力群, 2018. 罗伯特·特里弗斯: 桀骜不驯的卓越进化理论家[J]. 自然辩证法通讯, 40(05): 132-141.
- 罗力群, 2019. 人类行为和文化的演化研究[J]. 科学文化评论, 16(03): 29-42.
- 罗力群, 2020. 生物社会视角下的民族问题研究引论[J]. 社会科学论坛, 328(04): 187-194.
- 王健, 2009. 道德的经济学——亲缘选择与广义效用理论[J]. 晋阳学刊, 173(02): 62-66.
- 吴林林, 郭松涛, 齐晓光等, 2013. 非人灵长类的亲缘选择[J]. 兽类学报, 33(03): 267-273.
- 肖凤良, 陈晓平, 2012. 论社会生物学的亲缘选择理论及其学术评价[J]. 求索, 240(10): 80-82.
- 肖凤良, 陈晓平, 2013. 基因选择理论及其理论困境[J]. 科学技术哲学研究, 30(03): 45-48.
- 徐亮, 2022. 进化论下的生物个体性哲学研究[D]. 华中科技大学.
- 杨阳, 2020. 生物利他行为的演化基础及其对人类道德的意义[D]. 南京师范大学.
- 杨月, 龚钢明, 郑雪松, 2015. 基于微生物模型的合作行为进化研究进展[J]. 微生物学通报, 42(04): 749-757.
- 于书, 2019. 爱德华·威尔逊的基因—文化协同进化理论研究[D]. 中南财经政法大学.
- 于书, 2020. 社会生物学对广义适合度理论的批判和超越[J]. 当代经济, 513(09): 96-100.

- 赵斌, 2012. 生物学中的功能定义问题研究[J]. 山西大学学报(哲学社会科学版), 35(06): 8–13.
- 赵斌, 郭虹, 2019. 利他性进化理论中的模型多元论[J]. 东北农业大学学报(社会科学版), 17(01): 70–76.
- 邹笑笑, 2012. 群体选择理论的科学史考察[J]. 华中农业大学学报(社会科学版), 101(05): 108–112.
- Ågren J A, 2021. The gene's-eye view of evolution[M]. Oxford: Oxford University Press.
- Abbot P, Abe J I, Alcock J, et al., 2011. Inclusive fitness theory and eusociality[J]. *Nature*, 471(7339): E1–E4.
- Allen B L, Nowak M A, 2016. There is no inclusive fitness at the level of the individual[J]. *Current Opinion in Behavioral Sciences*, 12: 122–128.
- Allen B L, Nowak M A, Wilson E O, 2013. Limitations of inclusive fitness[J]. *Proceedings of the National Academy of Sciences*, 110(50): 20135–20139.
- Andersson M, 1984. The evolution of eusociality[J]. *Annual Review of Ecology and Systematics*, 15(1): 165–189.
- Birch J, 2014a. Gene mobility and the concept of relatedness[J]. *Biology & Philosophy*, 29(4): 445–476.
- Birch J, 2014b. Hamilton's Rule and Its Discontents[J]. *The British Journal for the Philosophy of Science*, 65(2): 381–411.
- Birch J, 2014c. Has Grafen formalized Darwin?[J]. *Biology and Philosophy*, 29(2), 175–180.
- Birch J, 2016. Natural selection and the maximization of fitness[J]. *Biological Reviews*, 91(3): 712–727.
- Birch J, 2017a. The inclusive fitness controversy: finding a way forward[J]. *Royal Society Open Science*, 4(7): 170335.
- Birch J, 2017b. The philosophy of social evolution[M]. Oxford: Oxford University Press.
- Birch J, 2019. Inclusive fitness as a criterion for improvement[J]. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 76: 101186.
- Birch J, Okasha S, 2015. Kin selection and its critics[J]. *BioScience*, 65(1): 22–32.
- Bourke A F G, 2011. Principles of social evolution[M]. New York: Oxford University Press.
- Bourke A F G, 2014. The gene's-eye view, major transitions and the formal darwinism project[J]. *Biology and Philosophy*, 29(2): 241–248.
- Bourke A F G, 2015. Sex investment ratios in eusocial Hymenoptera support inclusive fitness theory[J]. *Journal of Evolutionary Biology*, 28(11): 2106–2111.
- Bourke A F G, 2019. Inclusive fitness and the major transitions in evolution[J]. *Current Opinion in Insect Science*.
- Darwin C, 1859. On the origin of species[M]. London: John Murray. http://darwin-online.org.uk/converted/pdf/1859_Origin_F373.pdf

- Ewens W J, 2014. Grafen, the Price equations, fitness maximization, optimisation and the fundamental theorem of natural selection[J]. *Biology and Philosophy*, 29(2): 197–205.
- Foster K R, Wenseleers T, Ratnieks F L W, 2006. Kin selection is the key to altruism[J]. *Trends in Ecology & Evolution*, 21(2): 57–60.
- Foster K R, Wenseleers T, Ratnieks F L W, Queller D C, 2006. There is nothing wrong with inclusive fitness[J]. *Trends in Ecology & Evolution*, 21(11): 599–600.
- Gallow J D, 2022. The Metaphysics of Causation[OL]//Zalta E N, Nodelman U. The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/archives/fall2022/entries/causation-metaphysics/>
- Gardner A, 2014. Life, the universe and everything[J]. *Biology and Philosophy*, 29(2): 207–215.
- Gardner A, West S A, Barton N, 2007. The relation between multilocus population genetics and social evolution theory[J]. *The American Naturalist*, 169(2): 207–226.
- Gardner A, West S A, Wild G, 2011. The genetical theory of kin selection[J]. *Journal of Evolutionary Biology*, 24(5): 1020–1043.
- Grafen A, 1982. How not to measure inclusive fitness[J]. *Nature*, 298(5873): 425–426.
- Grafen A, 1999. Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging[J]. *Proceedings of the Royal Society B: Biological Sciences*, 266(1421): 799–803.
- Grafen A, 2000. Developments of Price's Equation and natural selection under uncertainty[J]. *Proceedings of the Royal Society B: Biological Sciences*, 267(1449): 1223–1227.
- Grafen A, 2002. A first formal link between the Price equation and an optimization program[J]. *Journal of Theoretical Biology*, 217(1): 75–91.
- Grafen A, 2006a. A theory of Fisher's reproductive value[J]. *Journal of Mathematical Biology*, 53: 15–60.
- Grafen A, 2006b. Optimization of inclusive fitness[J]. *Journal of Theoretical Biology*, 238(3): 541–563.
- Grafen A, 2007a. An inclusive fitness analysis of altruism on a cyclical network[J]. *Journal of Evolutionary Biology*, 20(6): 2278–2283.
- Grafen A, 2007b. Detecting kin selection at work using inclusive fitness[J]. *Proceedings of the Royal Society B: Biological Sciences*, 274: 713–719.
- Grafen A, 2007c. The formal Darwinism project: a mid-term report[J]. *Journal of Evolutionary Biology*, 20(4): 1243–1254.
- Grafen A, 2008. The simplest formal argument for fitness optimization[J]. *Journal of Genetics*, 87: 421–433.
- Grafen A, 2014a. The formal darwinism project in outline[J]. *Biology and Philosophy*, 29(2): 155–174.

- Grafen A, 2014b. The formal darwinism project in outline: response to commentaries[J]. *Biology and Philosophy*, 29(2): 281–292.
- Haig D, 2014. Genetic dissent and individual compromise[J]. *Biology and Philosophy*, 29(2): 233–239.
- Hamilton W D, 1964. The genetical evolution of social behaviour. I & II[J]. *Journal of Theoretical Biology*, 7(1), 1–52.
- Hamilton W D, 1971. Selection of selfish and altruistic behaviour in some extreme models[M]//Fisemberg J F, Dillon W S. *Man and Beast: Comparative Social Behavior*. Washington: Smithsonian Press, 1971: 57–91.
- Helanterä H, Bargum K, 2007. Pedigree relatedness, not greenbeard genes, explains eusociality[J]. *Oikos*, 116(2): 217–220.
- Hesslow G, 1976. Two notes on the probabilistic approach to causality[J]. *Philosophy of Science*, 43(2): 290–292.
- Hitchcock C, 2001. A tale of two effects[J]. *The Philosophical Review*, 110(3): 361–396.
- Hughes W O H, Oldroyd B P, Beekman M, et al., 2008. Ancestral monogamy shows kin selection is key to the evolution of eusociality[J]. *Science*, 320(5880): 1213–1216.
- Huneman P, 2014. Formal Darwinism as a tool for understanding the status of organisms in evolutionary biology[J]. *Biology and Philosophy*, 29(2): 271–279.
- Kay T, Keller L, Lehmann L, 2020. The evolution of altruism and the serial rediscovery of the role of relatedness[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 117(46): 28894–28898.
- Koliofotis V, Verreault-Julien P, 2022. Hamilton’s rule: A non-causal explanation?[J]. *Studies in History and Philosophy of Science*, 92: 109–118.
- Lehmann L, Rousset F, 2014. Fitness, inclusive fitness, and optimization[J]. *Biology and Philosophy*, 29(2): 181–195.
- Lehmann L, Rousset F, 2020. When do individuals maximize their inclusive fitness?[J]. *The American Naturalist*, 195(4): 717–732.
- Levin S R, Grafen A, 2019. Inclusive fitness is an indispensable approximation for understanding organismal design[J]. *Evolution*, 73(6): 1066–1076.
- Levin S R, Grafen A, 2021. Extending the range of additivity in using inclusive fitness[J]. *Ecology and Evolution*, 11(5): 1970–1983.
- Liao X, Rong S, Queller D C, 2015. Relatedness, Conflict, and the Evolution of Eusociality[J]. *PLOS Biology*, 13(3): e1002098.
- Marshall J A R, 2016. What is inclusive fitness theory, and what is it for?[J]. *Current Opinion in Behavioral Sciences*, 12: 103–108.
- Mas-Colell A, Whinston M D, Green J R, 1995. *Microeconomic theory*[M]. Oxford: Oxford University Press.
- Maynard Smith J, 1964. Group selection and kin selection[J]. *Nature*, 201(4924): 1145–1147.

- Maynard Smith J, Szathmáry E, 1998. *The Major Transitions in Evolution*[M]. New York: Oxford University Press.
- Michod R E, Hamilton W D, 1980. Coefficients of relatedness in sociobiology[J]. *Nature*, 288(5792): 694–697.
- Nowak M A, Allen B L, 2015. Inclusive Fitness Theorizing Invokes Phenomena That Are Not Relevant for the Evolution of Eusociality[J]. *PLOS Biology*, 13(4): e1002134.
- Nowak M A, McAvoy A, Allen B, et al., 2017. The general form of Hamilton's rule makes no predictions and cannot be tested empirically[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 114(22): 5665–5670.
- Nowak M A, Tarnita C E, Wilson E O, 2010. The evolution of eusociality[J]. *Nature*, 466(7310): 1057–1062.
- Okasha S, 2016a. On Hamilton's Rule and Inclusive Fitness Theory with Nonadditive Payoffs[J]. *Philosophy of Science*, 83(5): 873–883.
- Okasha S, 2016b. The Relation between Kin and Multilevel Selection: An Approach Using Causal Graphs[J]. *The British Journal for the Philosophy of Science*, 67(2): 435–470.
- Okasha S, 2018. *Agents and goals in evolution*[M]. Oxford: Oxford University Press.
- Okasha S, Martens J, 2016. Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games[J]. *Journal of Evolutionary Biology*, 29(3): 473–482.
- Okasha S, Paternotte C, 2014a. Adaptation, fitness and the selection-optimality links[J]. *Biology and Philosophy*, 29(2): 225–232.
- Okasha S, Paternotte C, 2014b. The formal Darwinism project: editors' introduction[J]. *Biology and Philosophy*, 29(2): 153–154.
- Orzack S H, 2014. A commentary on “The Formal Darwinism Project”: there is no grandeur in this view of life[J]. *Biology and Philosophy*, 29(2): 259–270.
- Pernu T K, Helanterä H, 2019a. Genetic relatedness and its causal role in the evolution of insect societies[J]. *Journal of Biosciences*, 44(4).
- Pernu T K, Helanterä H, 2019b. Social evolution and the two elements of causation[J]. *Oikos*, 128(7): 905–911.
- Queller D C, 1985. Kinship, reciprocity and synergism in the evolution of social behaviour[J]. *Nature*, 318(6044): 366–367.
- Queller D C, 1992. A general model for kin selection[J]. *Evolution*, 46(2): 376.
- Queller D C, 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 108(supplement_2): 10792–10799.
- Queller D C, 2016. Kin selection and its discontents[J]. *Philosophy of Science*, 83(5): 861–872.

- Queller D C, Strassmann J E, 1998. Kin selection and social insects[J]. *BioScience*, 48(3): 165–175.
- Queller D C, Strassmann J E, 2009. Beyond society: the evolution of organismality[J]. *Philosophical Transactions - Royal Society. Biological Sciences*, 364(1533): 3143–3155.
- Ratnieks F L W, 1988. Reproductive harmony via mutual policing by workers in eusocial hymenoptera[J]. *The American Naturalist*, 132(2): 217–236.
- Ratnieks F L W, Visscher P K, 1989. Worker policing in the honeybee[J]. *Nature*, 342(6251): 796–797.
- Rice S H, 2004. *Evolutionary theory: mathematical and conceptual foundations*[M]. Sunderland: Sinauer Associates.
- Rodrigues A M M, Gardner A, 2023. Inclusive fitness: a scientific revolution[M]//Dickins T E, Dickins B J A. *Evolutionary Biology: Contemporary and Historical Reflections Upon Core Theory*. Gewerbestrasse: Springer, 2023: 343–360.
- Rubin H, 2018. The debate over inclusive fitness as a debate over methodologies[J]. *Philosophy of Science*, 85(1): 1–30.
- Sarkar S, 2014. Formal Darwinism[J]. *Biology and Philosophy*, 29(2): 249–257.
- Smith J, van Dyken J D, Zee P C, 2010. A generalization of Hamilton's rule for the evolution of microbial cooperation[J]. *Science*, 328(5986): 1700–1703.
- Szathmáry E, 2015. Toward major evolutionary transitions theory 2.0[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33): 10104–10111.
- Shelton D, Michod R E, 2014. Levels of selection and the formal Darwinism project[J]. *Biology and Philosophy*, 29(2): 217–224.
- Tartina C E, Taylor P D, 2014. Measures of relative fitness of social behaviors in finite structured population models[J]. *The American Naturalist*, 184(4): 477–488.
- Trivers R L, Hare H, 1977. Haplodiploidy and the Evolution of the Social Insects[J]. *Science*, 191(4224): 249–263.
- van Veelen M, Allen B, Hoffman M, et al., 2017. Hamilton's rule[J]. *Journal of Theoretical Biology*, 414: 176–230.
- Walsh D M, Lewens T, Ariew A, 2002. The Trials of Life: Natural selection and random drift[J]. *Philosophy of Science*, 69(3): 452–473.
- Walsh D M, Ariew A, Matthen M, 2017. Four Pillars of Statisticalism[J]. *Philosophy, Theory, and Practice in Biology*, 9(20171201).
- Weinberger N, 2019. Path-Specific effects[J]. *The British Journal for the Philosophy of Science*, 70(1): 53–76.
- Wilson E O, 1975. Wilson, E. O. (1975). *Sociobiology: the new synthesis*[M]. Cambridge: Belknap Press of Harvard University Press.
- Wilson E O, 2005. Kin selection as the Key to altruism: its rise and fall[J]. *Social Research an International Quarterly*, 72(1): 159–166.

- Wilson E O, 2008. One giant leap: How insects achieved altruism and colonial life[J]. *BioScience*, 58(1): 17–25.
- Wilson E O, Hölldobler B, 2005. Eusociality: Origin and consequences[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 102(38): 13367–13371.
- Wilson E O, Nowak M A, 2014. Natural selection drives the evolution of ant life cycles[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35): 12585–12590.
- Woodward J, 2003. *Making things happen: a theory of causal explanation*[M]. New York: Oxford University Press.
- Wright S, 1922. Coefficients of inbreeding and relationship[J]. *The American Naturalist*, 56(645): 330–338.

致 谢

感谢导师王巍教授对论文的指导。论文开题报告、第 2 章和第 3 章的草稿都曾在王巍老师组会上报告，感谢王巍老师和在场同学的意见和建议。感谢陆俏颖老师对论文第 2 章初稿的修改意见。论文第 2 章的一个早期版本曾投稿至《自然辩证法通讯》，感谢审稿人提出的修改意见。论文第 2 章的一个版本曾在清华大学科学史系学术例会上报告，感谢在场听众的评论、提问和讨论。论文第 2 章的一个版本曾在第六届“科学与哲学青年学者工作坊”报告，感谢评论人张明君老师和在场听众的评论、提问和讨论，特别是卢天杨同学对论文的严厉批评。感谢布拉特（Pierrick Bourrat）老师和我就与论文第 2 章和第 3 章有关内容进行的有益交流。

感谢论文公开评阅人陈勃杭老师、陆俏颖老师和蒋澈老师对论文的评阅意见。

感谢论文答辩委员会杨舰老师、孙承晟老师、蒋澈老师、陆俏颖老师和王巍老师在论文答辩会上对论文提出的意见和建议。

感谢清华大学图书馆对本研究提供的文献资源支持。

感谢马悦老师、任昱霖同学、周敏萱同学和左宁宇同学对本研究提供的情感支持。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名： 王康桥 日 期： 2024年9月10日

个人简历、在学期间完成的相关学术成果

个人简历

1999 年 8 月 21 日出生于江苏省苏州市。

2017 年 9 月考入北京大学哲学系（宗教学系）哲学专业，2022 年 6 月本科毕业并获得哲学学士学位。

2022 年 9 月考入清华大学人文学院攻读科学技术史硕士至今。

在学期间完成的相关学术成果

学术论文：

- [1] 王康桥，录用待发表. 对广义适合度理论中汉密尔顿规则解释价值的质疑——驳乔纳森·伯奇[J]. 自然辩证法通讯.

指导教师评语

广义适合度理论是当代社会进化的重要理论，但也一直饱受争议。论文对广义适合度理论的数学公式与哲学诠释进行深入研究，选题很有理论意义与学术价值。

作者的哲学功底很好，并且自学了相关的生物学理论，全面分析了广义适合度理论，取得了三点创新：（1）汉密尔顿规则的解释价值有问题，引入了新的方案来理解；（2）分析了目前广义适合度理论之争的真正争议点在于 IBD 亲缘系数在真社会性进化中的地位；（3）批评了奥卡沙的混淆，论证了广义适合度最大化和能动者思维并无紧密联系。相关研究成果已有一篇论文被 CSSCI 期刊《自然辩证法通讯》录用，另有英文论文得到国内外生物学哲学家的好评。

论文的文献引用全面，数学细节清晰，分析深入，思路严谨，写作规范，是一篇较为优秀的硕士学位论文。

答辩委员会决议书

王康桥的硕士学位论文对广义适合度理论的数学公式与哲学诠释做了深入研究，选题前沿，有较高的理论意义和学术价值。

论文细致地分析了广义适合度理论的历史背景、理论框架以及汉密尔顿规则在学界引发的争论，阐述了汉密尔顿规则的解释价值，评述了真社会性进化这一案例与广义适合度理论的关系，论证了广义适合度最大化和能动者思维并无紧密联系。

论文结构清晰，逻辑缜密，论证充分，写作规范，兼具理论深度和创新性，是一篇优秀的硕士论文。

王康桥很好地回答了答辩委员会的提问，答辩委员会由五人组成，经无记名投票表决，一致同意通过王康桥的硕士学位论文答辩，建议授予其科学技术史硕士学位。