UAB
Universitat Autònoma de Barcelona

From Structure to Identity
The Computational Dynamics of Multi-Agent Cooperation

Alejandra López de Aberasturi Gómez

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: https://creativecommons.org/licenses/?lang=ca

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: https://creativecommons.org/licenses/?lang=es

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

From Structure to Identity
The Computational Dynamics of Multi-Agent Cooperation

Alejandra López de Aberasturi Gómez

Phd thesis

From Structure to Identity:

The Computational Dynamics of Multi-Agent Cooperation

UAB

Alejandra López de Aberasturi Gómez
Instituto de Investigación en Inteligencia Artificial
Consejo Superior de Investigaciones Científicas

A thesis submitted for the degree of
Doctor of Philosophy

January 2026

A mi familia.

El mundo era tan reciente, que muchas
cosas carecían de nombre,
y para mencionarlas había que señalarlas
con el dedo.

Gabriel García Márquez

In memoriam

A Ana María Garrido Acero.

Y si ya se estuvo en la nada,
o en la no existencia,
no es tan extraño ni grave
regresar a ella.

Javier Marías

Acknowledgements

My first words of gratitude go to my thesis supervisor, Carles Sierra. His guidance allowed me to explore my own lines of work and to take intellectual risks from the very beginning. That autonomy has been decisive for my training as a researcher, particularly in the formulation of relevant questions and in the pursuit of results with the methodological rigour that a research project requires. Carles' generosity has gone beyond the strictly academic sphere. I am also grateful for his hospitality throughout this period. From him I have learned that maintaining a constructive attitude when facing difficulties can be a reasonable choice. That lesson has had personal value for me as well.

My second acknowledgement goes to Jordi Sabater-Mir, who, in addition to co-supervising this thesis, facilitated my incorporation into the IIIA both during my Master's degree and the doctoral programme. His guidance was decisive in enabling me to carry out my work with a high degree of autonomy while incorporating an applied perspective. I have learnt from him the importance of evaluating the applicability of one's own ideas, a fundamental aspect of rigorous research.

I thank both my supervisor and co-supervisor for the opportunities they offered me to advance in my training, expand my scientific network, and participate in international environments. Within that network, I wish to mention Juan Antonio Rodríguez-Aguilar, head of the team with which I have collaborated during this period. His guidance and support contributed substantially to the development of this thesis and to my consolidation as a researcher, in close complementarity with the supervision of Carles Sierra and Jordi Sabater-Mir. I am also grateful to the team led by Juan Antonio for the discussions and comments that improved and helped to reshape the theoretical and computational models presented here. Among them, I wish to highlight the contribution of Maite López-Sánchez, whose analytical clarity has been particularly valuable. I hope that this collaboration may continue and strengthen in the future.

Moreover, my affiliation with the IIIA and my status as a doctoral student under the supervision of Carles Sierra provided me with professional development opportunities beyond the strictly academic context. Among them, I highlight my active participation in the design and implementation of an Artificial Intelligence exhibition at CaixaForum València. This experience was especially formative thanks to the leadership of Marta García-Matos, whose drive for science and public engagement was key to the project.

My gratitude also extends to those who accompanied me from the University of Oxford. Mike Wooldridge has been an academic reference whose ideas have significantly influenced this work. His course on Game Theory at the IIIA, as well as our subsequent conversations, guided the development of Chapter 3. I later had the opportunity to undertake a research stay under his supervision at the Department of Computer Science in Oxford, which contributed substantially to my research training. I am grateful for the chance to work in such an intellectually stimulating environment.

Among those I collaborated with during that period, I would like to mention Arnau Quera-Bofarull and Daniel Jarne-Ornia. Their contributions and our discussions enriched the ideas that eventually crystallised in Chapter 5. It is a collaboration that I continue to value and maintain today.

To all the researchers and colleagues who took part in the pilot study presented in this thesis, I thank you for your time and support.

I am likewise grateful to the friends who have accompanied me during this period. Among them, a few rare connections offer clarity and a quiet sense of home, even if I cannot list each one by name.

Finally, I recognise that it is difficult to adequately express the relevance that my family and my partner have had in my journey; their support exceeds what I can convey in these lines.

Abstract

Understanding how cooperation emerges among self-interested agents without central enforcement remains a fundamental challenge across the social and computational sciences. As artificial and human agents increasingly interact, misalignment between individual incentives and collective outcomes becomes both scientifically and practically consequential. Yet existing approaches provide only fragmented accounts: game theory offers structural insight but assumes enforceable cooperation, multi-agent learning captures adaptation but relies on domain-specific environments, and economic models rarely integrate dynamic social identity.

This thesis proposes a unifying perspective: cooperation arises when structural, temporal, and social incentives become aligned with group welfare. To examine these mechanisms, the work combines formal modelling with multi-agent simulations. The resulting contributions serve a dual purpose: they formalise psychosocial theories within game-theoretic and economic frameworks, advancing computational models of human-like teamwork; and, conversely, they leverage simulation environments as testbeds for these theories, enabling systematic evaluation and generating predictions that inform and refine their development. This dual perspective is operationalised through a sequence of modelling contributions that progressively integrate structural, temporal, and social dimensions of cooperation.

First, it introduces *Teamwork Games*, a non-cooperative model of teamwork grounded in group-productivity theory, and evaluates the learning dynamics of decentralised agents within this framework. The analysis is then extended through *Sequential Teamwork Games*, a Markov-game formulation that enables the identification of temporal social dilemmas. Finally, the thesis offers a micro-founded model of dynamic social identity and embeds it within an economic agent-based framework of teamwork and job mobility.

Together, these contributions advance a multilayered theory of cooperation, demonstrating how collective behaviour emerges—and collapses—through the interplay of structural, temporal, and social mechanisms.

Contents

1 Introduction **3**

1.1 Research Gaps . . . 7

1.1.1 Structural Alignment: Gaps in Game Theory . . . 7

1.1.2 Temporal Alignment: Gaps in Multi-Agent Reinforcement Learning . . . 9

1.1.3 Social Alignment: Gaps in Identity Economics . . . 9

1.2 Central Proposal of the Thesis . . . 10

1.3 Approach and Methodology . . . 11

1.4 Objectives . . . 13

1.5 Main Contributions . . . 13

1.6 Thesis Roadmap . . . 15

2 Methods **17**

2.1 Game Theory and Aggregative Games . . . 18

2.1.1 Social Dilemmas . . . 19

2.1.2 Aggregative Games . . . 20

2.1.2.a Generalised Aggregative Games . . . 20

2.1.2.b General Public-Good Games . . . 23

2.2 Learning in Games . . . 24

2.3 Markov Games and Multi-Agent Reinforcement Learning . . . 26

2.4 Deep Reinforcement Learning . . . 27

2.4.1 Policy Gradient Methods . . . 28

2.4.2 Generalised Advantage Estimation (GAE) . . . 30

2.4.3 Proximal Policy Optimization (PPO) . . . 31

3 To Work, or Not to Work: Teamwork Games **35**
3.1 Introduction: The Teamwork Dilemma . . . 38
3.2 Social-Psychological Theories of Group Performance . . . 39
3.2.1 Steiner's Model of Group Performance . . . 40
3.2.2 Nijstad's Model of Group Performance . . . 42
3.3 Game-Theoretic and Computational Perspectives on Teamwork . . . 43
3.4 Teamwork Games . . . 46
3.4.1 Formal Model . . . 46
3.4.2 Computing Nash Equilibria in Teamwork Games . . . 53
3.5 *CollabQuest*: A Multi-Agent Multi-Armed Bandit for Teamwork . . . 67
3.5.1 Teamwork Game (one-shot). . . . 68
3.5.2 Learning Setup . . . 69
3.6 Experiments . . . 70
3.6.1 Experiment 1: Do Multi-Agent Multi-Armed Bandit Teams Learn Nash-consistent policies? . . . 75
3.6.2 Experiment 2: Interaction Between Team Composition and Task Type . . . 78
3.6.3 Experiment 3: Interaction Between Team Composition and Evaluation Treatment . . . 81
3.6.4 Experiment 4: Individual Strategies . . . 83
3.6.5 Experiment 5: Binary Pass-Fail Evaluation . . . 87

4 Multi-Agent Sequential Teamwork: The Dynamics of Cooperation **93**
4.1 Introduction: Anna's Choice . . . 95
4.2 Teamwork Across Time: Dynamics and Determinants of Group Performance . . . 97
4.2.1 The Input-Mediator-Output-Input (IMOI) Model of Team Performance . . . 98
4.3 Sequential Teamwork Games . . . 99
4.3.1 Context and Related Work . . . 100
4.3.2 Formal Model . . . 103
4.4 *TEMPO*: A Markov Game Environment for Sequential Teamwork . . . 107
4.4.1 Markov Game. . . . 107
4.4.2 Learning Setup . . . 111
4.5 Experiments . . . 112

4.5.1 Experiment 1: Sequential Teamwork Games and Sequential Social Dilemmas . . . 115
4.5.2 Experiment 2: Emerging Behaviour in Sequential Teamwork Games . . . 127
4.5.3 Experiment 3: Teamwork Mediators in Sequential Teamwork Games . . . 138

5 Agents, Firms and Identity 149
5.1 Introduction: Times of Change . . . 152
5.2 Self-Categorisation Theory and Groups at Work . . . 154
5.3 Existing Multi-Agent and Identity-Based Models in Economics . . . 157
5.4 An Agent-Based Model of Firm Identity and Effort Contribution . . . 160
5.4.1 Setup and Notation . . . 161
5.4.2 Elements of a Self-Categorisation Theory-Compatible Utility Function . . . 162
5.4.3 Formalising the Meta-Contrast Ratio . . . 163
5.4.4 A Function for Social Identity Salience . . . 164
5.4.5 A Utility Formulation for Firm Identity and Effort Contribution . . . 165
5.5 Comparative Statics . . . 171
5.5.1 Behavioural comparative statics I: Identity salience . . . 173
5.5.2 Behavioural comparative statics II: Firm size . . . 176
5.5.3 Welfare and identity salience . . . 178
5.5.4 Welfare and firm size . . . 180
5.6 Simulation Design and Implementation . . . 183
5.6.1 One-Dimensional Traits . . . 183
5.6.2 Information, Network and Candidate Sets . . . 183
5.6.3 Local Evaluation Heuristic . . . 183
5.6.4 Calibration and Parameters . . . 183
5.7 Experiments . . . 185
5.7.1 Experiment: Endogenous Identity with Network-Homophily Traits . . . 187

6 Conclusion 199
6.1 Synthesis of Contributions . . . 200
6.1.1 Chapter 3: Structural Alignment . . . 201
6.1.2 Chapter 4: Temporal Alignment . . . 202
6.1.3 Chapter 5: Social Alignment . . . 203

6.2 Integrative Insights . . . 204
6.3 Limitations and Future Work . . . 207
6.4 Applied Directions and Transfer Opportunities . . . 208
6.5 Outstanding Questions and Concluding Remarks . . . 210

A Teamwork Games **219**
A.1 Proof of Lemma 3.9 . . . 220
A.2 Proof of Lemma 3.10 and Proposition 3.12 . . . 220
A.3 Analysis of a player's replacement function in teamwork games with conjunctive tasks . . . 222
A.4 Proof of Lemma 3.11 . . . 224
A.5 Teamwork Games as Social Dilemmas . . . 227
A.6 Proof of Proposition 3.14 . . . 231
A.7 Individual efforts in disjunctive and conjunctive tasks: tables . . . 232

B Multi-Agent Sequential Teamwork **237**
B.1 Implementation Details for Sequential Teamwork Games . . . 237
B.1.1 Environment Construction . . . 237
B.1.2 Training Loop . . . 238
B.1.3 Attribute Shuffling Procedure . . . 238
B.1.4 Neural Architectures . . . 239
B.1.5 Logging and Diagnostics . . . 239
B.2 Convergence Criterion for Empirical Payoff Estimation . . . 239
B.3 Mapping of Sequential Social Dilemma Presence Across Task Types . . . 240
B.4 Team Composition . . . 240
B.5 Cooperativeness in Heterogeneous Teams . . . 241
B.6 Moderators of Cooperativeness in Heterogeneous Teams . . . 242
B.7 Proxies of Implicit Coordination: Tables . . . 244

C The Island **249**
C.1 Introduction . . . 250
C.2 Game Design . . . 251
C.2.1 Platform Architecture . . . 251
C.2.2 User Interaction and Interface . . . 251

C.2.3 Player Assignment and Team Formation . . . 253
C.2.4 Administrator Dashboard . . . 254
C.2.5 Game Logic . . . 255
C.2.6 Mini-game Structure and Scoring . . . 257
C.2.7 Mini-game Descriptions . . . 259
C.2.8 Database . . . 260
C.2.9 Pilot test . . . 261
C.3 Conclusions . . . 262

D Agents, Firms and Social Identity **265**
D.1 Derivations for the Identity-Salience Comparative Statics . . . 266
D.2 Derivations for the Firm-Size Comparative Statics . . . 269
D.3 Detailed computation of $\frac{\partial V_i^{\mathrm{SCT}}}{\partial n}$ . . . 271
D.4 Justification of Proposition 5.7 . . . 272
D.5 Simulation Kernel . . . 274
D.6 Size and Utility in a Degenerated Case . . . 275
D.7 Scenario A (i.i.d. Traits): Endogenous Identity without Network Homophily . . . 276
D.8 Scenario B (Exogenous Identity): Salience Control via Sampled Meta-Contrast Ratio (MCR) . . . 277
D.8.1 Salience response . . . 277
D.8.2 Behavioural and meso-level effects . . . 278
D.8.3 Macro structure . . . 278

Bibliography **281**

List of Figures

3.1 Nijstad's model of group performance . . . 43
3.2 A player's utility in the three regimes of disjunctive tasks . . . 57
3.3 Total group product *vs.* individual gifts as per Equation (3.17) . . . 62
3.4 Components of player i's share correspondence . . . 65
3.5 Logistic evaluation function . . . 73
3.6 Utilities $\widetilde{u}_i$ of two agents in a team . . . 75
3.7 Representation of the convergence of Q-values during training of two agents. . . . 77
3.8 Learned aggregates *vs.* theoretical Nash Equilibria (NEs) aggregates . . . 77
3.9 Variation in team productivity across teams and tasks. . . . 79
3.10 Heaviside evaluation function . . . 88
3.11 Group product with Heaviside evaluation . . . 88

4.1 The Input-Mediator-Output-Input model of team performance [Ilgen et al., 2005] 98
4.2 Evolution of the task state over time, with rewards. . . . 105
4.3 Agent-environment interaction in a Sequential Teamwork Game. . . . 110
4.4 Per-agent training curves . . . 115
4.5 Schelling diagram for medium-difficulty, additive tasks . . . 121
4.6 Empirical mapping of Sequential Social Dilemmas (Sequential Social Dilemmas (SSDs)) under the additive task regime with medium difficulty. . . . 123
4.7 Average action trajectories for four agents performing an additive task ($\rho = 1$) of medium difficulty. . . . 130
4.8 Average team effort (percentage of time) across task type (ρ) and average expertise $\overline{p}$ . . . 131
4.9 Effect of the preference for leisure (α) on effort-expertise curves . . . 134
4.10 Effect of evaluation hardness ($G^{\mathcal{T}}$) on effort-expertise curves . . . 136

4.11 Distribution of anticipation coefficients (θ_1) across task types 144
4.12 Distribution of dynamic adjustment indices ($\overline{D}$) across task types 144
4.13 Distribution of role specialisation (NMI) across task types 145

5.1 Computational pipeline of the proposed identity-enriched extension of Axtell's firm-formation model. Firm composition determines the meta-contrast ratio (MCR_f), which endogenously generates workers' identity salience ϕ_i. This salience modulates the SCT-compatible utility function, shaping both effort provision and mobility decisions. These individual choices subsequently alter firm composition and social contrast, closing the feedback loop between organisational structure, identity formation, and economic behaviour over time. 172
5.2 Average firm size over time in the baseline economy 190
5.3 Average firm size over time in the degenerate social economy 191
5.4 Average social-identity salience for different clustering thresholds M_0 191
5.5 Cross-sectional effort distributions by salience 192
5.6 Maximum firm size and identity salience over time. 193
5.7 Individual effort (median) versus firm size (n, logarithmic scale). 194
5.8 Probability Mass Function (PMF) of firm size 194

B.1 Mapping of Sequential Social Dilemma Presence Across Task Types 245
B.2 Average team effort (percentage of time) in heterogeneous teams 246
B.3 Effect of the preference for leisure (α) on effort–expertise curves 246
B.4 Effect of the passing threshold ($G^{\mathcal{T}}$) on effort-expertise curves 246

C.1 Login screen of *The Island*. Players access the platform using a pre-assigned username and password, and can select their preferred language (English, Spanish, or Catalan) before proceeding. 252
C.2 Instruction screen presented to players after logging in. The general rules and objectives of the game are displayed until the user confirms understanding by pressing the `CONTINUE` button. 252
C.3 Forming team screen. While waiting for a team to be formed on the fly, players are reminded of their progress so far and their avatar's levels in each mini-game. 253

C.4 Administrator Dashboard. The administrator class has access to detailed game and player data, including current activity on the platform, currently connected players, and pending games. It is also possible to create backup copies, download or delete the database and broadcast messages to all the players at once. . . 255
C.5 Sequence of screens presented to a player during a mini-game in *The Island*, illustrating the decision, waiting, and feedback stages. **(a): Decision screen.** The player selects how much time to dedicate to the task by clicking on the white bar. Her skill level and those of the other players are displayed on the left. **(b): Decision confirmation.** Once the bar has been set, the corresponding contribution and rest time are shown below. **(c): Waiting screen.** After confirming her decision, the player waits while her teammates make their own selections. **(d): Turn results screen.** Once all decisions are made, the team's individual and collective outcomes for the turn are displayed, together with cumulative progress and the lower threshold. 256
C.6 Fishing mini-game (**additive task**). Each player decides what fraction of her time to devote to fishing. Individual catches depend on the avatar's fishing skill, and the team's total catch in each turn is the sum of all individual contributions. . . 259
C.7 Building the shelter mini-game (**conjunctive task**). Players allocate time to carrying building materials. The team's progress in each turn is limited by the smallest individual contribution, illustrating the conjunctive rule where collective output depends on the least-performing member. 260
C.8 Navigation mini-game (**disjunctive task**). Players devote time to performing calculations that help the team navigate. The team's progress each turn is determined by the highest individual contribution, reflecting a disjunctive task in which the most capable member drives team success. 260

D.1 Average salience over time for different normal trait distributions 276
D.2 Average salience over time for different log-normal means μ in the sampled-MCR scenario 277
D.3 Share of free-riders over time (sampled-MCR scenario) 278
D.4 Job changes over time (sampled-MCR scenario) 279
D.5 Maximum firm size over time (sampled-MCR scenario) 279
D.6 Firm-size distribution under exogenous MCR control 280

List of Tables

3.1 Steiner's task taxonomy . . . 42
3.2 Key variables and parameters in Teamwork Games . . . 47
3.3 Numerically computed Nash equilibria for a heterogeneous team . . . 66
3.4 Fixed hyperparameters and environment settings . . . 74
3.5 Task types used in the experiments . . . 76
3.6 Evaluation conditions used in the experiments . . . 76
3.7 Soft Evaluation Treatment . . . 85
3.8 Percentage increase in average dedication . . . 86
3.9 Percentage dispersion of group products across three independent trainings per team . . . 89

4.1 Structure of the agent's local observation vector . . . 108
4.2 Difficulty levels . . . 110
4.3 Optimised hyperparameters . . . 113
4.4 Fixed hyperparameters used across all runs. . . . 114
4.5 Experimental variables used in Experiment 1. . . . 118
4.6 Conditions under which different task types give rise to ssd, along with the dominant incentive . . . 122
4.7 Conceptual mapping of environment dials. . . . 128
4.8 One-factor-at-a-time experimental grid used for Experiment 2. . . . 129
4.9 Pairwise global comparisons of average team effort across task types (restricted to active regimes and homogeneous teams) . . . 132
4.10 Pairwise global comparisons of team effort dispersion across task types (restricted to active regimes and homogeneous teams) . . . 132
4.11 Global vertical shifts in cooperation across α levels . . . 135

4.12 Horizontal shifts in activation thresholds across α levels 135
4.13 Vertical shifts in cooperation across difficulty levels 137
4.14 Horizontal activation shifts across difficulty levels 137

5.1 Elements of a Self-Categorisation Theory-compatible utility function. 171
5.2 Simulation parameters, distributions, and initialisation. 185
5.3 Difficulty levels 195

A.1 Individual effort across different teams for a **conjunctive** task ($\rho = -10$) under three different assessment treatments 233
A.2 Individual effort across different teams for a **disjunctive** task ($\rho = 10$) under three different assessment treatments 234

B.1 Team composition in homogeneous teams 240
B.2 Team composition in heterogeneous teams 241
B.3 Anticipation proxy 244
B.4 Dynamic adjustment 244
B.5 Role specialisation proxy 244

C.1 Summary of pilot coordination proxies 262



1 Introduction

In an astonishing variety of contexts, apparently complex structures or behaviors emerge from systems characterized by very simple rules. These systems are said to be self-organized and their properties are said to be emergent. The grandest example is the universe itself, the full complexity of which emerges from simple rules plus the operation of chance.

Murray Gell-Mann

1.1 Research Gaps 7
1.2 Central Proposal of the Thesis 10
1.3 Approach and Methodology 11
1.4 Objectives 13
1.5 Main Contributions 13
1.6 Thesis Roadmap 15

Under what conditions will cooperation emerge in a world of egoists without central authority?

With this question, Axelrod inaugurated a computational research tradition that sought to explain the emergence of cooperation through simple interaction rules [Axelrod and Hamilton, 1981]. His work built on the theoretical foundations of *Evolutionary Game Theory*—which provides the formal conditions under which strategies resist invasion, as established by Maynard Smith & Price [Maynard Smith and Price, 1973] and Maynard Smith [Maynard Smith, 1982]—and introduced a constructive, simulation-based perspective, showing how cooperative behaviour can emerge and persist through processes of selection and differential success among self-interested agents.

On the theoretical side, early works by Maynard Smith & Price characterised the stability of strategies from a population-level perspective—*i.e.*, without explicitly modelling the underlying processes through which such states are reached. Subsequent developments in evolutionary game theory incorporated dynamic accounts of strategic change, most notably through the formalisation of *replicator dynamics* by Taylor & Jonker [Taylor and Jonker, 1978] and later work by Hofbauer & Sigmund [Hofbauer and Sigmund, 1998]

Both of these early evolutionary and computational approaches provided a macroscopic, long-run perspective in which cooperation emerges as an evolutionary outcome across generations. However, many cooperative phenomena—both in human societies and in artificial multi-agent systems—unfold within the much shorter timescale of individual interactions.

Börgers and Sarin [Börgers and Sarin, 1997] provided the pivotal formal bridge between these scales by proving that, in a continuous-time limit, a specific reinforcement learning model—known as the Cross model [Cross, 1973]—converges precisely to the asymmetric replicator equation. This mathematical foundation established a non-biological interpretation of game dynamics, allowing them to be viewed as the result of individual mental adaptation rather than genetic selection. Tuyls [Tuyls et al., 2003, 2004, Tuyls and Nowé, 2005, Tuyls and Parsons, 2007] subsequently extended this work by showing a formal correspondence between replicator dynamics and modern algorithms such as Q-learning, and by expanding the framework from stateless interactions to multi-state (stochastic) games [Bloembergen et al., 2015]. Within this framework, replicator dynamics and reinforcement learning serve as complementary perspectives: the former provides an analytic description of population-level dynamics, while the latter offers an algorithmic account of adaptive behaviour at the level of individual

agents, incorporating history-dependent learning and enabling the study of settings in which analytical solutions become intractable due to the complexity of the state and strategy spaces.

More than fifty years after Axelrod's seminal inquiry, the study of cooperation through computational and game-theoretic models has evolved substantially, yet the question of how cooperative behaviour arises from simple local interactions remains as relevant as ever. Modern multi-agent methods make it possible to examine such intragenerational adaptation directly, modelling agents whose behavioural policies evolve as a consequence of their own histories of interaction. For instance, recent work has begun to leverage multi-agent reinforcement learning to study complex social phenomena in temporally extended environments, including partner choice and the emergence of biases in social decision-making [Duéñez-Guzmán et al., 2025]. In this sense, the present work adopts an explicitly algorithmic perspective, leveraging contemporary multi-agent learning frameworks to study cooperation as an emergent property of adaptive, interacting agents.

Just as statistical physics provides micro-founded accounts of thermodynamic regularities, and complex-systems theory elucidates global organisation through simple interaction rules, this thesis seeks mechanistic explanations of how cooperation arises, stabilises, or collapses from the bottom up. From this perspective, it situates itself at the intersection of Artificial Intelligence (AI) and social simulation, within a line of research that develops new formal and computational tools to study cooperative phenomena—a line that, in recent years, has come to be known as *Cooperative AI* [Dafoe et al., 2020].

At the time of writing, many of the once-speculative promises of a society composed of both human and artificial agents have already been realised [Krakowski, 2025, Trappey et al., 2026, Wang et al., 2026]. The need for the latter to be aligned with the former is now widely recognised as both a scientific and a practical priority, to the extent that *AI alignment* has rapidly established itself as a distinct field within AI research [Amodei et al., 2016, Christiano et al., 2017, Russell, 2019, Di Langosco et al., 2022, Bai et al., 2022]. Its primary focus lies in incorporating human goals and values into artificial systems—typically a single agent or model—in order to make them as helpful, safe, and reliable as possible.

However, the scope of *Cooperative AI* is considerably broader [Dafoe et al., 2020, 2021, Conitzer and Oesterheld, 2023]. The term encompasses and connects a wide range of areas, including game theory [Von Neumann and Morgenstern, 1947, Nash, 2024] and social choice [Arrow, 1950, 1951, Olson Jr, 1971], multi-agent systems [Weiss, 1999, Jennings, 1993, Wooldridge, 2009], evolutionary game theory [Maynard Smith and Price, 1973, Maynard Smith,

1982, Hofbauer and Sigmund, 1998, Taylor and Jonker, 1978], behavioural economics [Simon, 1955, Kahneman, 1979, 2002, Ostrom, 2003a] and human–machine interaction [Hollan et al., 2000, Ritter et al., 2014, Dix, 2017, Card, 2018].

The present work proposes using multi-agent simulations as a specific application of AI techniques to study the factors that induce or inhibit cooperation—in short, to understand why an agent with particular motivations chooses to cooperate when it is not strictly required.

Concretely, this thesis focuses on *teamwork under mixed motives*, where agents contribute voluntarily to a shared task without the possibility of enforcing cooperative behaviour. Such settings, which are ubiquitous in human collaboration—from workplace environments to collective-action problems—are characterised by a fundamental tension: while group success depends on joint effort, individual incentives may favour free-riding or partial contribution. They therefore provide a natural testbed for studying cooperation as an emergent phenomenon shaped by task structure, interaction dynamics, and social factors.

This focus is not merely of theoretical interest. Contemporary societies increasingly rely on intelligent systems embedded in everyday life—systems that, while often aligned with individual users, can nonetheless generate undesirable collective outcomes. Personalised recommender systems, for example, may reinforce existing beliefs, amplify false-consensus biases, or contribute to long-run polarisation dynamics [Bakshy et al., 2015, Areeb et al., 2023]. These phenomena illustrate a broader point central to Cooperative AI: **alignment at the individual level does not guarantee beneficial behaviour at the collective level**.

This discrepancy has long been recognised in behavioural economics, where cooperation is understood to depend not only on material payoffs, but also on social preferences and boundedly rational heuristics [Fehr and Schmidt, 1999, Camerer, 2018]. In teamwork settings, these factors directly shape individual incentives to contribute, making the emergence of cooperative behaviour among locally adapting agents both a scientific challenge and a practical necessity.

In adopting this perspective, the thesis treats AI not only as a tool for engineering systems, but as an instrument for scientific inquiry. A central feature of the approach is its bidirectional perspective on the relationship between human behaviour and artificial agents. On the one hand, multi-agent simulations are used as controlled environments in which to formalise and test theories of human cooperation, translating insights from social psychology into computational models. On the other hand, these simulations serve as generative tools: by analysing the behaviour of learning agents, they potentially provide new hypotheses and conceptual

insights that can inform and refine our understanding of human teamwork. In this sense, *a central aim of the thesis is to formalise psychological theories of human cooperation into computational frameworks that can be analysed and instantiated in artificial agents*.

In this sense, the thesis draws on human collaboration as a source of inductive structure for modelling, while using artificial agents as epistemic instruments for studying the mechanisms that sustain—or undermine—cooperation in complex systems.

The next section identifies the key gaps in the literature that are most relevant to this perspective, focusing on models of cooperation that aim to capture teamwork under mixed motives. The gaps are presented in an order that mirrors the structure of the contributions developed throughout this work: from structural, to temporal, to social forms of cooperative alignment. This organisation reflects the conceptual progression that motivates the modelling work that follows.

1.1 Research Gaps

1.1.1 Structural Alignment: Gaps in Game Theory

Classical Game Theory has traditionally approached teamwork through the lens of Cooperative Game Theory. In this literature, teamwork is understood as a setting in which all parties share a joint intention to work together—either because their preferences are aligned or because binding contracts can be enforced [Von Neumann and Morgenstern, 1947, Elkind and Rothe, 2016]. This shifts the focus of analysis: rather than asking whether a team will cooperate and why, Cooperative Game Theory typically addresses how to fairly divide the total utility generated by collective action [Shapley, 1953b, Chalkiadakis et al., 2022].

Similarly, the field of Multi-Agent Systems (MAS) [Weiss, 1999, Shoham and Leyton-Brown, 2008, Wooldridge, 2009] has devoted substantial effort to the design of models [Levesque et al., 1990, Cohen and Levesque, 1991b, Jennings, 1995, Grosz and Kraus, 1996] and systems [Shoham, 1993, Shoham and Tennenholtz, 1995, Tambe, 1997, Pynadath et al., 1999] in which agents coordinate effectively, often under assumptions that facilitate or enforce cooperation. While this line of work has produced powerful frameworks for engineered coordination,

it typically adopts a normative perspective, treating cooperation as an objective to be achieved through appropriate system design.

By contrast, this thesis adopts a perspective that ties in with the descriptive agenda formulated by Shoham [Shoham et al., 2007], focusing on settings in which cooperation cannot be taken for granted. In many real-world contexts—from workplace collaboration to collective-action problems—agents have mixed motives and cannot be compelled to cooperate. Similar tensions arise in systems where artificial agents are embedded in human interaction loops, such as personalised recommender systems, where individually aligned behaviour can give rise to undesirable collective dynamics. In such settings, cooperation must emerge from the interaction between individual incentives and the structure of the task itself.

Analysing these environments calls for a non-cooperative, behaviourally grounded approach to teamwork. Insights from behavioural economics and evolutionary game theory show that cooperative behaviour depends not only on material payoffs, but also on social preferences, bounded rationality, and adaptive dynamics [Simon, 1955, Kahneman, 1979, Hofbauer and Sigmund, 1998, Kahneman, 2002, Ostrom, 2003a]. However, these perspectives have not yet been systematically integrated into formal models of teamwork that explicitly capture how task structure shapes strategic interdependence.

In parallel, social psychology offers a rich tradition of group productivity models that classify tasks according to how individual contributions combine to produce collective outcomes—what we refer to as patterns of *structural interdependence*. Likewise, in economics and game theory, public-good models distinguish between different types of production technologies beyond the additive case, reflecting the underlying geometry of contribution.

Bridging these two strands of work remains an open challenge: developing formal frameworks that translate insights from group psychology into models of strategic interaction among self-interested agents. Advancing such a bridge is key to understanding cooperation in human-like teamwork settings, where collective outcomes depend critically on how individual incentives are shaped by task structure, and more broadly, to informing both the modelling of human collaboration and the design of multi-agent systems operating under human-aligned preferences.

1.1.2 *Temporal Alignment: Gaps in Multi-Agent Reinforcement Learning*

Agent teamwork is not a foreign concept within Multi-Agent Reinforcement Learning (MARL). Benchmark suites such as *Melting Pot* and *SocialJax* [Agapiou et al., 2022, Guo et al., 2025] provide rich collections of multi-agent environments featuring a wide variety of social interactions, including cooperation, competition, deception, reciprocation, trust, and stubbornness. These environments have significantly advanced the empirical study of emergent social behaviour by enabling researchers to observe learning dynamics in complex, temporally extended settings.

However, their strength as richly instantiated environments also introduces certain limitations. Because each scenario is defined through specific game mechanics and environmental details, the insights obtained are often closely tied to the particular structure of the environment in which they arise. As a result, it can be difficult to disentangle which aspects of observed behaviour are driven by general mechanisms of sequential decision-making, and which depend on domain-specific features.

While related work has explored coordination among self-interested agents in temporally extended settings, including approaches based on system-level optimisation and incentive shaping [Wolpert and Tumer, 1999, 2001, Zheng et al., 2020], existing models do not typically provide a unified, domain-independent representation of teamwork that explicitly captures different forms of task interdependence.

This points to a more specific gap: current MARL research lacks general models of teamwork that simultaneously capture its temporally extended nature, its grounding in human-like task structures, and its roots in well-established theories of group productivity. While existing approaches model coordination among learning agents, they do not typically provide abstract, domain-independent formulations of teamwork in which different forms of task interdependence—central in social psychology—are explicitly represented.

1.1.3 *Social Alignment: Gaps in Identity Economics*

Finally, relatively few utility-maximisation models—whether in Game Theory, Reinforcement Learning, or economics—explicitly incorporate the social dimension of decision makers in a way that directly shapes their evaluative framework. This is particularly the case in MARL. Although existing work has begun to incorporate elements aligned with psychosocial theories—such as social learning [Jaques et al., 2019], models of implicit bias in learning agents [Duéñez-Guzmán et al., 2025], or constructs such as Social Value Orientation (SVO) [McKee et al.,

2020]—these approaches do not typically model social identity as an endogenous component of agents' preferences.

In particular, current MARL formulations rarely represent social identity as part of the reward structure through which agents evaluate outcomes. Instead, social behaviour is often modelled through external signals, inductive biases, or auxiliary objectives, rather than as a transformation of the utility function itself.

An important exception in the broader literature is *Identity Economics* [Akerlof and Kranton, 2000b], which studies how group membership can modify an agent's incentives through non-pecuniary terms in their utility function. This work laid the foundations for a research stream that, in many cases, predicts behaviour more accurately than classical economic models. However, existing formulations of Identity Economics typically treat identity as static or exogenously given.

This contrasts with well-established theories in social psychology, such as Self-Categorisation Theory [Turner, 1985, 1987], in which identity is understood as a dynamic, endogenous process that evolves over time and feeds back into how individuals evaluate outcomes in strategic settings, including teamwork.

This points to a specific gap: we lack formal models that integrate identity formation as a dynamic process, evaluative change, and strategic interaction within a unified framework. Bridging this gap is essential for understanding how social identities shape cooperative behaviour in both human-like teamwork and multi-agent systems more broadly.

1.2 Central Proposal of the Thesis

In the chapters that follow, teamwork provides the core setting in which cooperative behaviour is analysed. The underlying idea is that we can only expect cooperation without enforcement when it lies within an agent's self-interest. On this view, a system becomes more aligned—*i.e.*, group incentives more closely coincide with agents' perceived self-interest—when conditions for cooperation are ripe, not when enforcement is more intense.

Put differently, behaviour can be interpreted as the revealed maximisation of a utility function that is not strictly pecuniary. The preferences encoded in such a function are shaped by static and dynamic incentives of various kinds, whose effects intertwine, moderate, and feed

back into agents' behaviour. When individuals engage in a collective task, they operate within a network of interdependence that shapes both their motivations and the consequences of their actions—for themselves and for the group.

Cooperation is best understood not as a fixed state, but as a **hierarchy of alignment mechanisms** Nowak [2006], ranging from structural to temporal to social. These mechanisms shape how individual incentives relate to collective outcomes, and how this relationship evolves through interaction and learning.

From this perspective, cooperation emerges when agents come to perceive alignment between their own incentives and those of the group, a process stabilised through experience and adaptation over time. Understanding cooperation, therefore, requires analysing how alignment is generated, maintained, or disrupted across these different levels.

This central proposal is operationalised through a combination of formal modelling, multi-agent simulation, and behavioural analysis. Each level of alignment is instantiated in a dedicated modelling framework designed to isolate its core mechanisms while allowing interactions across levels to be examined. The next section details the methodological approach overarching this thesis, explaining how these models are constructed, how agents behave within them, and how cooperative behaviour is evaluated in the light of theory and simulation.

1.3 Approach and Methodology

A key feature of this work is that it operates on both sides of the equation—theoretical formulation and practical implementation. It alternates between successive cycles of abstraction and instantiation: analytical demonstration of a mechanism is followed by its operationalisation in a computational environment. This logic permeates the structure of each chapter. Alongside each theoretical model proposed, a companion simulation environment has been built. These environments serve as mirrors: they test whether the theory holds when instantiated in interacting agents, and they reveal aggregate, evolving behaviours that no equation alone can predict.

Consequently, Chapter 3 opens the analysis of alignment mechanisms by examining the most directly designable factors that foster cooperation in teamwork settings. It introduces a novel class of mixed-motive games—*Teamwork Games*—that capture situations in which **agents**

decide their voluntary contributions to a shared task. Formally, the **Nash equilibria** of these games are analytically characterised. Computationally, the chapter develops a **Multi-Agent Multi-Armed Bandit** framework in which agents learn contribution strategies that approximate these equilibria, instantiated in the simulation environment *CollabQuest*.

Chapter 4 extends the analysis of alignment mechanisms by **introducing temporality as a central dimension of teamwork**. Conceptually, the chapter embeds Teamwork Games within a dynamic model, offering a formalisation of cooperative tasks under temporal pressure, heterogeneous roles, and partial observability. The result is a **Markov Game formulation**—*Sequential Teamwork Games*—that extends the static framework of Chapter 3 into a sequential decision-making environment, capturing how team processes can arise endogenously even in the absence of explicit communication or centralised coordination. Computationally, the chapter introduces *TEamwork under Multi-agent Policy Optimisation (TEMPO)*, a **Multi-Agent Reinforcement Learning** simulation environment that operationalises Sequential Teamwork Games.

Chapter 5 adds a final layer of abstraction to the study of cooperation by **examining how workers' sense of belonging to an in-group shapes their behaviour**. Conceptually, the chapter offers a formal, micro-founded representation of *social identity as a scalar field* in the physical sense: a value assigned to each worker, generated endogenously through the dynamic contrast between the focal firm and the set of alternative firms visible to the worker. Through comparative statics, the chapter characterises the joint effects of identity salience and firm size on workers' optimal effort and welfare. Computationally, these cognitive processes are embedded into agents' optimisation rules within Axtell et al. [1999]'s **labour-mobility agent-based model**. The Self-Categorisation Theory (SCT)-compatible formulations introduced in Chapter 5 are, in principle, applicable to any utility function, opening the way to a more comprehensive identity-based economics grounded in social psychology.

Finally, Appendix C begins translating this hierarchy of alignment mechanisms into **human data**. It presents a pilot platform—***The Island***—designed to empirically reproduce the same structural conditions instantiated in *CollabQuest* and *TEMPO*, with the aim of developing a testbed in which to assess the ecological realism of the model's predictions and to obtain behavioural traces that can feed back into the models to improve their validity.

Overall, the approach to cooperation adopted moves fluidly between the mathematical, the computational, and the empirical—from equations, to increasingly sophisticated artificial agents, to people. In doing so, it pursues the following objectives:

1.4 Objectives

The overarching objective of this thesis is to develop a unified framework for understanding cooperation as an emergent phenomenon arising from the alignment of individual and collective incentives in teamwork settings without enforcement.

To achieve this goal, the thesis pursues three interconnected objectives:

(O1) **To develop a formal representation of teamwork grounded in socio-psychological theories of group productivity.** This includes bridging task taxonomies from social psychology with general formulations of public-good games, and analysing the resulting strategic equilibria.

(O2) **To extend this representation to temporally extended settings with adaptive agents.** This involves modelling teamwork as a dynamic process, instantiating it within multi-agent learning environments, and identifying the conditions under which cooperation emerges or breaks down.

(O3) **To incorporate social identity as an endogenous driver of behaviour.** This includes formalising identity formation, integrating it into agents utility functions, and analysing its implications at both the micro and macro levels.

These objectives define a research programme that combines formal modelling, multi-agent simulation, and behavioural analysis to study cooperation across structural, temporal, and social dimensions.

1.5 Main Contributions

The contributions of this thesis are organised around three interconnected dimensions of alignment: structural, temporal, and social. Each dimension combines formal modelling with computational instantiation, and together they provide a multi-level account of cooperation.

Structural Alignment

- **Introduction of a non-cooperative model of teamwork grounded in group productivity theory: *Teamwork Games*.** This model captures teamwork as a voluntary-contribution problem with heterogeneous production technologies derived from social-psychological task taxonomies. Its strategic solutions are analytically characterised.
- **Development of a Multi-Agent Multi-Armed Bandit environment, *CollabQuest*.** This environment instantiates Teamwork Games with boundedly rational agents and exhibits empirical convergence towards the predicted equilibria, providing a bridge between analytical and computational perspectives.

Temporal Alignment

- **Extension to *Sequential Teamwork Games*.** This Markov Game formulation captures temporally extended interaction, evolving task structure, and constraints such as deadlines, transforming static teamwork into a dynamic process.
- **Implementation of the *TEMPO* environment with independent Proximal Policy Optimization agents.** This framework enables the emergence of dynamic teamwork policies and supports empirical game-theoretic analysis of sequential social dilemmas across parameter regimes.
- **Development of an empirical platform—*The Island*.** This platform enables the collection of human behavioural data under controlled structural and temporal conditions, laying the groundwork for future empirical validation.

Social Alignment

- **Formalisation of social identity as an endogenous process.** This includes the introduction of the Meta-Contrast Ratio (MCR) and a scalar salience function grounded in Self-Categorisation Theory, capturing how identity strength evolves with intergroup contrast.
- **Integration into an economic model of labour and mobility.** Identity is embedded directly into agents utility functions, refining Identity Economics through a dynamic, micro-founded account of identity-based motivation.

1.6 Thesis Roadmap

The chapters that follow develop the contributions above in a consequential and cumulative manner. Chapter 2 equips the reader with the technical foundations required to follow the computational theory of cooperation advanced from Chapter 3 onwards.

Chapter 3 proposes a one-shot, non-cooperative model of teamwork and pairs it with a Multi-Agent Multi-Armed Bandit environment to simulate learning dynamics. The results offer insights into the structural drivers and inhibitors of cooperation—that is, the *tangible* variables that shape decision-makers' willingness to contribute to a collective task, such as the geometry of the task and the incentives induced by its evaluation.

These static foundations are extended temporally in Chapter 4, which introduces a Markov Game formulation of teamwork. This dynamic model is instantiated in a MARL environment with iPPO agents, whose emergent policies are analysed to understand how the structural alignment mechanisms of Teamwork Games interact with experience accumulation and state visitation patterns—in other words, how temporality and learning reshape the predictions of the static framework.

The thesis also includes Appendix C, which introduces *The Island*: a pilot platform for collecting human behavioural data in tasks aligned with the structural and temporal conditions studied in *CollabQuest* and *TEMPO*, thereby laying groundwork for future empirical validation.

Chapter 5 incorporates social identity directly into agents' utility—or reward—functions within a setting where team members may also move between firms. In doing so, the chapter *opens* the system: teamwork is now embedded within a broader environment featuring entry and exit. The chapter examines how group identification affects both contribution incentives and retention dynamics within firms.

Finally, Chapter 6 provides an integrative overview of the main findings, discusses the limitations of the work, and outlines future research directions that naturally follow from the models and simulations developed in the thesis.



2 Methods

How can equilibrium be established? The attainment of equilibrium requires a disequilibrium process. What does rational behavior mean in the presence of disequilibrium?

Kenneth J. Arrow

2.1 Game Theory and Aggregative Games 18
2.2 Learning in Games 24
2.3 Markov Games and Multi-Agent Reinforcement Learning 26
2.4 Deep Reinforcement Learning 27

The goal of this chapter is to equip the reader with the formal and computational tools required to understand the models of cooperation developed in the remainder of the thesis.

These tools are primarily used in Chapters 3 and 4, which introduce the structural and temporal dimensions of cooperation, respectively. The social dimension developed in Chapter 5 builds on these same foundations and does not require additional methodological machinery.

Chapter 3 focuses on how individual contributions combine into collective outcomes. This requires a formal framework for representing interdependence, provided by Aggregative Game Theory. To examine whether agents converge towards the predicted equilibria, the analysis is complemented with a dynamic perspective from Learning in Games, implemented computationally in the environment *CollabQuest*.

Chapter 4 extends this setting to temporally structured interaction. Here, agents' actions affect not only immediate outcomes but also the future state of the task. This requires a shift to Markov Games, which model sequential interaction among multiple agents. To study how behaviour emerges in such environments, agents are trained using Deep Reinforcement Learning methods, giving rise to the simulation platform *TEMPO*.

This chapter introduces these tools in turn. It begins with Game Theory and Aggregative Games, followed by Learning in Games in the context of *CollabQuest*. It then introduces Markov Games and concludes with the principles of Deep Reinforcement Learning used to implement *TEMPO*.

2.1 Game Theory and Aggregative Games

Teamwork in Chapter 3 is modelled as a public good game, a type of social dilemma. Consider, for instance, a group of workers contributing effort to a shared task: while the success of the project depends on collective input, each individual has an incentive to reduce their own effort and free-ride on the contributions of others. This tension between individual and collective incentives lies at the core of social dilemmas.

To formalise such situations, this section begins with a concise definition of social dilemmas. It then introduces aggregative games—a subset of normal-form games in which individual payoffs depend on the aggregate of all agents' actions. This framework provides a natural

way to model teamwork settings, where outcomes are determined by how individual contributions combine into a collective result, and serves as the analytical foundation for the models developed in the next chapter.

2.1.1 *Social Dilemmas*

Many problems of collective action can be characterised as *social dilemmas*: situations in which individually rational behaviour leads to collectively suboptimal outcomes [Rapoport and Chammah, 1965, Olson Jr, 1971, Beckenkamp, 2002]. Suppose that, given such a problem, two possible courses of action exist: one corresponding to *cooperation* (C) and one to *defection* (D).

According to Dawes [1980], all social dilemmas share two properties: **i)** the social payoff to each individual from defecting is higher than the payoff from cooperating, and **ii)** the payoff to all individuals is lower if all defect than if all cooperate.

Formally, let $\mathcal{I} = \{1, \ldots, n\}$ denote a set of players and $a_i \in \mathcal{A}_i = \{C, D\}$ their individual strategies. Denote $u_i(D, m)$ the payoff to a defector in an n-person game with m cooperators, and $u_i(C, m)$ the payoff to a cooperator when m *other* players cooperate. A social dilemma is characterised by two inequalities:

$$\begin{cases} u_i(D,\, m) > u_i(C,\, m), \\ u_i(D,\, 0) < u_i(C,\, n-1) \end{cases}$$

That is, the payoff when $m \in [0, n-1]$ others cooperate is always higher for an individual who defects than for one who becomes the $m{+}1^{\text{st}}$ cooperator, yet universal cooperation among the n players yields a higher payoff than universal defection. In short, a dilemma game is one in which all players possess dominant strategies that lead to a deficient equilibrium.

In what follows, this notion serves primarily as a philosophical anchor: teamwork is interpreted—with some caveats—as a particular instance of such a dilemma. This interpretation motivates the need for a formal framework that can capture symmetric strategic interactions. The next subsection introduces *Aggregative Games*, an alternative formulation of normal-form games that leverages game symmetry to simplify the computation of equilibria. *Teamwork Games*, the formal model developed in Chapter 3, instantiates teamwork as a social dilemma within the aggregative-game framework.

2.1.2 *Aggregative Games*

Aggregative games are (pure-strategy) non-cooperative games in which each player's payoff can be written as a function of their own action and an *aggregate* that summarises all players' actions [Corchón, 1994, Kukushkin, 1994, Selten, 2013]. This representation simplifies analysis by avoiding the combinatorial growth of strategy profiles as the number of players increases [Jensen et al., 2018]. By aggregating interactions, one can compute best responses without tracking each opponent's action separately. Several classic economic models exhibit an aggregative structure, including the voluntary provision of public goods [Corchón, 2021], which provides the analytical foundation for Chapter 3. In this literature, emphasis is placed on the existence of equilibria, their computation, and comparative statics of Nash Equilibria (NEs) with respect to exogenous parameters [Acemoglu and Jensen, 2013, Jensen et al., 2018].

2.1.2.a Generalised Aggregative Games

Consider games with a finite set of players $\mathcal{I} = \{1, \ldots, n\}$. Unless stated otherwise, each player i has a d-dimensional action set $\mathcal{A}_i \subseteq \mathbb{R}^d$, with a typical action $a_i \in \mathcal{A}_i$. The joint action set is $\mathcal{A} \equiv \prod_{i=1}^{n} \mathcal{A}_i$, and a joint action is $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}$. For fixed i, opponents' actions are $\boldsymbol{a}_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \in \mathcal{A}_{-i}$ with $\mathcal{A}_{-i} \equiv \prod_{j \neq i} \mathcal{A}_j$. Player i's payoff is $u_i : \mathcal{A}_i \times \mathcal{A}_{-i} \to \mathbb{R}$. A (pure-strategy) Nash equilibrium is a profile $\hat{\boldsymbol{a}} = (\hat{a}_1, \ldots, \hat{a}_n)$ such that

$$u_i(\hat{a}_i, \hat{\boldsymbol{a}}_{-i}) \;\geq\; u_i(a_i, \hat{\boldsymbol{a}}_{-i}) \quad \forall a_i \in \mathcal{A}_i,\ \forall i \in \mathcal{I} \tag{2.1}$$

The distinctive feature of aggregative games is that each $u_i(\cdot)$ admits a *reduced* form depending only on a_i and a scalar aggregate $G(\boldsymbol{a})$:

$$\widetilde{u}_i\big(a_i,\, G(\boldsymbol{a})\big) \;\equiv\; u_i(a_i, \boldsymbol{a}_{-i}) \tag{2.2}$$

To build intuition, consider a setting in which each player contributes effort a_i, and the outcome depends on the total effective contribution of the group. Individual actions are first transformed through functions $h_i(a_i)$—capturing, for instance, differences in productivity or influence—and then combined through an aggregate function

$$G(\boldsymbol{a}) = H\left(\sum_{i=1}^{n} h_i(a_i)\right).$$

If H is linear, this reduces to a standard public good. More generally, concave or convex forms of H allow the model to capture diminishing returns or complementarities between contributions.

The focus here will be on generalised aggregative games, which align with the public-good structure in Chapter 3.

Definition 2.1 (Generalised Aggregative Game [Jensen et al., 2018])**.** A non-cooperative game $\Gamma = \left(\mathcal{A}_i, u_i\right)_{i\in\mathcal{I}}$ is *generalised aggregative* if:

1. There exists an aggregator $G : \mathcal{A} \to \mathbb{R}$ and reduced utilities $\widetilde{u}_i : \mathcal{A}_i \times \mathbb{R} \to \mathbb{R}$ such that $u_i(\boldsymbol{a}) = \widetilde{u}_i\big(a_i, G(\boldsymbol{a})\big)$.

2. The aggregator is additively separable: $G(\boldsymbol{a}) = H(\sum_{i=1}^{n} h_i(a_i))$, where $H : \mathbb{R} \to \mathbb{R}$ and $h_i : \mathcal{A}_i \to \mathbb{R}$ are strictly increasing.

3. The set of feasible aggregates is $\mathcal{G} \equiv \{G(\boldsymbol{a}) : \boldsymbol{a} \in \mathcal{A}\}$.

A central tool for equilibrium analysis in aggregative games is the *best-response correspondence*, which specifies how a player chooses their optimal action given the behaviour of others.

Definition 2.2 (Best-response correspondence)**.** In a generalised aggregative game, player i's best-response correspondence is

$$b_i(\boldsymbol{a}_{-i}) \;=\; \underset{a_i\in\mathcal{A}_i}{\operatorname{argmax}}\; \widetilde{u}_i\left(a_i,\, H\left(\sum_{j=1}^{n} h_j(a_j)\right)\right) \tag{2.3}$$

In an aggregative game, this optimisation problem depends on opponents' actions only through the aggregate. To make this explicit, recall that each action is first transformed through $h_j(\cdot)$ before being aggregated. It is therefore convenient to summarise the contribution of all players other than i through the transformed aggregate

$$T_{-i}(\boldsymbol{a}_{-i}) \equiv \sum_{j\neq i} h_j(a_j),$$

which represents the total (transformed) contribution of all players except i.

The best response can then be expressed as a function of this single quantity, yielding the *reduced best response*:

$$\widetilde{b}_i\big(T_{-i}\big) \;=\; \underset{a_i\in\mathcal{A}_i}{\operatorname{argmax}}\; \widetilde{u}_i\big(a_i,\, H\big(h_i(a_i) + T_{-i}\big)\big) \tag{2.4}$$

Rather than asking how a player reacts to others, one can instead ask which actions are consistent with a given level of the aggregate. This shift in perspective leads to the *replacement correspondence*:

Definition 2.3 (Replacement correspondence [Jensen et al., 2018, Acemoglu and Jensen, 2013])**.**

$$r_i(G) \equiv \left\{ a_i \ : \ a_i \in \widetilde{b}_i\big(H^{-1}(G) - h_i(a_i)\big) \right\} \tag{2.5}$$

Finally, the aggregate replacement correspondence maps each aggregate into the set of aggregates that can be sustained by individual best responses:

Definition 2.4 (Aggregate replacement correspondence)**.**

$$R(G) \equiv \left\{ G(\boldsymbol{a}) \ : \ a_i \in r_i(G) \ \ \forall i \in \mathcal{I} \right\} \tag{2.6}$$

With these definitions in place, equilibria in aggregative games can be defined based on the aggregate replacement correspondence:

Proposition 2.5 (Equilibrium aggregates [Jensen et al., 2018])**.** *An aggregate $\hat{G}$ is an equilibrium aggregate if and only if it is a fixed point of R:*

$$\hat{G} \in R(\hat{G}) \tag{2.7}$$

Intuitively, a fixed point $\hat{G} \in R(\hat{G})$ corresponds to a self-consistent aggregate: given the total contribution $\hat{G}$, each player's optimal response leads to individual actions that reproduce the same aggregate. Equilibrium is therefore characterised by consistency between individual incentives and the aggregate outcome they collectively generate.

Practical route to equilibria. The previous constructions allow the equilibrium problem to be reformulated entirely in terms of the aggregate. Rather than searching over the full action space, one can identify equilibria by working directly in the space of aggregates. In practice, equilibria in a generalised aggregative game can be computed through the following algorithm:

1. Compute $R(G)$ as in Eq. (2.6).
2. Find fixed points $\hat{G} \in R(\hat{G})$.
3. For each fixed point $\hat{G}$, recover equilibrium profiles from the product of replacement correspondences:
$$\{\hat{\boldsymbol{a}}\} = \big\{(a_1, \ldots, a_n) \in r_1(\hat{G}) \times \cdots \times r_n(\hat{G})\big\}$$

In some settings, it is useful to express equilibrium behaviour in relative terms, focusing on how the aggregate is divided among players rather than on absolute contributions. This leads to the notion of share correspondences:

Definition 2.6 (Share correspondence)**.** Let Γ be a generalised aggregative game with replacement correspondence $r_i(G)$ for each i. For any $G > 0$, player i's share correspondence $s_i : \mathcal{G} \to [0,1]$ is

$$s_i(G) \;=\; \left\{ \frac{h_i(a_i)}{H^{-1}(G)} \;:\; a_i \in r_i(G) \right\} \tag{2.8}$$

Whenever the replacement correspondence r_i is single-valued, so is s_i. Writing $S(G) \equiv \sum_{j=1}^{n} s_j(G)$, an equilibrium with aggregate $\hat{G}$ satisfies

$$S(\hat{G}) \;=\; \sum_{j=1}^{n} s_j(\hat{G}) \;=\; 1 \tag{2.9}$$

(see Jensen et al. [2018] and Hartley and Cornes [2000] for further details).

Put differently, the share correspondence specifies the fraction of the total effective contribution attributable to each player. In equilibrium, these shares must sum to one, reflecting the fact that the aggregate is fully accounted for by individual contributions.

With the core elements of aggregative games in place, the next subsection turns to a particular class of interest: General Public-Good Games.

2.1.2.b General Public-Good Games

Several classic economic models exhibit an aggregative structure, including the voluntary provision of public goods [Corchón, 2021], which provides a natural foundation for modelling situations in which individual contributions combine into a collective outcome. The use of aggregative games for voluntary public good provision goes back to Cornes and Hartley [2007a], who adapted the canonical model of Bergstrom et al. [1992] (public goods as the unweighted sum of individual contributions) to the aggregative framework. Earlier, Hirshleifer [1983] had already emphasised two empirically relevant technologies not captured by simple summation: *best-shot* (team output equals the maximum individual contribution) and *weakest-link* (team output equals the minimum contribution). Subsequent work considers more general, non-summative technologies [Cornes, 1993, Cornes and Hartley, 2007b]; see also Cornes [2016] and Buchholz and Rübbelke [2017].

A *general public-good game* is a generalised aggregative game in which the aggregator (public good) is a Constant Elasticity of Substitution (CES) function of contributions—*a.k.a.* gifts:

$$G \;=\; \Big(\sum_{i=1}^{n} \beta_i \, g_i^{\rho} \Big)^{1/\rho} \tag{2.10}$$

where $\rho \in \mathbb{R}^*$ is the substitution parameter, $g_i \in \mathbb{R}_0^+$ is player i's contribution, and $\beta_i > 0$ is a weight.[1] In the notation of Definition 2.1, $G = H(\sum_{i=1}^n h_i(g_i))$ with $H(z) = z^{1/\rho}$ and $h_i(g_i) = \beta_i g_i^\rho$.

Finally, an assumption regarding the preferences of players in a general public-good game will be useful in the modelling framework developed later:

Assumption 2.7 (Well-behaved preferences)**.** In a general public-good game, the reduced utility of player i depends on a private good x_i and the public good G:

$$\widetilde{u}_i(x_i, G)$$

The function $\widetilde{u}_i$ is assumed to be continuously differentiable, strictly increasing, and strictly quasi-concave in both arguments. Moreover, both the private and the public goods are normal for every player, in the sense that their demand increases with income under these preferences.

These definitions provide the formal backbone for the models developed in the next chapter. The following section complements this framework by introducing the dynamic perspective needed to study how agents adapt their behaviour over time within a MA-MAB environment.

2.2 *Learning in Games*

Classical Game Theory (GT) addresses questions about prescribed behaviour under the assumption of fully rational players. In contrast, the focus of the literature on learning in games is on expected behaviour: how agents adapt and what outcomes emerge when rationality is limited.

In the context of this thesis, the central question is not how cooperation should be prescribed, but how it emerges among self-interested agents engaged in a shared task without enforcement. In studying such behaviour, this work adopts a perspective in which theory and simulation are used in tandem. From the simulation side, agents are modelled as boundedly

[1] $\mathbb{R}^+$ denotes the set of positive real numbers, $\mathbb{R}_0^+$ the set of non-negative real numbers including zero, and $\mathbb{R}^*$ the set of all real numbers except zero.

rational, adapting their behaviour through experience and limited information. This perspective motivates the use of learning dynamics to connect equilibrium analysis with observed behaviour.

Decision-makers in a team can be viewed as engaging in a search process over actions that yield favourable outcomes, balancing the discovery of new options against the use of known ones. This perspective is commonly formalised through multi-armed bandit (MAB) models, in which agents repeatedly select among actions with uncertain rewards [Reverdy et al., 2014, Maghsudi and Hossain, 2016, Wu et al., 2018].

In the classical bandit problem, decision-making is reduced to a single-agent setting with stochastic payoffs. Each action (arm) a is associated with an unknown expected reward $q(a)$, and selecting an action yields a noisy observation of this value [Robbins, 1952, Auer et al., 2002, Sutton and Barto, 2018]. The agent must learn these values over time in order to act effectively.

To learn it, each agent maintains an estimate of the value of each action a. Let $Q^N(a)$ denote the estimate after N updates based on observations of rewards obtained from selecting a. Given the N^{th} observed reward R^N, the estimate is updated incrementally as:

$$Q^{N+1}(a) = Q^N(a) + l_r^N\big(R^N - Q^N(a)\big), \tag{2.11}$$

where $l_r^N \in (0, 1]$ denotes the learning rate. Higher learning rates place more weight on recent observations, enabling faster adaptation, while lower values favour stability.

A decaying learning rate ensuring asymptotic stochastic convergence is assumed to satisfy:

$$\sum_{N=1}^{\infty} l_r^N = \infty, \qquad \sum_{N=1}^{\infty} (l_r^N)^2 < \infty.$$

This formulation corresponds to a standard Q-learning update rule in a stateless setting. Its use is motivated not only by its simplicity, but also by the role of the learned Q-values in defining behavioural policies. In particular, these value estimates can be directly used within a Boltzmann (softmax) action-selection rule, which induces smooth best responses [Tokic and Palm, 2011, Chapman et al., 2013].

Under Boltzmann action selection [Sutton and Barto, 2018], the probability of selecting action a at time t is given by:

$$\mathbb{P}_t(a) = \frac{\exp\big(Q^t(a)/\tau\big)}{\sum_{a'} \exp\big(Q^t(a')/\tau\big)}, \tag{2.12}$$

where $\tau > 0$ is the temperature parameter controlling the exploration–exploitation trade-off. Low values of τ lead to near-deterministic behaviour, while higher values induce greater exploration.

Under these dynamics, Nash equilibria are no longer strict fixed points. Instead, the adaptive process induced by Boltzmann action selection may converge—under suitable conditions—to stationary distributions over actions. These stationary points, referred to as *Nash distributions* [Leslie and Collins, 2005], characterise the long-run outcome of the learning dynamics.

Crucially, Nash distributions are mathematically equivalent to logit Quantal Response Equilibrium (QRE) [McKelvey and Palfrey, 1998], a solution concept capturing stable behaviour under bounded rationality. As the temperature parameter $\tau \to 0$, QRE approach Nash equilibria [Govindan et al., 2003].

In this way, Q-learning provides a principled bridge between adaptive behaviour in stateless settings and equilibrium concepts, allowing the outcomes of learning dynamics to be interpreted in formal game-theoretic terms. In particular, it enables us to assess whether the outcomes observed in simulation align with the equilibrium structures predicted by the underlying game-theoretic models in settings where cooperation must emerge from the interaction of self-interested agents.

2.3 Markov Games and Multi-Agent Reinforcement Learning

To extend teamwork analysis into temporally extended settings, sequential interaction is modelled using Markov Games—the multi-agent generalisation of Markov Decision Processes (MDPs) Shapley [1953a], Littman [1994]. Following the notation of Sutton and Barto [2018], realised quantities are written in uppercase throughout this dissertation, and functions in lowercase.

Consider an n-agent partially observable Markov Game $\mathcal{M}$ defined over a finite set of states $\mathcal{S}$, where $\mathcal{A}_i$ denotes agent i's action set. The observation function O specifies each agent's d-dimensional view on the state space:

$$O : \mathcal{S} \times \{1, 2, \ldots, n\} \longrightarrow \mathbb{R}^d \tag{2.13}$$

Thus, agent i's observation space is

$$\mathcal{O}_i = \{O(s, i) \mid s \in \mathcal{S}\}$$

A joint action of the agents is an element $\boldsymbol{a} \in \mathcal{A} \stackrel{\text{def}}{=} \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_n$, where all agents' action sets are identical. At each timestep, each agent i chooses an action a_i^t. Agents act simultaneously, so given an environment state s^t and a joint action $\boldsymbol{a}^t$, the state changes to s^{t+1} following the transition function, which is taken as deterministic:

$$T : \mathcal{S} \times \mathcal{A} \longrightarrow \mathcal{S} \tag{2.14}$$

As a result of this state transition, agent i receives a reward at time $t+1$ given by the reward function:

$$r_i : \mathcal{S} \times \mathcal{A} \longrightarrow \mathbb{R} \tag{2.15}$$

denoted by $R_i^{t+1} = r_i(s^t, \boldsymbol{a}^t)$. Learning is decentralised: each agent optimises its own policy based solely on its local experience, without access to the observations or rewards of others. The result of learning is a stochastic policy $\pi_i : \mathcal{O}_i \longrightarrow \Delta(\mathcal{A}_i)$ (written $\pi_i(a_i^t \mid o_i^t)$). To represent joint action selection, write $\boldsymbol{a}^t \sim \boldsymbol{\pi}$. It is assumed that agents act independently conditioned on their local observations, so the joint policy factorises as:

$$\boldsymbol{\pi}(\cdot \mid \boldsymbol{o}^t) \stackrel{\text{def}}{=} \big(\pi_1(\cdot \mid o_1^t), \ldots, \pi_n(\cdot \mid o_n^t)\big) \tag{2.16}$$

All episodes have the same duration H, and agent i's goal is to maximise its total expected discounted future reward:

$$v_{i,\boldsymbol{\pi}}(S_0) \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{k=0}^{H-1} \gamma^k \, R_i(S^k, \mathbf{A}^k) \;\middle|\; \mathbf{A}^k \sim \boldsymbol{\pi}(\cdot \mid \mathbf{O}^k),\ S^{k+1} \sim T(\cdot \mid S^k, \mathbf{A}^k)\right] \tag{2.17}$$

The next section explores the algorithms used in this dissertation to train independent agents under this Markov Game setting.

2.4 Deep Reinforcement Learning

The formalism of Markov Games provides a framework to describe multi-agent interaction over time, but it does not prescribe how agents acquire optimal policies in practice. In high-dimensional or partially observable environments, analytical solutions are intractable, and explicit representations of value functions or policies become computationally prohibitive. Deep

Reinforcement Learning (DRL) addresses this challenge by using neural networks as function approximators for policy and value estimation, enabling agents to learn effective behaviours directly from experience.

Within this dissertation, DRL provides the computational backbone for implementing the policies π_i defined in the previous section. Each agent is modelled as an independent learner whose policy and value functions are parameterised by neural networks and trained via gradient-based optimisation. Specifically, Chapter 4 employs an independent multi-agent variant of Proximal Policy Optimization (PPO), following recent evidence on the effectiveness of PPO-based methods in cooperative multi-agent settings [Yu et al., 2022].

Both applications of neural networks in DRL—value and policy estimation—follow the principle of stochastic gradient descent (or ascent). The underlying idea is to iteratively adjust the parameters of the function being learned in the direction that improves a given performance metric. In value approximation, the objective is a quadratic loss that measures the discrepancy between the estimated and target values obtained from experience. In contrast, policy gradient methods update the parameters of a stochastic policy in the direction that maximises the expected return, thereby increasing the probability of selecting rewarding actions over time.

The next section formalises this principle by introducing the policy gradient theorem, which provides the mathematical foundation for actor–critic algorithms such as PPO.

2.4.1 Policy Gradient Methods

For simplicity, this subsection considers a single agent and omits agent subscripts. All results generalise directly to the multi-agent case introduced earlier.

Policy gradient methods learn a parameterised policy that selects actions without directly consulting a value function. Although a value function may also be used to guide the learning of policy parameters, it is not required for action selection [Williams, 1992, Sutton et al., 1999]. Following the notation of Sutton and Barto [2018], the basic components of these methods are introduced below.

Let $\boldsymbol{\theta} \in \mathbb{R}^{d'}$ denote the policy parameter vector. The policy of the agent is defined as the conditional probability distribution

$$\pi(a \mid s, \boldsymbol{\theta}) = \mathbb{P}\{A^t = a \mid S^t = s, \boldsymbol{\theta}^t = \boldsymbol{\theta}\}$$

The objective of policy gradient methods is to learn the parameter vector $\boldsymbol{\theta}$ that induces a rewarding policy. Formally, given a scalar performance measure $J(\boldsymbol{\theta})$, these methods aim to maximise performance by progressively updating $\boldsymbol{\theta}$ in the direction of a **stochastic approximation of the gradient** of $J(\boldsymbol{\theta})$, denoted $\widehat{\nabla J(\boldsymbol{\theta})} \in \mathbb{R}^{d'}$:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha_\theta \, \widehat{\nabla J(\boldsymbol{\theta}_t)} \tag{2.18}$$

When the action space is discrete and not prohibitively large, a common parameterisation of the policy is a softmax distribution:

$$\pi(a \mid s, \boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\exp\big(h(s, a, \boldsymbol{\theta})\big)}{\sum_b \exp\big(h(s, b, \boldsymbol{\theta})\big)}$$

where $h(s, a, \boldsymbol{\theta}) \in \mathbb{R}$ denotes the *numerical preference* associated with the state–action pair (s, a). This is the parameterisation employed in Chapter 4, where the preferences correspond to the unnormalised logit outputs of a neural network with parameters $\boldsymbol{\theta}$.

For episodic settings—those with a predefined time horizon—the performance measure is typically defined as

$$J(\boldsymbol{\theta}) \stackrel{\text{def}}{=} v_{\pi_{\boldsymbol{\theta}}}(S^0)$$

that is, the expected return when following policy $\pi_{\boldsymbol{\theta}}$ from the initial state S^0. In this case, the *policy gradient theorem* establishes that

$$\nabla J(\boldsymbol{\theta}) \propto \mathbb{E}_{\pi_{\boldsymbol{\theta}}}\left[\sum_a q_{\pi_{\boldsymbol{\theta}}}(S^t, a) \, \nabla_{\boldsymbol{\theta}} \pi(a \mid S^t, \boldsymbol{\theta})\right] \tag{2.19}$$

In words, the theorem states that the gradient of $J(\boldsymbol{\theta})$ is proportional to the expected contribution of each state–action pair to the return, weighted by how the probability of selecting that action changes under infinitesimal variations of $\boldsymbol{\theta}$.

More concretely, for each state S^t visited under the policy, the value of each action a under that policy ($q_{\pi_{\boldsymbol{\theta}}}(S^t, a)$) is multiplied by the sensitivity of the policy's probability of selecting a in S^t with respect to the parameter vector ($\nabla_{\boldsymbol{\theta}} \pi(a \mid S^t, \boldsymbol{\theta})$). Summing these contributions—weighted by state visitation frequencies—yields the direction in parameter space that most increases expected performance.

In practice, policy gradient methods **sample** this expectation. Multiplying Eq. (2.19) by

$$\frac{\pi(A^t \mid S^t, \boldsymbol{\theta})}{\pi(A^t \mid S^t, \boldsymbol{\theta})},$$

replacing a by the sampled action $A^t \sim \pi_{\boldsymbol{\theta}}$, and using the identity

$$\mathbb{E}_{\pi_{\boldsymbol{\theta}}}[G^t \mid S^t, A^t] = q_{\pi_{\boldsymbol{\theta}}}(S^t, A^t)$$

yields the well-known expression [Sutton and Barto, 2018]:

$$\mathbb{E}_{\pi_{\boldsymbol{\theta}}}\left[G^t \, \frac{\nabla_{\boldsymbol{\theta}} \pi(A^t \mid S^t, \boldsymbol{\theta})}{\pi(A^t \mid S^t, \boldsymbol{\theta})} \right]$$

where G^t denotes the empirical return. Substituting this into Eq. (2.18) yields the stochastic ascent rule:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha_\theta \, G^t \frac{\nabla_{\boldsymbol{\theta}} \pi(A^t \mid S^t, \boldsymbol{\theta})}{\pi(A^t \mid S^t, \boldsymbol{\theta})}$$

This update rule is known as the *REINFORCE* algorithm. It constitutes the most direct application of the policy gradient theorem to stochastic approximation and serves as a baseline for all policy gradient methods. Subsequent approaches aim to reduce the bias and variance of this estimator by incorporating baselines or more structured advantage estimators, which are addressed next.

2.4.2 *Generalised Advantage Estimation (GAE)*

Variance reduction is a central concern in policy gradient methods, given the stochastic nature of the policy and the delayed influence of actions on rewards. The *Generalised Advantage Estimation (GAE)*, introduced by Schulman et al. [2015], provides a smoothed, lower-variance estimator of the advantage function, which is a key component in modern actor–critic algorithms such as PPO.

The *advantage function* $A_\pi(s^t, a^t)$ quantifies how much better (or worse) taking action a^t in state s^t is relative to the expected return under the current policy π:

$$A_\pi(s^t, a^t) = q_\pi(s^t, a^t) - v_\pi(s^t)$$

GAE estimates this quantity using a weighted sum of temporal-difference (TD) residuals. The TD residual (or TD error) at time step t is defined as

$$\delta^t = R^t + \gamma V(S^{t+1}) - V(S^t) \tag{2.20}$$

where the capital $V(\cdot)$ denotes a learned approximation to the true value function $v_\pi(\cdot)$.

The generalised advantage estimate at time t is then given by

$$\hat{A}^t = \sum_{l=0}^{T-t-1} (\gamma\lambda)^l \, \delta^{t+l} \tag{2.21}$$

with

- $\gamma \in [0, 1)$ the discount factor,
- $\lambda \in [0, 1]$ the smoothing (or bias–variance trade-off) parameter,
- T the episode length or rollout horizon.

The parameter λ interpolates between:

- high-bias, low-variance estimates when $\lambda \approx 0$, corresponding to one-step TD,
- low-bias, high-variance estimates when $\lambda \approx 1$, corresponding to Monte Carlo returns.

In practice, GAE is computed efficiently through recursive formulations and vectorised operations over entire rollouts. It has become a standard component of modern actor–critic algorithms due to its effectiveness in stabilising policy learning while retaining informative gradient signals.

2.4.3 Proximal Policy Optimization (PPO)

Proximal Policy Optimization [Schulman et al., 2017] is an on-policy policy gradient method that achieves stable and efficient learning by constraining how much the policy may change at each update. It can be regarded as a simplified alternative to Trust Region Policy Optimisation (TRPO), providing comparable stability without the complexity of constrained optimisation.

Although PPO allows several gradient updates over the same batch of data, it remains on-policy, as the collected trajectories are generated by the most recent version of the policy, denoted $\pi_{\boldsymbol{\theta}_{\text{old}}}$. To account for the small mismatch between the updated policy $\pi_{\boldsymbol{\theta}}$ and the data-generating policy $\pi_{\boldsymbol{\theta}_{\text{old}}}$, PPO employs importance sampling. The surrogate objective can therefore be written as

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\pi_{\boldsymbol{\theta}_{\text{old}}}} \left[r(\boldsymbol{\theta}) \, \hat{A}^t \right] \tag{2.22}$$

where the importance ratio is

$$r(\boldsymbol{\theta}) = \frac{\pi(A^t \mid S^t, \boldsymbol{\theta})}{\pi(A^t \mid S^t, \boldsymbol{\theta}_{\text{old}})}$$

While the expected value of Eq. (2.22) is consistent with the original on-policy objective, in practice the estimator may exhibit high variance when the new and old policies diverge substantially. Large deviations in the importance ratio amplify the contribution of a few samples, leading to unstable or inefficient updates.

To mitigate this effect, PPO applies a clipping mechanism to the surrogate loss, restricting parameter updates within a trust region and balancing stability with learning progress. The clipped objective is defined as

$$L^{\text{CLIP}}(\boldsymbol{\theta}) = \mathbb{E}_{\pi_{\theta_{\text{old}}}}\left[\min\Big(r(\boldsymbol{\theta})\,\hat{A}^t,\, \text{clip}(r(\boldsymbol{\theta}),\, 1-\epsilon,\, 1+\epsilon)\,\hat{A}^t\Big)\right] \tag{2.23}$$

where $\epsilon > 0$ defines the allowable deviation from the previous policy. This formulation discourages excessively large policy updates, promoting more stable learning in practice.

Together, these methods ground the computational theory of cooperation to come. From extrinsic to intrinsic motives, from tangible constraints to intangible identities, from structure to learning, the next chapters explore how alignment mechanisms—spanning task structure, interaction and identity—can make cooperation the optimal policy for utility-maximising agents.



3 To Work, or Not to Work: Teamwork Games

We cannot enter into alliances until we are acquainted with the designs of our neighbors.

Sun Tzu

3.1 Introduction: The Teamwork Dilemma 38
3.2 Social-Psychological Theories of Group Performance 39
3.3 Game-Theoretic and Computational Perspectives on Teamwork 43
3.4 Teamwork Games 46
3.5 CollabQuest: A Multi-Agent Multi-Armed Bandit for Teamwork 67
3.6 Experiments 70

Having introduced the mathematical and computational tools underpinning this dissertation in Chapter 2, this chapter turns to the first team alignment mechanism: *structure*. By *structure* we refer to the tangible constraints that can be externally engineered to shape cooperation within teams.

More fundamentally, this chapter takes a step towards the central objective of the thesis: the **formalisation of psychological models of teamwork into computational frameworks**. In doing so, it develops representations that are not restricted to human-only settings, but are applicable across human-human, human-agent, and agent-agent interactions.

Specifically, the chapter introduces *Teamwork Games* [López de Aberasturi Gómez et al., 2025], a formal framework that translates core constructs from Steiner's *model of group performance* into a game-theoretic setting. Rather than proposing an alternative to existing models on purely game-theoretic grounds, the contribution lies in making psychologically grounded variables—such as task interdependence, team composition, and evaluative thresholds—amenable to formal analysis and computational instantiation.

Within this framework, Steiner's key constructs are given a precise operational interpretation. The notion of *task type* is mapped to different forms of *task interdependence*, implemented through the aggregator that links individual contributions to the collective outcome. *Team composition* is represented by the distribution of capabilities—or expertise—across agents, while the *evaluation function* formalises the *evaluative threshold* that determines whether a team's performance is deemed successful.

By embedding these psychologically grounded elements into a non-cooperative game, the framework captures voluntary contribution decisions by self-interested agents while preserving the structural features identified in group psychology. In this sense, its novelty does not lie in outperforming existing models, but in providing a unified formal language through which psychological theories of teamwork can be analysed, simulated, and extended.

To close the theory-implementation loop, the second part of the chapter develops a Multi-Agent Multi-Armed Bandit (MA-MAB) environment to examine whether utility-maximising agents, operating under these psychologically grounded structures, reproduce the strategies predicted by the analytical solutions through their learned policies.

Chapter Contributions. As the first step in the analysis of alignment mechanisms, this chapter advances the modelling of teamwork without enforced cooperation by introducing a new class of games and analysing their equilibria and learnability.

(C1) **Introduction of Teamwork Games:** A novel class of aggregative games is formalised to capture teamwork settings where agents decide their **voluntary contributions** to a shared task. Building on Steiner's *theory of group performance*, the model incorporates variables often overlooked in game-theoretic treatments of teamwork, such as *team roles*, *task type*, *expertise*, *preferences*, and *evaluation functions*.

(C2) **Equilibrium Characterisation:** The Nash Equilibria (NEs) of Teamwork Games are analytically characterised, showing how they extend the equilibria of general public good games. The analysis highlights the influence of teamwork-specific elements such as evaluation functions and agent expertise.

(C3) **Learning Framework with Multi-Agent Bandits:** A Multi-Agent Multi-Armed Bandit (MA-MAB) framework is proposed in which agents learn contribution strategies that approximate the equilibria of Teamwork Games. Contributions are conceptualised as bandit arms, and agents adapt their behaviour using Boltzmann exploration.

(C4) **Empirical Validation of Equilibrium Convergence:** The proposed framework is shown to reliably converge to approximations of the analytically derived equilibria. Across more than 140 experiments, the model exhibits a near-perfect fit between predicted and observed team productivity.

(C5) **Analysis of Teamwork Determinants:** An empirical analysis is conducted of how task type, evaluation functions, and agent expertise shape group products in agent simulations. The findings align with established theories in social psychology, providing computational evidence for Steiner's hypotheses about *task interdependence*, *evaluative thresholds*, and *team composition*.

(C6) **Extension to Non-Analytical Settings:** The analysis is extended to tasks with discontinuous evaluation functions, showing that the framework generalises to settings lacking closed-form solutions.

Together, these contributions establish Teamwork Games as a tractable yet expressive model for analysing voluntary teamwork, setting the foundation for the dynamic extensions developed in Chapter 4.

3.1 Introduction: The Teamwork Dilemma

Anna is taking an algebra course as part of her undergraduate degree in engineering. One of the requirements is a group assignment: she and three classmates must collectively solve a mathematical proof. Only one solution is submitted on behalf of the entire group, and all members receive the same grade. Passing requires completing the proof, and there are several ways to approach it. Depending on the rigour and originality of the chosen method, they may earn extra points.

Each teammate has different commitments and interests. Some are enrolled in a demanding deep learning course with a project due the same day, while Anna attends ballet classes she does not want to miss. Beyond these differences in schedules, the team also varies in ability. Emanuele has struggled with algebra, failing the last three exams. Subham is capable and quick, but often distracted. Sanutha has become diligent and steady, with no competing deadlines. Anna, by contrast, excels at mathematics and has a proven record of success.

The group has decided to meet on the afternoon before the deadline to complete the assignment. With limited time to coordinate and no opportunity to revise decisions, each member must choose how much effort to contribute. Faced with these differences, a set of natural questions arises: how much effort should each teammate contribute? Should they aim only to pass, or also pursue additional marks for elegance and originality? Should effort be distributed evenly, or concentrated in the most capable members? More fundamentally, what is the rational course of action for each individual, given the characteristics of the others?

This simple scenario captures the core elements of teamwork as a strategic problem. The outcome is shared, resembling a public good, yet contributions are individual and costly. Team members differ in ability and motivation, and the final result depends on how their efforts combine. In such settings, cooperation cannot be taken for granted; rather, it must emerge from the interaction between individual incentives and task structure.

These questions lie at the heart of social-psychological theories of group performance. Steiner, whose work will play a central role in this chapter, argued that the productivity of a team depends not only on the motivation of its members, but also on the structure of the task and the way individual contributions combine to produce collective outcomes. His framework distinguishes between different forms of task interdependence and emphasises the role of team composition and evaluative criteria in shaping group performance.

Despite its richness, this body of work has remained largely conceptual and empirical, lacking a formal representation that allows for systematic analysis of strategic behaviour. This chapter addresses this gap by translating these psychological constructs into a computational and game-theoretic framework.

Specifically, it introduces *Teamwork Games* [López de Aberasturi Gómez et al., 2025], a class of non-cooperative games in which self-interested agents decide how much to contribute to a shared task under different forms of task interdependence, team heterogeneity, and evaluation rules. In doing so, the chapter formalises key elements of Steiner's theory—task structure, team composition, and evaluative thresholds—in a way that makes them amenable to analytical and computational study.

The next section reviews the social-psychological foundations that underpin this formalisation.

3.2 Social-Psychological Theories of Group Performance

In social psychology, teamwork is often understood as a form of social dilemma [Dawes, 1980, Bowles and Gintis, 2003, Karau and Williams, 2001]. In particular, the public good dilemma has been extensively studied through the lens of interdependence theory [De Cremer, 1999, Balliet and Van Lange, 2013]. Originally inspired by Lewin [1948] and further formalised by Thibaut and Kelley [1959], interdependence theory conceptualises groups as dynamic entities shaped by the mutual *dependence* of their members.

Within this school of thought, Steiner proposed a theory of group performance that explicitly conceives groups in terms of interdependence [Steiner, 1972]. More recently, Nijstad [Nijstad, 2009] reformulated Steiner's ideas in light of the empirical evidence accumulated over the subsequent three decades. This section delineates the basic constructs common to both models, paving the way for their formalisation in Section 3.4. Before turning to these models, it is useful to clarify what *interdependence* means in this context. Following Wageman [2014], two main forms can be distinguished:

- **Structural interdependence**, referring to contextual features that link team members' work. This includes **task interdependence** (the extent to which members must rely on

one another's information, resources, or expertise) and **outcome interdependence** (the degree to which individual rewards depend on collective performance). These two forms interact in complex ways, but a consistent finding is that highly interdependent tasks require equally interdependent outcomes to sustain good performance.

- **Behavioural interdependence**, referring to the actual interactions team members engage in when working together. While often shaped by structural conditions, it also depends strongly on individual differences and cultural context.

In this chapter, the two main components of structural interdependence are explicitly modelled: *task* and *outcome* interdependence. As detailed in the next section, task interdependence is instantiated through the *task type* and represented via a Constant Elasticity of Substitution (CES)-based parameterisation, which enables precise control over how agents' efforts combine—whether teams benefit from redundancy, bottleneck avoidance, or specialisation. Outcome interdependence is encoded in the utility (reward) functions of the game within the multi-agent framework, acting as the dial that tunes the degree of alignment between individual and collective outcomes. In this setting, rewards are only partially congruent with goal attainment and moderately dependent on co-workers' performance, giving the environment a distinctly mixed-motive character.

While acknowledging that team performance encompasses more than mere productivity [McGrath, 1964], this chapter focuses on productivity as the primary metric of teamwork [Campion et al., 1993, Kerr and Tindale, 2004]. Accordingly, attention is restricted to the internal variables within a group that affect—and best predict—group productivity (*i.e.*, a closed-group paradigm [Casal et al., 2011]).

3.2.1 Steiner's Model of Group Performance

It is hard to overstate the impact that Steiner's model of group performance had on social-psychology. In identifying the four key elements within a group that influence productivity, it set the foundations of current teamwork paradigms:

1. **Individuals' profiles:** Individuals contribute various resources to the task, ranging from skills and expertise levels to effort and personality traits such as predisposition towards work—*i.e.*, preferences.

2. **Task Type:** Steiner categorised tasks based on how individual contributions combine to shape group performance (see Table 3.1). In addition, he distinguished between **unitary** tasks—those that cannot be subdivided into smaller subtasks and require a single skill—and **divisible** tasks, which can be broken down into subtasks performable by different individuals. For simplicity, unitary tasks are the ones analysed here, leaving divisible tasks to future work.

3. **Group Interaction Processes:** Team members' preferences and intentions drive their action choices. Introducing an evaluation system that assesses group production is likely to alter those preferences, reshaping interdependencies among members and thereby affecting productivity [Wageman, 2014].

4. **Group Outcome:** Finally, individual decisions are influenced by the foreseen group's outcome, particularly whether the group's outcome will satisfy the individuals' preferences or not. Likewise, group product feeds back into the group in future interactions.

Another important contribution by Steiner was his distinction between a group's *potential performance*, representing its capabilities, and its *actual performance*, reflecting what is achieved in practice. Actual performance falls short of potential due to "process losses", most notably in motivation and coordination.

- **Motivation losses** include phenomena such as *social loafing*—the tendency to reduce effort when individual contributions are not identifiable [Kerr, 1983, Karau and Williams, 1993]—and the "*sucker effect*", by which members lower their effort to avoid being exploited. Conversely, *social compensation* occurs when individuals exert greater effort to offset perceived weaknesses among teammates, a phenomenon known as the Köhler effect [Smith, 2022].

- **Coordination losses** arise when teams fail to combine efforts optimally; for example, in disjunctive tasks where the correct solution is difficult to identify, groups may fall short of their potential. However, when tasks are highly demonstrable, group performance can even exceed that expected from the "typical" group member, which is known as *synergy* [Larson Jr, 2013].

These insights on motivation and coordination losses highlight mechanisms that reduce and increase actual group performance relative to potential performance, and they later inform how equilibrium outcomes are interpreted in the model.

Table 3.1: **Steiner's task taxonomy** based on how individual contributions combine into a group product [Steiner, 1972].

Task Type	Characteristics	Example
Additive	Group success results from the sum of individual efforts, with each member's contribution directly adding to the group product.	A team of researchers preparing a joint report, where each person writes a section that contributes to the final document; or a tug-of-war contest, where the total force exerted is the sum of all participants' efforts.
Conjunctive	Group performance is determined by the weakest or least effective member.	A ballet performance where the least synchronised dancer affects the overall impression of the ensemble; or a mountaineering expedition that can only progress at the pace of its slowest climber.
Disjunctive	Group performance is determined by the strongest or most effective member.	A research group in which the breakthrough idea of one scientist secures success for the entire team; or a quiz competition where only one team member needs to know the correct answer for the group to win the round.

3.2.2 Nijstad's Model of Group Performance

Over the years, Steiner's model has been refined and extended. The revision examined here is that of Nijstad [2009], which explicitly distinguishes between individual- and group-level factors influencing performance, and maps the feedback loops among them (Figure 3.1).

At the individual level (left side of the figure), team members contribute their knowledge, skills, and abilities (KSAs), along with motives and personal characteristics that shape their engagement. Individuals can also affect the broader context in which the group operates, while the context in turn influences individual behaviour (Arrow 1).

The group-level elements (right side of the figure) encompass group interactions and their group product. Distinguishing the two levels clarifies top-down (group-to-individual) and bottom-up (individual-to-group) processes. Individual contributions feed into group-level processes (bottom-up, Arrow 2), while the behaviour of other group members feeds back into individual decision-making (top-down, Arrow 3). These social dynamics ultimately shape the group's output (Arrow 4), which may also influence the group's external environment (Arrow 5).

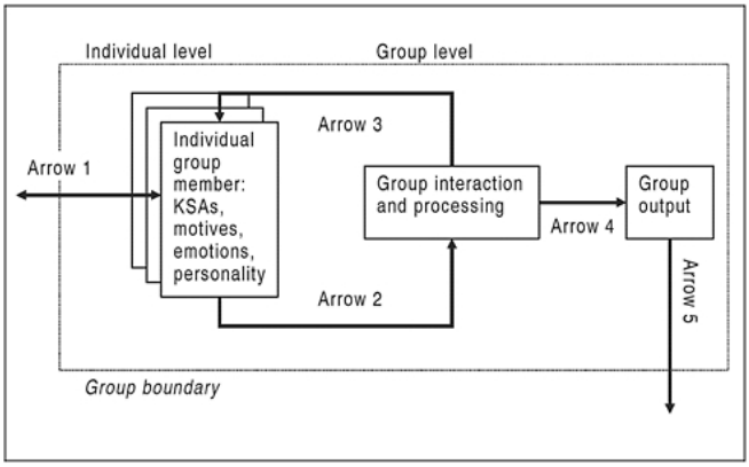


Figure 3.1: **Nijstad's model of group performance**. Individual- and group-level variables interact through processes that condition team performance. In turn, team performance feeds back into individual decision-making. *Source:* [Nijstad, 2009].

Notice that in contrast to Steiner's model, this framework adopts a more dynamic perspective by incorporating bidirectional influences between the group and its external context (Arrow 1).

3.3 *Game-Theoretic and Computational Perspectives on Teamwork*

From a classical game theoretic perspective, teamwork has been broadly modelled within cooperative game theory, assuming that team members share common goals and rewards and face no strategic conflict [Von Neumann and Morgenstern, 1947, Elkind and Rothe, 2016]. Classical solution concepts, such as the Shapley Value, distribute the total utility generated by the team in a fair way [Shapley, 1953b, Branzei et al., 2008, Maschler et al., 2020, Chalkiadakis et al., 2022]. This approach works well for closely knit teams where cooperation is assured, but in typical human teamwork, securing binding collaboration agreements is often impossible. As a result, cooperative models offer limited insight into performance when cooperation is voluntary and fragile.

Similarly, research in Multi-Agent Systems (MAS) [Weiss, 1999, Shoham and Leyton-Brown, 2008, Wooldridge, 2009] has focused extensively on teamwork architectures where agents share

common goals [Shoham, 1993, Shoham and Tennenholtz, 1995, Tambe, 1997, Pynadath et al., 1999, Wooldridge, 2009, Dunin-Keplicz and Verbrugge, 2010]. These approaches address prescriptive questions such as: *How should agents be designed to cooperate effectively, and how should their interactions be coordinated to maximise group performance?*. By contrast, predicting the performance of human teams—or more generally of teams without guaranteed cooperation—requires a descriptive perspective. This leads to a different question: *What outcomes emerge when self-interested agents interact under varying conditions of interdependence, task type, and team composition?*.

Beyond classical cooperative formulations, a related line of work has explicitly addressed teamwork as a problem of aligning self-interested agents through reward design. In particular, the Collective Intelligence (COIN) framework introduced by Wolpert and Tumer [1999, 2001] models multi-agent systems in which individual agents optimise private utility functions that are designed to promote desirable global outcomes. Within this framework, constructs such as the Wonderful Life Utility (WLU) aim to reduce negative side-effects of decentralised optimisation and improve coordination without requiring explicit cooperation.

This perspective shares with the present work the view that agents cannot be assumed to act cooperatively by default. However, it differs in its objective: rather than designing individual reward functions to enforce alignment, the approach developed here focuses on how structural properties of the task—such as interdependence, evaluation criteria, and team composition—shape incentives and give rise to cooperative or non-cooperative behaviour.

Recognising that team members may shirk collective responsibilities and maximise their own utility by free-riding calls for the use of non-cooperative Game Theory (GT). This perspective is typically studied through mixed games, especially the voluntary provision of public good games [Bergstrom et al., 1992, Cornes and Sandler, 1996, Myatt and Wallace, 2009, Carpenter et al., 2009, Baños and Rojas, 2019]. In its most basic version, participants in a public goods game privately decide how many tokens to contribute to a communal pool. The tokens are aggregated into a public good whose benefits are evenly distributed, while each participant retains the tokens not contributed. Teamwork can thus be seen as a public good—presenting a dilemma between contributing to the collective task or exploiting others' efforts (social loafing/free-riding). Even when cooperation is beneficial, individuals may differ in their preferences for the common good, and practical constraints often fail to prevent free-riding [Kameda et al., 1992]. This perspective provides the starting point for the present work,

where teamwork is modelled as a mixed-motive setting in which individual contributions to a shared task exhibit the characteristics of a public good.

Chapter 2 showed how the field of learning in games offers a clue: it examines how players who are less than fully rational adapt their strategies over time, and what obstacles may arise in reaching equilibrium. Reinforcement Learning (RL) provides a natural tool in this context. Although RL agents rarely converge to a Nash equilibrium, they can approximate Nash distributions by using *smooth best responses*. In these, each player selects actions with probabilities biased toward higher payoffs rather than deterministically choosing the best response (see Chapter 2 for greater detail).

Beyond their mathematical appeal, RL architectures also capture agent-level constructs such as beliefs, preferences, intentions, and actions [Jara-Ettinger, 2019]. When such agents interact in a multi-agent system, their adaptive dynamics can generate complex social behaviours [Baker, 2020, Ndousse et al., 2021, Albrecht et al., 2024]. This makes Multi-Agent Reinforcement Learning (MARL) a powerful framework for studying the emergence of cooperation and conflict in team settings [Lowe et al., 2017, Rashid et al., 2018, Foerster et al., 2016]. Similar to MAS, much of MARL research has focused on fully cooperative teams where agents share a common reward [Busoniu et al., 2008, Zhang et al., 2021], although recent advances extend its scope to mixed-motive scenarios and social dilemmas beyond zero-sum interactions [Leibo et al., 2017, Willis et al., 2025]. This shift has also encouraged richer modelling of agents' motivations [Ndousse et al., 2021, McKee et al., 2020, Madhushani et al., 2023], and the study of negotiation and bargaining protocols in multi-agent contexts [Kramár et al., 2022].

These efforts have converged into a broader research agenda known as *Cooperative Artificial Intelligence (AI)* [Dafoe et al., 2020, 2021, Conitzer and Oesterheld, 2023], which bridges artificial intelligence with the natural, social, and behavioural sciences. Cooperative AI uses computational models and RL techniques to study cooperation, and to design systems that promote beneficial collective outcomes in both human and artificial settings. While recent work in MARL has made significant progress in modelling cooperation and social dilemmas among learning agents, existing approaches typically focus on specific environments, reward structures, or interaction protocols. As a result, they do not generally provide abstract, domain-independent representations of teamwork that explicitly capture how structural features of the task—such as interdependence, evaluation criteria, or team composition—shape incentives and behaviour.

To address this gap, teamwork is framed here as a mixed-motive setting in which the task functions as a public good. The contribution of this chapter does not lie in proposing a new solution concept or optimisation procedure, but in providing a formal representation of teamwork grounded in social-psychological theory.

In contrast to cooperative game theory, which assumes perfectly aligned incentives, and to much of the MARL and Cooperative AI literature, which often relies on shared rewards or explicitly designed coordination mechanisms, the model explicitly incorporates psychologically grounded variables—such as task interdependence, team composition, and evaluative thresholds—within a non-cooperative framework. This design enables the systematic analysis and computational instantiation of these constructs in settings where cooperation is voluntary, incentives are only partially aligned, and free-riding remains possible, thereby offering a unified framework for studying teamwork across human and artificial agents.

3.4 Teamwork Games

Building on the psychosocial foundations reviewed above, this section develops a formal model of teamwork. More specifically, it provides a computational and game-theoretic formalisation of core constructs from social-psychological theories of group performance, translating them into a framework that can be analysed and instantiated in interacting agents.

Through this formalisation, the key structural variables identified by the models reviewed above are examined to assess how they shape group performance and individual incentives. The approach is game-theoretical (see Chapter 2): the decision to cooperate or not is framed as a strategy in a one-shot aggregative game.

3.4.1 Formal Model

When individuals collaborate in teams, they face a fundamental trade-off: allocate time to privately beneficial activities (*e.g.*, leisure) while relying on others to complete the task, or sacrifice that time to contribute to the common goal. The former risks universal defection and the consequent failure of the collective task; the latter risks the "sucker effect," in which one member bears the entire burden alone. From this perspective, teamwork can be regarded as a **public**

good to which members may (or may not) contribute. The framework developed here builds directly on Steiner's and Nijstad's models of group performance, translating its core constructs into a formal game-theoretic representation. In particular, task interdependence, team composition, and evaluative thresholds are encoded as explicit components of the model, allowing their effects on individual incentives and group outcomes to be analysed systematically.

For clarity, Table 3.2 summarises the main variables and parameters used throughout the formal model.

Table 3.2: Key variables and parameters in Teamwork Games

Symbol	Description
$\mathcal{I} = \{1, \ldots, n\}$	Set of players (team members)
$a_i \in [0, 1]$	Effort level of player i (fraction of time allocated to the task)
Δ	Duration of the task (time endowment)
$p_i \in [0, 1]$	Expertise of player i (productivity per unit of time)
$l_i \in [0, 1]$	Leisure capacity of player i
g_i	Contribution (gift) of player i in work units
x_i	Private good (leisure) of player i
$\boldsymbol{g}$	Vector of contributions $(g_1, \ldots, g_n)$
$\boldsymbol{p}$	Vector of expertise levels
$G(\boldsymbol{g})$	Group product (aggregated contributions)
ρ	Substitution parameter (task interdependence)
β_i	Weight of player i in the aggregation (role/importance)
Θ	Task parameters $(\rho, \beta_1, \ldots, \beta_n)$
$G^{\mathcal{T}}$	Minimum threshold required to complete the task
$\sigma(G)$	Evaluation function of the group product
$\widetilde{u}_i$	Utility function of player i

In Teamwork Games, agency lies in the time allocated to the task. Let $\mathcal{I} = \{1, \ldots, n\}$ be the set of agents working on the task. The actions of player $i \in \mathcal{I}$ are abstract and dimensionless, representing the effort as the **percentage of time** devoted to teamwork, $a_i \in \mathcal{A}_i \subseteq [0, 1]$. A joint action is the vector $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}$, where $\mathcal{A} = \prod_{i=1}^{n} \mathcal{A}_i$.

Similar to authors in social psychology [Campion et al., 1993], productivity is adopted as the metric to evaluate team performance, measured in *work units*. Players differ in their ability to complete tasks. Two players devoting the same amount of time to a team task may end up contributing very differently. More precisely, a player's *expertise* is a measure of their ability:

Definition 3.1 (Expertise). For player i, expertise $p_i \in [0, 1]$ quantifies the amount of work i can accomplish per unit of time. In that sense, a player's expertise can be thought of as

a **conversion factor of effort a_i into productivity**. Throughout this chapter, p_i is assumed fixed for each i. Given a set of n players working together, their *joint expertise* is the vector $\boldsymbol{p} = (p_1, \ldots, p_n)$.

Consequently, a player i with expertise p_i who chooses an action a_i in a single-round episode of duration Δ makes a contribution (or *gift*) of g_i work units. In this framework, a contribution represents the effective amount of work that a player adds to the collective task as a result of their time allocation decision:

$$g_i = a_i \, p_i \, \Delta \tag{3.1}$$

where g_i is measured in units of work.

Leisure is defined as any activity pursued while delegating the task to others, yielding utility outside of team responsibilities. For completeness, just as expertise measures a player's productivity per unit of time, *leisure capacity* l_i measures efficiency in deriving personal benefit from free time. Leisure is assumed to be quantifiable in the same units as work. Defining leisure capacity and expertise in comparable terms provides a consistent basis for evaluating how players allocate time between group- and individual-oriented activities.

Definition 3.2 (**Leisure Capacity**)**.** Leisure capacity is the rate at which player i obtains leisure benefits when allocating free time to personal activities. Denoted $l_i \in [0, 1]$, it is a conversion factor of the complement of effort a_i into leisure.

With this definition, the *private contribution* x_i—the leisure units allocated to private activities by i—can be written as:

$$x_i = (1 - a_i) \, l_i \, \Delta \; = \; l_i \, \Delta - g_i \frac{l_i}{p_i} \tag{3.2}$$

where the second equality follows from Eq. (3.1). Note that $l_i \Delta$ plays the role of i's exogenous income, leading to the following budget constraint:

Assumption 3.3 (**Individual budget constraints**)**.** Suppose i's exogenous income is $l_i \, \Delta$. Then player i's budget constraint is:

$$l_i \, \Delta = x_i + g_i \frac{l_i}{p_i}$$

For simplicity, the convention $l_i = 1$ is adopted for all i throughout this chapter (*i.e.*, all players are assumed capable of fully transforming their free time into leisure benefits):

$$\Delta = x_i + \frac{g_i}{p_i} \tag{3.3}$$

Group Product. Depending on the type of task involved, the gifts $\boldsymbol{g}$ are aggregated differently into a final outcome—henceforth the *group product*. A single parameterisation, the Constant Elasticity of Substitution (CES), is used as the aggregator across Steiner's task types. This aggregator is widely used in the aggregative-games literature, as detailed in Chapter 2. Hence, the group product (in work units) is denoted G and defined by

$$G(\boldsymbol{g}) = \left(\sum_{i=1}^{n} \beta_i \, g_i^{\rho} \right)^{\frac{1}{\rho}} \tag{3.4}$$

The parameter $\rho \in \mathbb{R}^*$ in Eq. (3.4) sets the level of task interdependence[1]. In terms of Chapter 2, ρ controls the degree of substitutability across teammates' gifts. On the other hand, $\beta_i > 0$ is the weight given to the gift of team member i—signalling i's role in the team. High values of β_i imply i's leadership. Together, these parameters encode the task type, $\Theta = (\rho, \beta_1, \ldots, \beta_n)$.

In the standard CES formulation, setting $\beta_i = 1/n$ yields a generalised (Hölder) mean of $\boldsymbol{g}$. This normalisation is conventional—though not compulsory—in economics as it provides a measure of mean productivity.

In contrast, allowing β_i to remain unnormalised suggests an interpretation of G as a group measurable product. Likewise, it enables the weights to reflect structural asymmetries among teammates (*e.g.*, leadership or relative task importance) without altering the scale of G.

Throughout, $\beta_i = 1$ is intentionally set for all i (no role weights), unless stated otherwise, so that G represents the total production of a team without role differentiation rather than a per-capita normalisation. Future work will explore teamwork in hierarchical teams by choosing $\beta_i \neq 1$.

With all these elements in place, *unitary team tasks* can now be defined:

Definition 3.4 (**Unitary Team Task**)**.** Consider a group of n agents who must jointly complete a shared task within **one** round of duration $\Delta > 0$. A unitary team task $\mathcal{T}$ is defined as the tuple

$$\mathcal{T} \stackrel{\text{def}}{=} \left(G^{\mathcal{T}}, \Delta, \Theta \right) \tag{3.5}$$

Given a vector of gifts $\boldsymbol{g} = (g_1, \ldots, g_n)$ producing a group product G as per Eq. (3.4), the task is *accomplished* if and only if productivity meets or exceeds the minimum requirement or

[1] $\mathbb{R}^+$ denotes the set of positive real numbers, $\mathbb{R}_0^+$ the set of non-negative real numbers including zero, and $\mathbb{R}^*$ the set of all real numbers except zero.

threshold $G^{\mathcal{T}}$:

$$G \geq G^{\mathcal{T}} \tag{3.6}$$

By setting the parameters appropriately, the function $G(\boldsymbol{g})$ in Eq. (3.4) yields a group production model that incorporates key elements of Steiner and Nijstad's theories of group performance:

1. β_i is the weight given to the gift of team member i.

2. Along with the set $\{\beta_i\}_{i\in\mathcal{I}}$, the substitution parameter ρ determines the **task type**:

   - **Additive tasks**: $\rho = 1$ and $\beta_i = 1$ for all i. Gifts are aggregated additively, reproducing the canonical public good model [Bergstrom et al., 1992, Cornes and Hartley, 2007a].

   - **Conjunctive tasks**: $\rho < 1$ and $\beta_i = 1$ for all i. Lower gifts weigh more heavily on the group product. As $\rho \to -\infty$, the task becomes strongly conjunctive, where the weakest performance fully determines productivity. Mathematically, an aggregative public good with $\rho < 1$ is concave in g_i [Cornes and Hartley, 2007b].

   - **Disjunctive tasks**: $\rho > 1$ and $\beta_i = 1$ for all i. Higher gifts weigh more heavily on the group product. As $\rho \to +\infty$, the task becomes strongly disjunctive, where the best performance fully determines productivity. In other words, the group product depends on the player producing the most work units g_i. Mathematically, an aggregative public good with $\rho > 1$ is not concave in g_i [Cornes and Hartley, 2007b].

Solving Eq. (3.3) for x_i and adding the group product G to both sides yields player i's full income M_i:

Definition 3.5 (**Full Income**). A player's full income is the sum of their free time and the accessible public good, encompassing all resources available to them:

$$M_i \stackrel{\text{def}}{=} x_i + G = \Delta - \frac{g_i}{p_i} + \left(G_{-i}^{\rho} + \beta_i g_i^{\rho}\right)^{\frac{1}{\rho}} \tag{3.7}$$

where $G_{-i}^{\rho} = \sum_{j\neq i} \beta_j g_j^{\rho}$ and $G = \left(G_{-i}^{\rho} + \beta_i g_i^{\rho}\right)^{\frac{1}{\rho}}$. Notice that M_i incorporates others' gifts explicitly as part of player i's income endowment, a feature that will be useful in later sections.

Finally, the last pending piece to define Teamwork Games is the notion of an *evaluation function* in this context. The evaluation function is essential because Teamwork Games differ from general public good games in a key respect: a player's utility depends on the public good through its evaluation.

Definition 3.6 (Evaluation Function). A function $\sigma : \mathcal{G} \to [\underline{\sigma}, \overline{\sigma}] \subset \mathbb{R}_0^+$ is an evaluation function if:

1. σ is strictly increasing;
2. σ is differentiable;
3. $\sigma'(G)^2 - \sigma''(G)\,\sigma(G) > 0$ for all G in its domain.

where $\underline{\sigma}$ denotes the infimum (minimum attainable evaluation) and $\overline{\sigma}$ the supremum (maximum attainable evaluation).

Notice how Condition **3)** is equivalent to $(\log \sigma(G))'' < 0$, so that σ is strictly log-concave. In particular, the ratio $\sigma(G)/\sigma'(G)$ is strictly increasing in G, a property that plays a central role in the analysis of best responses and replacement correspondences. Log-concavity of σ does not by itself guarantee quasi-concavity of $\widetilde{u}_i(x_i, \sigma(G))$ for arbitrary $\widetilde{u}_i$. However, as shown in Lemma 3.9 below, it does preserve the curvature of preferences for the class of Cobb-Douglas-inspired utilities used in explicit equilibrium computations. A *Teamwork Game* can now be defined:

Definition 3.7 (Teamwork Game). Consider a set of players $\mathcal{I} = \{1, \ldots, n\}$ engaged in a general public good game $\Gamma = (\mathcal{A}_i, \widetilde{u}_i)_{i \in \mathcal{I}}$—see Chapter 2 for further reference. Given a one-shot task $\mathcal{T} = (G^{\mathcal{T}}, \Delta, \Theta)$, an evaluation function $\sigma : \mathcal{G} \to [\underline{\sigma}, \overline{\sigma}] \subset \mathbb{R}$, a joint expertise vector $\boldsymbol{p} = (p_1, \ldots, p_n)$, and a joint leisure capacity vector $\boldsymbol{l} = (l_1, \ldots, l_n)$, the triplet $(\Gamma, \mathcal{T}, \sigma)$ is a *Teamwork Game* if it satisfies:

1. The game is static (simultaneous) and non-cooperative, where strategic choices are percentages $(a_1, \ldots, a_n) \in [0, 1]$ of the total available time Δ allocated by each player to the task.
2. These strategic choices $(a_1, \ldots, a_n)$ correspond to individual gifts $(g_1, \ldots, g_n)$ via Eq. (3.1), with $g_i \geq 0$ for all i.
3. The aggregator G is the CES function with parameters $\Theta = \{\rho, \beta_1, \ldots, \beta_n\}$:
$$G = \left(\sum_{i=1}^{n} \beta_i \, g_i^{\rho} \right)^{\frac{1}{\rho}}$$
4. For player i, the private good x_i is given by Eq. (3.2).

5. Player i's reduced utility (or simply utility) is a function of their private good x_i and the evaluation $\sigma(G)$, of the form

$$\widetilde{u}_i(x_i, \sigma(G)) \tag{3.8}$$

where $\widetilde{u}_i(x_i, G)$ was the original utility in the general public good game Γ.

In addition, well-behaved preferences are assumed:

Assumption 3.8 (Well-behaved preferences in Teamwork Games)**.** In a Teamwork Game $(\Gamma, \mathcal{T}, \sigma)$, player i's reduced utility is

$$\widetilde{u}_i(x_i, \sigma(G))$$

It is assumed that $\widetilde{u}_i$ is continuously differentiable, strictly increasing, and strictly quasi-concave in $(x_i, \sigma(G))$, and that both x_i and $\sigma(G)$ are normal goods under these preferences. We restrict attention to interior allocations with $x_i > 0$ and $\sigma(G) > 0$, so that logarithms are well defined whenever they are used in the analysis.
Remark: Since σ is strictly increasing, normality of $\sigma(G)$ is equivalent to normality of the underlying public good G.

Assumption 3.8 does not hold automatically. To close this section, Definitions 3.6 and 3.7 are used to introduce a lemma. Its goal is to propose a working example of a utility function ensuring that preferences in Teamwork Games are well-behaved:

Lemma 3.9 (Example of well-behaved preferences under evaluation)**.** *Consider a Teamwork Game $(\Gamma, \mathcal{T}, \sigma)$ as in Definition 3.7. Suppose that σ is an evaluation function and that player i's reduced utility takes the Cobb-Douglas-inspired form*

$$\widetilde{u}_i(x_i, \sigma(G)) \;=\; x_i^{\alpha}\, \sigma(G), \qquad \alpha > 0$$

Then:

1. *$\widetilde{u}_i$ is continuously differentiable and strictly increasing in both x_i and $\sigma(G)$.*
2. *$\widetilde{u}_i$ is strictly quasi-concave in $(x_i, \sigma(G))$.*

Hence, Cobb-Douglas preferences combined with an evaluation function σ satisfy Assumption 3.8.

The proof of Lemma 3.9 is provided in Appendix A.1.

Teamwork Games and classical social dilemmas. So far the formal model of Teamwork Games has been introduced and these mixed-motive interactions have been referred to as social dilemmas. Yet no explicit analysis has been given of *why* or *in what sense* they correspond to classical dilemma types (see Appendix A.5). There, two-player Teamwork Games are reduced to simple 2×2 normal-form games and it is shown how each task type (additive, conjunctive, disjunctive) maps onto a standard social dilemma. This static characterisation is not needed for the equilibrium analysis that follows, but it situates the model within existing theory and provides a useful reference for the comparison with Sequential Social Dilemmas in Chapter 4.

3.4.2 Computing Nash Equilibria in Teamwork Games

Teamwork Games provide a direct translation of key psychological primitives—*task type*, *expertise*, and *evaluation*—into a formal strategic model of voluntary cooperation in teams. Building on this parallelism, the solution concepts can be read as a set of prescriptive benchmarks for rational behaviour in collaborative settings.

This subsection focuses on one such benchmark: the Nash Equilibrium. Analysing NEs reveals the game's structural alignment mechanisms—what incentives foster cooperation, which discourage it, and under what conditions cooperation constitutes a rational course of action.

In this framework, each player's reduced utility $\widetilde{u}_i$ depends only on two arguments: their own strategy and the aggregate group product G formed by the gifts of all players. This feature allows the best responses to be computed without tracking each opponent's move individually. Instead, we rely on an alternative object: the *replacement correspondence* $r_i(G)$. This correspondence gives the set of gifts that player i would rationally choose if the group product—already including their own gift—were G.

Knowing the joint expertise of a team, there is an unambiguous mapping between a joint action $\boldsymbol{a}$ and the corresponding gift vector $\boldsymbol{g}$. Accordingly, when computing Nash Equilibria in the rest of the chapter, we express best responses in terms of gifts g_i rather than percentages of time a_i, unless stated otherwise. Specifically, we adopt a mild abuse of notation and let both the reduced best-response correspondence $\widetilde{bi}(G-i)$ (Eq. (2.4)) and the replacement correspondence $r_i(G)$ (Def. 2.3) output gifts g_i rather than actions a_i. This choice, though slightly non-standard, simplifies the discussion of the evaluation function's role in the game.

To connect this with the *reduced best-response function* $\widetilde{b}_i(G_{-i})$, recall from Chapter 2, Section 2.1, that $\widetilde{b}_i$ maps the gifts of others into player i's best strategy. The replacement corre-

spondence instead internalises the player's own action into the aggregate: for additive tasks with $G = \sum_i g_i$, one can show that $g_i = r_i(G)$ satisfies:

$$g_i = \widetilde{b}_i(G - g_i)$$

Whereas for the general case in Eq. (3.4), the equality becomes:

$$r_i(G) = \widetilde{b}_i\left(\left[G^\rho - \beta_i g_i^\rho\right]^{\frac{1}{\rho}}\right)$$

This formulation highlights a methodological distinction across task types. For additive and conjunctive tasks, the replacement correspondence is a function which selects a unique best response for each feasible group product. For disjunctive tasks, by contrast, non-concavity of the aggregator implies that the replacement correspondence maps into a set of strategies rather than a single one. As we will see, this difference is central, since it opens the possibility of multiple equilibria with free-riding behaviour.

Before turning to the formal derivations, it is helpful to recall the motivating algebra-assignment example introduced earlier. Its disjunctive task structure anticipated the equilibria that emerge when one member's contribution dominates the outcome. As we shall see, this translates into one or a few members carrying the task while the others remain passive. The following analysis shows how such outcomes can be derived formally within the replacement-correspondence framework.

The analysis unfolds in two steps. First, we identify the conditions under which equilibria are well defined by studying the existence of the replacement correspondence $r_i(G)$ across different task types. Second, once these conditions are established, we derive explicit equilibrium outcomes by specifying a functional form for preferences.

Existence and domain of the replacement correspondence

Establishing when the replacement correspondence is well defined is a prerequisite for any equilibrium analysis. The following cases illustrate how this depends on the task's structure.

Additive tasks. When the task is additive ($\rho = 1$), the group product reduces to

$$G = \sum_{i=1}^{n} \beta_i g_i. \tag{3.9}$$

Under Assumptions 3.8 and 3.3, and with the additive aggregator in Eq. (3.9), the Teamwork Game satisfies the conditions A1–A4 in Cornes and Hartley [2007a]: preferences are well-behaved (A1), the budget constraint is linear (A2), the public good aggregates additively (A3),

and both x_i and $\sigma(G)$ are normal goods (A4). Therefore, by Proposition 2.2 in Cornes and Hartley [2007a], the replacement correspondence of player i is a function $r_i : \mathcal{G} \to \mathcal{A}_i$ with the following properties:

1. There exists a finite standalone level $\overline{G}_i$ such that $r_i(\overline{G}_i) = \overline{G}_i$ for all i.
2. $r_i(G)$ is defined for all $G \geq \overline{G}_i$.
3. $r_i(G)$ is continuous.
4. $r_i(G)$ is everywhere non-increasing in G, and strictly decreasing wherever $r_i(G) > 0$.

The constant $\overline{G}_i$ is the *standalone value*: the level of G that player i would produce as sole contributor. This threshold marks the lower bound of the domain where $r_i(G)$ is defined. Below this value, the replacement function is undefined and no Nash Equilibrium can arise.

Conjunctive tasks. When $\rho < 1$ and $\rho \neq 0$, the task is conjunctive: lower gifts weigh more heavily in determining the group product. Here, G remains concave in each g_i.

Lemma 3.10. *Let $(\Gamma, \mathcal{T}, \sigma)$ be a Teamwork Game with a conjunctive task. If x_i and G are normal and desirable goods, then the replacement correspondence $r_i(G)$ is a function satisfying*

$$\left.\frac{\partial \widetilde{u}_i}{\partial g_i}\right|_{g_{-i}} \geq 0 \tag{3.10}$$

with equality whenever $g_i < p_i\Delta$. Moreover:

1. *if $0 < \rho < 1$, then $r_i : [\overline{G}_i, \infty) \to \mathcal{A}_i$;*
2. *if $\rho < 0$, then $r_i : [0, \overline{G}_i] \to \mathcal{A}_i$.*

Thus, in *quasi-conjunctive* tasks ($0 < \rho < 1$) the replacement function is defined from the standalone level upwards, whereas in *strongly conjunctive* tasks ($\rho < 0$) it spans the interval up to $\overline{G}_i$. Proofs are provided in Appendix A.2.

Disjunctive tasks. When $\rho > 1$, the task is disjunctive: higher gifts dominate in the aggregation of the group product. In this regime, G is non-concave in g_i. To obtain explicit best-response expressions under disjunctive tasks, we specify a utility consistent with Assumption 3.8. A convenient choice is the Cobb-Douglas form validated in Lemma 3.9:

$$\widetilde{u_i}(x_i, \sigma(G)) = x_i^{\alpha}\, \sigma(G), \qquad \alpha > 0 \tag{3.11}$$

Under these preferences, the **best-response set** $\mathcal{B}_i(G_{-i})$ exhibits three behavioural regimes (Lemma 3.11): a *no-gift regime* when G_{-i} exceeds a threshold, an *indifference regime* at the threshold, and a *gift regime* when G_{-i} falls short of it.

Lemma 3.11. *Let $(\Gamma, \mathcal{T}, \sigma)$ be a Teamwork Game featuring a disjunctive task ($\rho > 1$):*

$$G = \left(\sum_{i}^{n} \beta_i\, g_i^{\rho} \right)^{\frac{1}{\rho}} \text{ with } \rho > 1$$

Suppose that player i has preferences as in Equation (3.11), where $\sigma(G)$ denotes an evaluation function. Then, there exists a threshold value $G^*_{-i} > 0$ and a positive real-valued function $\widetilde{b}_i$ on $\left[0, G^*_{-i}\right]$ such that:

$$\mathcal{B}_i(G_{-i}) = \begin{cases} \{0\} & \text{if } G_{-i} > G^*_{-i} \\ \left\{0, \widetilde{b}_i(G_{-i})\right\} & \text{if } G_{-i} = G^*_{-i} \\ \left\{\widetilde{b}_i(G_{-i})\right\} & \text{if } G_{-i} < G^*_{-i} \end{cases} \tag{3.12}$$

The proof of this lemma can be found in Appendix A.4; Figure 3.2 depicts the cases it covers.

The existence of these three regimes implies that disjunctive tasks may yield multiple equilibria and partial free-riding, as predicted by the motivating example. The next subsection derives these equilibria explicitly.

Nash Equilibria Computation

Having established the conditions for the replacement correspondence $r_i(G)$ to exist, we now turn to its explicit computation and, consequently, to the derivation of Nash Equilibria. From this point onwards, all derivations assume the Cobb-Douglas-inspired utility function introduced in Lemma 3.9, which offers analytical tractability while remaining consistent with the evaluation-function framework. The equilibrium computation proceeds as follows:

1. Derive $r_i(G)$ for each task type;
2. Aggregate these into $R(G)$;
3. Solve for fixed points $\hat{G}$;

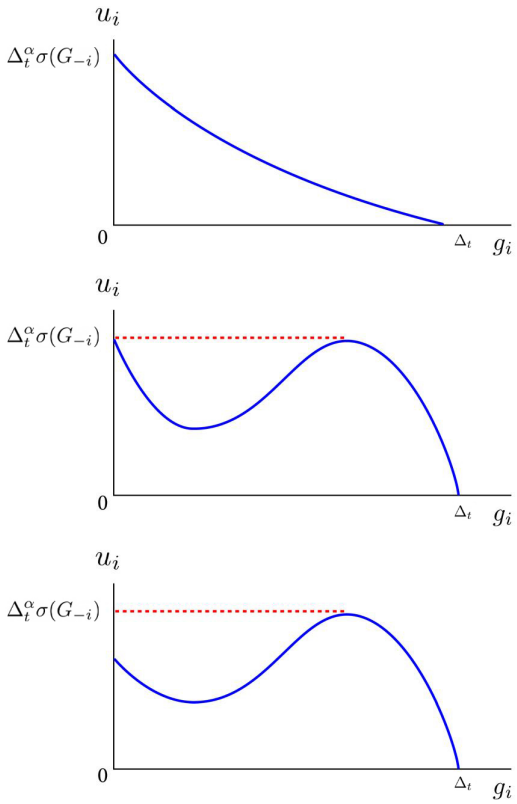


Figure 3.2: Player i's utility $\widetilde{u}_i(x_i, \sigma(G_{-i}))$, across three regime values of G_{-i}. Upper panel: $G_{-i} > G^*_{-i}$. Middle panel: $G_{-i} = G^*_{-i}$. Bottom panel: $G_{-i} < G^*_{-i}$. Adapted from [Cornes and Hartley, 2007b].

4. Each fixed point yields a corresponding set of equilibrium strategy profiles $\{\hat{\boldsymbol{a}}\}$.

For further details, please refer to Chapter 2.

Additive tasks. For additive tasks ($\rho = 1$), we showed in Section 3.4.2 that if the group product G is at least as large as each player's standalone value $\overline{G}_i$, then the replacement correspondence is a well-defined function. Under these circumstances, the existence and uniqueness of a NE are guaranteed (see Cornes and Hartley, 2007a for a full justification). Moreover, because G is concave in each g_i [Cornes and Hartley, 2007b], the equilibrium can be obtained from standard first-order conditions. Substituting $\rho = 1$ into the CES production function

(see Eq. (3.9)), applying the Cobb-Douglas-inspired preferences from Eq. (3.11), and using first-order-conditions, one obtains the following expression for the replacement function:

$$r_i(G) = \max\left\{0,\ p_i\Delta - \frac{\alpha}{\beta_i}\,\frac{\sigma(G)}{\sigma'(G)}\right\} \tag{3.13}$$

The interpretation of Eq. (3.13) is direct. The minuend in the non-zero term represents the potential productivity that player i could achieve if they dedicated an entire turn Δ to contributing to the task. The subtrahend quantifies the extent to which this is not possible: higher values render the subtrahend more negative, indicating a greater impediment to the agent to contribute. The ratio $\sigma(G)/\sigma'(G)$ measures the local responsiveness of the evaluation function: a larger ratio indicates a flatter or more easily attained evaluation profile.

Thus, when $p_i\Delta$ is less than $\frac{\alpha}{\beta_i}\,\frac{\sigma(G)}{\sigma'(G)}$, the result of $r_i(G)$ is zero and player i chooses not to contribute. That is, when the expertise is sufficiently low, the best a player can do is free-ride on the efforts of others with greater expertise than themselves. Conversely, if $p_i\Delta$ is greater than $\frac{\alpha}{\beta_i}\,\frac{\sigma(G)}{\sigma'(G)}$, then $r_i(G) = p_i\Delta - \frac{\alpha}{\beta_i}\,\frac{\sigma(G)}{\sigma'(G)}$, indicating that player i contributes a positive amount determined by this difference. Holding other parameters constant, as p_i increases, it becomes more likely for player i to contribute more, as $p_i\Delta$ becomes larger in comparison to the negative term in the equation. The same applies to the behaviour of $r_i(G)$ as Δ increases. In short, the higher a player's expertise, the more likely they are to contribute. Another way to encourage a player to contribute is by providing them with more time Δ (longer turns).

As for α, it captures the relative importance that players give to their private time compared to the group product. For the case $r_i(G) \neq 0$, it becomes evident that, for a fixed quantity of group product G and a fixed weight β_i, the optimal gift g_i diminishes as $\alpha > 0$ increases. This is an intuitive outcome, as higher values of α signify a greater emphasis on leisure in the preferences of player i, thereby discouraging gifts.

Conversely, as β_i increases (with all other factors held constant), the absolute value of the negative term diminishes. Put differently, the greater the importance or weight β_i assigned to a player's gifts, the more the player will contribute, assuming other factors remain constant.

Regarding the teamwork evaluation function, higher values of the fraction $\frac{\sigma(G)}{\sigma'(G)}$ render the subtrahend in Eq. (3.13) more negative, indicating a greater impediment to the agent's gift. This ratio can increase either due to a rise in the numerator or a decrease in the denominator. For simplicity, let us consider for a moment that either the numerator or the denominator changes while the other remains constant.

If the numerator increases at a point G_0 while $\sigma'(G)$ remains constant, it means that the evaluation has become more lenient, or in other words, the passing threshold has been lowered, making it easier to pass. Counter-intuitively, this lowering of the passing threshold inhibits player gifts rather than incentivising them, as the same utility is obtained with less effort than before.

If we fix the value of the evaluation at some group product value, $\sigma(G_0)$, but decrease the slope of the evaluation function $\sigma'(G)$, then more group product units ΔG will be needed to increase the evaluation from $\sigma(G_0)$ to $\sigma(G_0)+1$. In this situation, a player would prefer an evaluation function with a steeper slope, as an increase in their gifts would result in greater utility. In other words, decreasing the slope of the evaluation at a point on the curve $\sigma(G_0)$ inhibits gifts.

In both cases, the ratio $\frac{\sigma(G)}{\sigma'(G)}$ rises. Hence, a more lenient or flatter evaluation function discourages contributions, whereas a steeper and stricter evaluation fosters them: as the ratio **decreases**, the opportunity cost of effort falls and players find it easier to raise their utility through gifting.

Conjunctive tasks. For conjunctive tasks ($\rho < 1$, $\rho \neq 0$), G remains concave in each g_i, so the NE can again be derived from first-order conditions. Lemma 3.10 established that the replacement correspondence is in fact a function, with its domain depending on whether the task is quasi-conjunctive ($0 < \rho < 1$) or strongly conjunctive ($\rho < 0$). Building on this, Cornes and Hartley [2007b] proved the uniqueness of equilibrium in general public good games with concave G using the share function. Extending their result to Teamwork Games yields the following proposition:

Proposition 3.12. *Let $(\Gamma, \mathcal{T}, \sigma)$ be a Teamwork Game with a conjunctive task. Under the conditions of Lemma 3.10, there is a unique equilibrium whenever the utility function $\widetilde{u}_i(x_i, \sigma(G))$ has indifference curves asymptotic to the axes.*

Applying the first-order condition in Eq. (3.10) and substituting the Cobb-Douglas-inspired utility function from Eq. (3.11) yields:

$$\frac{\sigma(G)}{\sigma'(G)}\, G^{\rho-1} = \left(\Delta - \frac{r_i(G)}{p_i}\right) \frac{\beta_i p_i}{\alpha}\, r_i(G)^{\rho-1} \tag{3.14}$$

Equation (3.14) characterises the replacement function $r_i(G)$ in conjunctive Teamwork Games. Its comparative statics are discussed in Appendix A.3.

To illustrate, consider strongly conjunctive tasks where $\rho \to -\infty$. If the team is homogeneous ($p_i = 1$, $\beta_i = 1$ for all i), then

$$G = \min_{i \in \mathcal{I}} g_i$$

So the group product is determined by the weakest gift. By symmetry, all players contribute equally in equilibrium, and the group product $\hat{G}(\rho)$ satisfies

$$\hat{g} = \left(\frac{1}{n}\right)^{1/\rho} \hat{G}(\rho) \tag{3.15}$$

Substituting $\hat{g} = r_i(G)$ from Eq. (3.15) into Eq. (3.14) and taking the limit as $\rho \to -\infty$ yields

$$\hat{G}(-\infty) = \Delta - n\alpha \frac{\sigma(G)}{\sigma'(G)} \tag{3.16}$$

That is: in strongly conjunctive tasks with homogeneous teams, **enlarging the group reduces productivity at equilibrium**. This aligns with Steiner [1972]'s observation that conjunctive tasks are bottlenecked by the weakest member, implying that larger teams tend to underperform.

Disjunctive tasks. The disjunctive case ($\rho > 1$) requires more care. Here, the group product is non-concave in gifts, meaning that standard first-order conditions cannot be relied upon alone. Instead, we examine both local and global optimality conditions. Once again, we start from the preferences in Eq. (3.11). The condition for a local maximum is the same as in Eq. (3.14). Rearranging yields:

$$G = \left(\frac{\beta_i p_i}{\alpha}\left(\Delta - \frac{r_i(G)}{p_i}\right)\frac{\sigma'(G)}{\sigma(G)}\right)^{\frac{1}{\rho-1}} r_i(G) \tag{3.17}$$

Having characterised the local maximum, we next establish the global condition required for equilibrium. For this point to be a global maximum, the payoff must exceed or equal that at $g_i = 0$:

$$\left(\Delta - \tfrac{g_i}{p_i}\right)^{\alpha} \sigma(G) \;\geq\; \Delta^{\alpha}\, \sigma(G_{-i}) \tag{3.18}$$

Combining the local and global conditions defines the replacement correspondence $r_i(G)$. Notice that by definition, the values of G^*_{-i} from Lemma 3.11 and the corresponding critical point g^*_i are obtained when Eqs. (3.17) and (3.18) are satisfied with equality. Because in this case, the set of best-responses can have more than one element according to Lemma 3.11, a replacement **correspondence** has to be defined in turn:

$$r_i(G) = \{\, g_i : g_i \in \mathcal{B}_i([G^\rho - \beta_i g_i^\rho]^{1/\rho}) \,\} \tag{3.19}$$

Hence, in a disjunctive task, player i might choose from multiple possible strategies—*i.e.*, gifts—in an equilibrium that results in total group product G. From Lemma 3.11, $g_i \in r_i(G)$ if and only if either:

$$\begin{cases} g_i = 0 \text{ and } G_{-i} \geq G^*_{-i}, \text{ or} \\ g_i > 0 \text{ and } \begin{cases} \frac{\sigma(G)}{\sigma'(G)}\, G^{\rho-1} = \left(\Delta - \frac{g_i}{p_i}\right) \frac{\beta_i p_i}{\alpha}\, g_i^{\rho-1} \\ \beta_i^{1/\rho} g_i \leq G \leq \left(\sigma^{-1}\left(\sigma(G)\left(1 - \frac{g_i}{p_i \Delta}\right)^\alpha\right)^\rho + \beta_i g_i^\rho\right)^{\frac{1}{\rho}} \end{cases} \end{cases} \tag{3.20}$$

Where, for the second inequality of the case $g_i > 0$, the global maximum condition from Eq. (3.18) has been used. The above expressions tell us that the replacement correspondence $r_i(G)$ in disjunctive tasks has two components:

1. A component extending along the X- axis to the right of G^*_{-i}.

2. A positive component, which is the reflection across the 45° line of the segment of the group product G that simultaneously fulfils the local and global maximum conditions (Eqs. (3.17) and (3.18) respectively).

To understand the geometry of the positive component of r_i, notice that the right-hand side of Equation (3.14) vanishes when g_i is either equal to 0 or to $p_i\,\Delta$, and it has a peak at $g_i = p_i\,\Delta\left(1 - \frac{1}{\rho}\right)$. Additionally:

1. The boundary line of the left-hand inequality in Eq. (3.20), crosses the plot of Eq. (3.17) at the point corresponding to the standalone value of player i (the level of the group product that player i would contribute were they the sole contributor, $\overline{G}_i$):

$$(\overline{g}_i, \overline{G}_i) = \left(\Delta\, p_i - \frac{\sigma(\overline{G}_i)}{\sigma'(\overline{G}_i)} \frac{\alpha}{\beta_i^{1/\rho}},\ \beta_i^{1/\rho}\, \Delta\, p_i - \frac{\sigma(\overline{G}_i)}{\sigma'(\overline{G}_i)}\alpha\right)$$

2. By definition, the boundary line of the right-hand inequality in Eq. (3.20) crosses the graph of Eq. (3.17) to the left of $(\overline{g}_i, \overline{G}_i)$, at the critical point (g_i^*, G_i^*), where

$$G_i^* = (G_{-i}^{*\rho} + \beta_i g_i^{*\rho})^{1/\rho}$$

The top panel in Figure 3.3 illustrates G from Equation (3.17) and how the different boundary lines mentioned intersect with its graph. The bottom panel represents the positive component of r_i as the reflection across the 45° line of the segment of G that fulfils the conditions for $g_i > 0$ in Eq. (3.20).

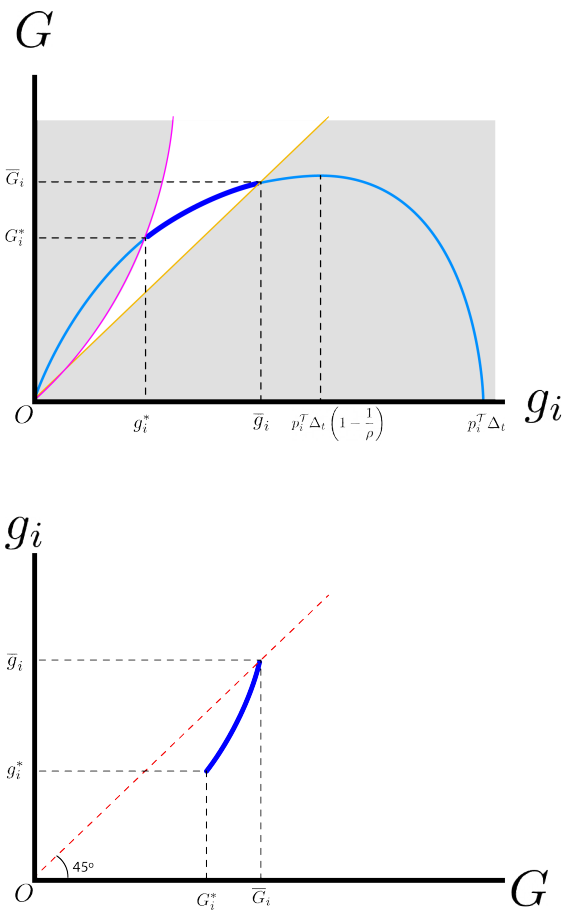


Figure 3.3: **Top panel: The group product G, as given by Eq. (3.17), is depicted in two different shades of blue and has a bell shape.** The boundary line of the left-hand inequality in Eq. (3.20), shown in fuchsia, intersects this graph at (g_i^*, G_i^*). The boundary line of the right-hand inequality in Eq. (3.20), shown in light orange, is to the right of the fuchsia line and intersects the graph at $(\overline{g}_i, \overline{G}_i)$. The bold blue segment within the non-shaded area indicates the specific region of G that satisfies both the local and global maximum conditions simultaneously. **Bottom panel: Reflection of the group product G across the 45° line, showing only the region where G meets the local and global maximum conditions.** Figures adapted from [Cornes and Hartley, 2007b].

If both (g_i^*, G_i^*) and $(\overline{g}_i, \overline{G}_i)$ are located to the left of (or at) the maximum of $G(g_i)$, both points will lie within the increasing portion of this function. Examining the 45° reflection of

that function in the bottom panel of Figure 3.3, this indicates that the positive component of the replacement correspondence is single-valued on its domain: $\left[G_i^*, \overline{G}_i\right]$. To ensure this condition is met, it must satisfy:

$$\frac{\rho}{p_i \Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)} \frac{\beta_i^{1/\rho}}{\alpha}$$

The next corollary follows naturally:

Corollary 3.13. *In a Teamwork Game involving a disjunctive task, if*

$$\frac{\rho}{p_i \Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)} \frac{\beta_i^{1/\rho}}{\alpha}$$

then the positive component of the replacement correspondence is single-valued on its domain: $\left[G_i^*, \overline{G}_i\right]$.

In essence, this result tells us that although a player in a Teamwork Game involving a disjunctive task does not have a replacement function but rather a replacement correspondence as per Eq. (3.19), when examining **the two possible components** of the replacement correspondence individually, they **are single-valued**. One component outputs zero, and the other is a function with a domain of $\left[G_i^*, \overline{G}_i\right]$. As a consequence, Teamwork Games involving disjunctive tasks may have multiple equilibria. In each of these equilibria, some players will contribute ($g_i > 0$), while all others will defect ($g_i = 0$). Players who positively contribute to teamwork are called the "active" players in that equilibrium, and they form an **equilibrium set** $\mathcal{J} \subseteq \mathcal{I}$.

By Definition 2.6, whoever the active players are, their share correspondences will sum up to 1 in equilibrium:

$$\sum_{j \in \mathcal{J}} s_j(\hat{G}) = 1 \tag{3.21}$$

Whereas $\hat{g}_i = 0$ for all $i \notin \mathcal{J}$. A unique equilibrium for each set of active players follows if $s_j(G)$ are increasing functions for the active players:

Proposition 3.14. *If* $\frac{\rho}{p_i \Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)} \frac{\beta_i^{1/\rho}}{\alpha}$, *the graph of* $s_i(G)$ *in a Teamwork Game featuring a disjunctive task is the disjoint union of two sets:* $\{(G, 0) : G \geq G_{-i}^*\}$ *and* $\{(G, s_i(G)) : G_i^* \leq G \leq \overline{G}_i\}$. *Plus,* s_i *is continuous and strictly increasing, and satisfies*

$$\begin{cases} s_i(G_i^*) = \beta_i \frac{g_i^{*\rho}}{G_i^{*\rho}} > 0, \text{ and} \\ s_i(\overline{G}_i) = 1 \end{cases}$$

The interested reader can find proof of this Proposition in Appendix A.6. Figure 3.4 depicts this result.

We now have all the ingredients to delineate equilibria in disjunctivet tasks using share functions. According to Proposition 3.14, for any active player $j \in \mathcal{J}$, $s_j(G)$ is an increasing function in G and hence there is one equilibrium per set of active players. The level of that equilibrium is denoted by $\hat{G}(\mathcal{J})$, and the equilibrium is obtained through the following procedure:

1. Set $s_i(\hat{G}(\mathcal{J})) = 0$ if $i \notin \mathcal{J}$.
2. Compute $\{s_j(\hat{G}(\mathcal{J}))\}$ for all $j \in \mathcal{J}$ such that Eq. (3.21) is satisfied.

Therefore, characterising equilibria is equivalent to characterising equilibrium sets, with necessary conditions for an equilibrium set following from Proposition 3.14. That is, in an equilibrium set $\mathcal{J}$, the following conditions are fulfilled:

$$G_j^* \leq \hat{G}(\mathcal{J}) \leq \overline{G}_i \text{ for all } j \in \mathcal{J}$$
$$G_{-i}^* \leq \hat{G}(\mathcal{J}) \text{ for all } i \notin \mathcal{J}$$

We can use these conditions to identify equilibrium sets systematically:

1. Consider each possible subset $\mathcal{J} \subseteq \mathcal{I}$.
2. Within each subset, select the critical value G_j^* corresponding to the player with the lowest tolerance for undersupplying the group product.
3. Simultaneously, choose the critical value G_{-i}^* of the player outside the subset who exhibits the least tolerance for the group product being undersupplied by others.
4. The maximum of these two values yields $G^*(\mathcal{J})$.
5. For $\mathcal{J}$ to qualify as an active set, it must consist of players whose individual standalone values $\overline{G}_j$ all exceed or equal this critical value $G^*(\mathcal{J})$, while also satisfying $\sum_{j\in\mathcal{J}} s_j(G^*(\mathcal{J})) \leq 1$.

Formally,

$$G^*(\mathcal{J}) = \max\left\{\max_{j\in\mathcal{J}} G_j^*,\ \max_{i\notin\mathcal{J}} G_{-i}^*\right\}.$$

With this, the characterisation of equilibrium sets in non-concave public goods in Cornes and Hartley [2007b] extends to Teamwork Games:

Proposition 3.15. *Suppose* $\frac{\rho}{p_i\Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)}\frac{\beta_i^{1/\rho}}{\alpha}$ *for all* $i \in \mathcal{I}$ *in a disjunctive Teamwork Game. A non-empty set* $\mathcal{J} \subseteq \mathcal{I}$ *is an equilibrium set if and only if*

$$\begin{cases} G^*(\mathcal{J}) \leq \min\limits_{j\in\mathcal{J}} \overline{G}_j, \\ \sum\limits_{j\in\mathcal{J}} s_j(G^*(\mathcal{J})) \leq 1. \end{cases}$$

It follows that non-empty subsets of equilibrium sets also satisfy equilibrium conditions (see See Corollary 6.1 in [Cornes and Hartley, 2007b]):

Corollary 3.16. *If* $\frac{\rho}{p_i\Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)}\frac{\beta_i^{1/\rho}}{\alpha}$ *for all* $i \in \mathcal{I}$*, then for any equilibrium set* $\mathcal{J}$*, every subset* $\mathcal{K} \subset \mathcal{J}$ *is also an equilibrium set, with* $\hat{G}(\mathcal{K}) > \hat{G}(\mathcal{J})$.

Intuitively, when fewer players contribute, each active member must cover a larger share of the task, so the group product increases. Free-riders then enjoy a larger public good without contributing themselves. As noted by Cornes and Hartley [2007b], this explains why in disjunctive tasks, free-riders prefer equilibria with fewer contributors.

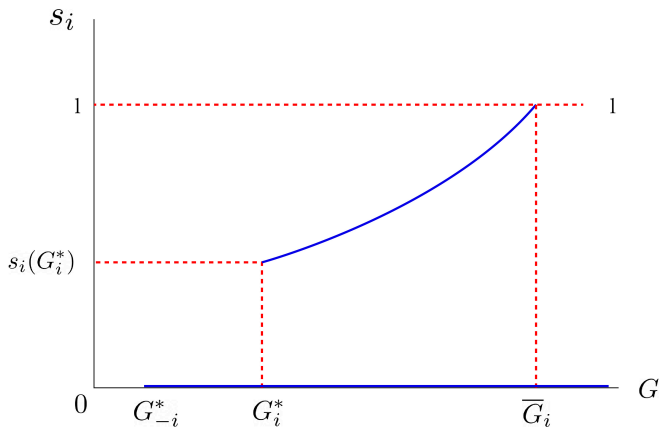


Figure 3.4: The two components of player i's share correspondence when $\frac{\rho}{p_i\Delta} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)}\frac{\beta_i^{1/\rho}}{\alpha}$

An illustrative example. To conclude this section, we now illustrate the types of equilibria that emerge in Teamwork Games by means of a simple example. For this purpose, we retain the Cobb-Douglas-inspired preferences introduced earlier:

$$\widetilde{u}_i(x_i, \sigma(G)) = x_i^{\alpha}\,\sigma(G)$$

To make the comparison more concrete, we assume that players value their private time quadratically more than the group product, setting $\alpha = 2$ for all i. For the evaluation function, we use a logistic specification:

$$\sigma(G) = \frac{d}{1 + e^{-\kappa(G - G^{\mathcal{T}})}}$$

which maps group products into evaluations within the interval $[0, d]$. The steepness parameter κ controls how quickly the evaluation rises with G. The parameter $G^{\mathcal{T}}$ represents the passing point: when $G = G^{\mathcal{T}}$, the evaluation equals $d/2$.

Table 3.3 shows numerically computed equilibria for a heterogeneous team with expertise vector $\mathbf{p} = (0.3, 0.8)$, and evaluation parameters $\kappa = 2$ and $d = 10$. Please notice that this choice of the evaluation parameters is arbitrary and it was made for presentation purposes. In general, the only requirement is that both are strictly positive: $d > 0$, $\kappa > 0$. We consider three evaluation treatments: easy ($G^{\mathcal{T}} = 3$), medium ($G^{\mathcal{T}} = 5$), and hard ($G^{\mathcal{T}} = 7$) across additive, conjunctive, and disjunctive tasks.

Table 3.3: **Numerically computed Nash equilibria for a heterogeneous team** with $\mathbf{p} = (0.3, 0.8)$ across nine Teamwork Games of varying difficulty: easy, medium, and hard, for additive (add.), conjunctive (conj.), and disjunctive (disj.) tasks.

Task Type	**Easy**	**Medium**	**Hard**
Add. ($\rho = 1$)	(0%, 45%)	(5%, 64%)	(33%, 75%)
Conj. ($\rho = -500$)	(56.8%, 21.4%)	(60.1%, 22.6%)	(60.2%, 22.6%)
Disj. ($\rho = 500$)	(0%, 45.1%)	(0%, 65.9%)	(0%, 82.2%)

Each equilibrium is represented as a vector (a_1, a_2), where a_i is the percentage of player i's available time devoted to the teamwork task. These time allocations are then weighted by each player's expertise to determine the group product.

Several patterns are immediately visible. First, as tasks become harder (from left to right in the table), players allocate more time to teamwork across all task types. Second, the nature of the task substantially shapes equilibrium behaviour:

- **Additive tasks.** Gifts from both players add linearly to the group product. As difficulty increases, both players steadily raise their input. In the hardest additive case, both players are contributing substantially.

- **Conjunctive tasks.** Group product is determined by the lowest-performing member. Here, the rational strategy is for players to match the smallest gift, since excess effort

is wasted if another player contributes less. Indeed, multiplying a_i by p_i shows that both players contribute approximately the same effective amount in equilibrium. The small discrepancies visible in Table 3.3 are due to numerical methods in equilibrium computation.

- **Disjunctive tasks.** Group product is determined by the best-performing member. Equilibria therefore involve one player contributing while the other free-rides. In this example, the higher-expertise player always ends up contributing, while the weaker player contributes nothing. This reflects the multiplicity of equilibria in disjunctive games: different subsets of players can, in principle, be active. Here, however, only the stronger member contributes in equilibrium.

In summary, additive and conjunctive Teamwork Games admit unique and well-behaved Nash equilibria, while disjunctive tasks may generate multiple equilibria corresponding to different sets of active contributors. The type of teamwork task critically shapes equilibrium patterns: additive tasks spread effort across members, conjunctive tasks equalise gifts at the weakest link, and disjunctive tasks concentrate effort in the strongest member with scope for free-riding by others. The evaluation function $\sigma(G)$ plays a crucial role in shaping these outcomes by affecting the incentives to contribute through the ratio $\sigma(G)/\sigma'(G)$.

This concludes the theoretical contribution of the chapter: a formal model of teamwork that elucidates how different structural parameters influence the individual decision to cooperate when participation is voluntary. The next section introduces a simulation environment designed to close the theory-simulation loop of this chapter.

3.5 CollabQuest: A Multi-Agent Multi-Armed Bandit for Teamwork

Having solved Teamwork Games analytically, this section now introduces ***CollabQuest***, a psychologically informed teamwork environment cast as a Multi-Agent Multi-Armed Bandit (MA-MAB), in which independent agents learn gift strategies by trial and error. Whereas the previous sections formalised the theoretical mechanisms of structural alignment, this section tests whether multi-armed bandits can reconstruct the equilibria derived above through learning. The multi-agent framework underlying *CollabQuest* is publicly available and can

be consulted at the project's GitHub repository: https://github.com/aloaberasturi/CollabQuest/.

In contrast to much prior work in MAS and MARL, which typically assumes cooperative interpretations of teamwork and is seldom grounded in psychological theory, *CollabQuest* models mixed-motive, decision-focused teamwork with **i)** heterogeneous abilities, **ii)** outcome interdependence, and **iii)** discrete effort choices, while retaining analytical transparency. It rests on three conceptual pillars:

1. **Group product via a CES aggregator.** A single Constant Elasticity of Substitution (CES) function serves as the aggregator across Steiner's task types, parameterising additive, conjunctive, and disjunctive tasks through the elasticity parameter ρ and agent weights β_i.

2. **Independent learning under heterogeneous team composition.** Agents differ in expertise p_i and role weights β_i and learn independently from bandit feedback, treating other agents as part of the environment dynamics.

3. **Dual motivational structure.** Individual utility combines private leisure and a shared evaluation $\sigma(G)$ of the group product, reflecting the coexistence of selfish and collective motives in teamwork.

3.5.1 Teamwork Game (one-shot).

Let the underlying Teamwork Game be $(\Gamma, \mathcal{T}, \sigma)$, where the unitary team task $\mathcal{T} = (G^{\mathcal{T}}, \Delta, \Theta)$ has one episode of duration $\Delta > 0$ and typology parameters $\Theta = \{\rho, \beta_1, \ldots, \beta_n\}$. Each agent $i \in \{1, \ldots, n\}$ selects a discrete effort allocation $a_i \in \mathcal{A}_i = \{0.00, 0.01, \ldots, 1.00\}$, interpreted as the fraction of Δ devoted to the task. With fixed expertise $p_i > 0$ and weight $\beta_i > 0$, the induced gift is

$$g_i \;=\; a_i \, p_i \, \Delta$$

Group product aggregates through a CES operator:

$$G \;=\; \left(\sum_{i=1}^{n} \beta_i \, g_i^{\rho} \right)^{\frac{1}{\rho}}$$

and the (one-shot) pass condition is $G \geq G^{\mathcal{T}}$. Outcome interdependence is captured by an evaluation function $\sigma(G)$ common to all agents.

Rewards

Let x_i denote the leisure (private contribution) enjoyed by agent i during the turn. Under the convention $l_i = 1$ for all i (Section 3.4):

$$x_i \;=\; \Delta - \frac{g_i}{p_i} \;=\; \Delta\,(1 - a_i)$$

Individual utility combines private leisure with the shared evaluation of the group product:

$$r_i(G, \boldsymbol{a}) \;=\; x_i^{\alpha}\,\sigma(G) \;=\; \Delta^{\alpha}\,(1 - a_i)^{\alpha}\,\sigma(G) \in \mathbb{R}$$

where $\alpha > 0$ tunes the curvature of the leisure component.

Action and Observation Spaces

Each action space $\mathcal{A}_i$ comprises $10^2{+}1 = 101$ discrete effort levels with 0.01 granularity. The observable state is the scalar group product G (common to all), which is input to the evaluation function $\sigma(G)$.

3.5.2 *Learning Setup*

With the Teamwork Game formalised as a one-shot decision problem, agent learning is specified within a MA-MAB framework:

Environment

The environment executes a single interaction per episode. Agents simultaneously choose $a_i \in \mathcal{A}_i$. This generates gifts $g_i = a_i p_i \Delta$, which are aggregated via CES to produce G. The evaluation $\sigma(G)$ is computed, and pay-offs $r_i(G, \boldsymbol{a}) = x_i^{\alpha}\sigma(G)$ are returned. Episodes terminate immediately after this turn, and subsequent episodes provide fresh bandit feedback for learning. This formulation accommodates both smooth and binary outcome interdependence by choice of σ.

Agents

The set $\mathcal{I} = \{1, \ldots, n\}$ consists of n independent k-armed bandits, one per agent. Each maintains action-value estimates $Q_i^t(a)$ over $a \in \mathcal{A}_i$ and selects actions via a smooth best response (Boltzmann) with temperature $\tau > 0$:

$$\mathbb{P}_i(a) \;=\; \frac{\exp\!\big(Q_i^t(a)\,/\,(Q_t^{\max}\,\tau)\big)}{\sum_{a' \in \mathcal{A}_i} \exp\!\big(Q_i^t(a')\,/\,(Q_t^{\max}\,\tau)\big)}, \qquad Q_t^{\max} \overset{\text{def}}{=} \max_{a' \in \mathcal{A}_i} Q_i^t(a') \tag{3.22}$$

Let R_i^t denote the *realised* reward, drawn from expectation $r_i(G, \boldsymbol{a})$. After observing the realised reward R_i^t, agent i updates only the action value of the arm actually played, denoted $\tilde{a}_i^t$:

$$Q_i^{t+1}(\tilde{a}_i^t) \;=\; Q_i^t(\tilde{a}_i^t) \;+\; l_r^t \left(R_i^t - Q_i^t(\tilde{a}_i^t)\right),$$

while $Q_i^{t+1}(a) = Q_i^t(a)$ for all $a \neq \tilde{a}_i^t$. The learning rate parameter, denoted by $l_r^t \in (0, 1]$, is the same for all agents and decays with the inverse of time as per Eq. (3.23), where $\gamma > 0$ is determined using an automatic hyperparameter optimisation framework [Akiba et al., 2019]:

$$l_r^t = \left(\frac{\gamma}{\gamma + t}\right) \tag{3.23}$$

A constant learning rate could in principle have been used in the proposed MAMAB implementation. In that case, agents would retain a persistent sensitivity to recent rewards, allowing their action-value estimates to track changes in the empirical reward landscape induced by the simultaneous adaptation of the other agents. This is a common choice in non-stationary multi-agent learning settings.

On the other hand, a decaying learning rate progressively reduces stochastic fluctuations in the Q-value estimates and favours stabilisation of the learned Boltzmann policies. In other words, the decaying step size trades off adaptivity for stability, which is appropriate here because the objective of the experiment is to assess whether agents settle into stable behaviour consistent with the predictions of the underlying static game, rather than to track a continually changing environment. In combination with Boltzmann exploration, this leads to smoothed best-response behaviour, with the learning rate primarily affecting the stability of the dynamics and the temperature parameter controlling the degree of stochasticity in action selection.

3.6 Experiments

CollabQuest instantiates one-shot teamwork among heterogeneous agents learning independently via a discrete Multi-Agent Multi-Armed Bandit (MA-MAB) framework, under a configurable CES interdependence structure and psychologically grounded incentives. Instead

of adopting a normative or prescriptive approach, it does not assume non-competing interests, shared states, or additive combinations of individual effort. This section asks five core questions about this setting and its connection to the game-theoretic analysis in this chapter:

- **Experiment 1: Do MA-MAB learners recover the Nash predictions of Teamwork Games?**
  Pairs of bandit agents are trained across task types and passing thresholds, then the empirical group product $\widetilde{G}$ at convergence is compared against the theoretical equilibrium aggregate $\hat{G}$ derived earlier. Alignment is quantified with scatter and density analyses and Q-table convergence is monitored.

- **Experiment 2: How do team composition and task interdependence shape group performance?**
  Holding the evaluation treatment fixed within each panel, the CES exponent $\rho \in \{-10, 1, 10\}$ (conjunctive, additive, disjunctive) is swept and all two-agent teams drawn from $p_i \in \{0.3, 0.5, 0.7, 0.9\}$ are evaluated. Resulting heatmaps summarise how G varies across different pairings, directly contrasting the qualitative predictions derived from Steiner's taxonomy of tasks.

- **Experiment 3: How does evaluation hardness interact with composition?**
  For each task type, the passing threshold $G^{\mathcal{T}} \in \{3, 5, 7\}$ (soft, medium, hard) is varied and column-wise changes in G are tracked. Feasibility boundaries, diminishing returns at higher difficulty, and composition-specific sensitivity are identified.

- **Experiment 4: What individual effort patterns underlie the aggregate effects?**
  The learned strategies a_i by expertise and evaluation level are reported, focusing on the additive case and providing summaries for conjunctive/disjunctive, examining analogues of the Köhler effect (conjunctive) and social loafing (disjunctive), and quantifying how mean effort increases as the passing threshold $G^{\mathcal{T}}$ becomes stricter.

- **Experiment 5: What changes under a binary pass-fail evaluation?**
  Finally, by replacing the logistic with a Heaviside reward, robustness and strategic polarisation under binary evaluation are studied, assessing whether the qualitative patterns observed under continuous evaluation persist.

Ultimately, the goal of these five experiments is to examine whether learning dynamics in *CollabQuest* reproduce (Experiment 1) or extend (Experiment 5) the equilibrium predictions of

Teamwork Games. Experiments 2-4 delve into the impact of task types, team compositions, and evaluation functions on group and individual behaviour.

Simulations and Hyperparameters. Each episode is one-shot by default: agents act simultaneously, their gifts are aggregated via the CES operator, the evaluation $\sigma(G)$ is computed, and then payoffs are returned. Actions are discretised effort levels—represented as percentages of time—with granularity `gr_act`, corresponding to work bursts in $[0, \Delta]$, thereby mirroring the continuous effort variables of the analytical model. Unless stated otherwise, agent weights are equal ($\beta_1 = \beta_2 = 1$), and the standard leisure-time parameterisation $l_1 = l_2 = 1$ is adopted. The per-episode (one-shot) reward for agent i is inherited from the teamwork utility:

$$r_i \equiv \widetilde{u}_i(x_i, \sigma(G)) = x_i^{\alpha}\, \sigma(G) = \Delta^{\alpha}(1 - a_i)^{\alpha}\, \sigma(G)$$

Throughout the following five experiments, a single turn lasts $\Delta = 10$ time units and $\alpha = 2$ for all agents (free time is squaredly more important than the evaluation of group product). *Both* $\Delta = 10$ *and* $\alpha = 2$ *are convenient choices adopted to enhance clarity in presenting the results; in principle, any* $\Delta, \alpha > 0$ *could be used.* The observable group product G is stored at a discretisation set by `gr_st` for indexing the value table.

Agents learn independently as k-armed bandits and action selection is Boltzmann with a low temperature ($\tau = 0.1$) as in Leslie and Collins [2005]—see Chapter 2 for further reference—so that smooth best responses approximate best replies. The soft-max temperature is fixed and the learning rate $l_r^t \in (0, 1]$ decays with the inverse of time as per Eq. (3.23), with the hyperparameter γ being tuned via Bayesian optimisation with Optuna [Akiba et al., 2019].

The evaluation function should offer a continuous mapping from group product to perceived merit. As discussed below, the logistic function (Fig. 3.5) meets these requirements:

$$\sigma(G) = \frac{d}{1 + \mathrm{e}^{-\kappa(G - G^{\mathcal{T}})}} \tag{3.24}$$

The logistic function takes the group product G as input and returns an evaluation within the interval $[0, d] \subset \mathbb{R}$. The parameter κ indicates how many units of the group product are necessary to raise the evaluation by one point. When $G = G^{\mathcal{T}}$, the group product receives an evaluation of $\frac{d}{2}$, making $G^{\mathcal{T}}$ the passing threshold. Notably, the logistic function allows creating a continuous approximation of qualitative evaluations, similar to a gradable pass-fail system. The main properties of the logistic function are as follows:

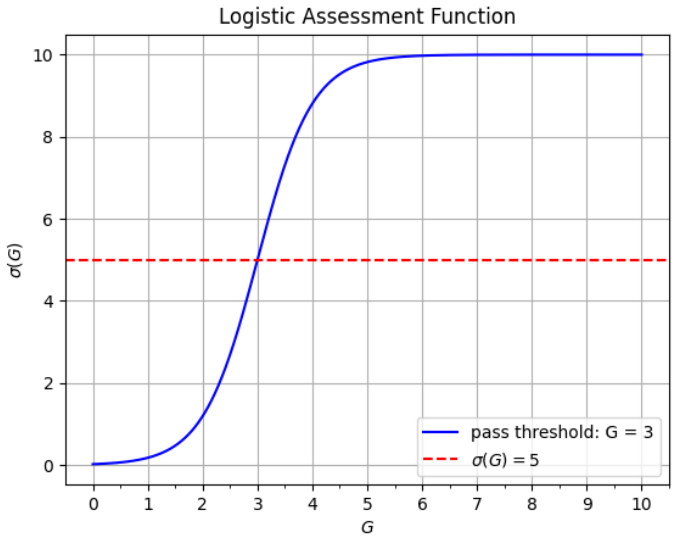


Figure 3.5: Logistic evaluation function with passing threshold $G^{\mathcal{T}} = 3$, steepness parameter $\kappa = 2$, and right-hand asymptote $d = 10$.

- Non-negative evaluation: The function approaches the right horizontal asymptote, $\bar{\sigma} = d$, representing the maximum attainable reward. The left asymptote is $\underline{\sigma} = 0$.
- *Evaluation steepness*: The curve steepness, controlled by parameter κ, determines the rate of evaluation change concerning G.
- *Passing threshold*: The parameter $G^{\mathcal{T}}$ of the curve represents the midpoint of the evaluation criteria, and it determines the passing threshold.

As discussed in Section 3.4.2, the ratio $\frac{\sigma(G)}{\sigma'(G)}$ indicates how costly it is for players to invest their time in improving their evaluation rather than enjoying their leisure time. In the case of the logistic function, the parameters $G^{\mathcal{T}}$ and κ jointly provide this information. If $G^{\mathcal{T}}$ decreases, then the passing threshold lowers, making it easier to pass (and making it less worthwhile to work on the task). If κ decreases, then the slope of the curve decreases, and more group product units are needed to increase the evaluation by one point (and also inhibiting gifts). Therefore, $G^{\mathcal{T}}$ and κ provide information about the *opportunity cost* of the evaluation in terms of enjoyed leisure. In this chapter, κ is kept constant in all experiments for the sake of parsimony. Consequently, changes in $G^{\mathcal{T}}$ across experiments modulate the ratio $\frac{\sigma(G)}{\sigma'(G)}$. This effect is hereafter referred to as *evaluation hardness*.

Table 3.4: **Fixed hyperparameters and environment settings** used across simulations.

Parameter	Value
Temperature (soft-max)	$\tau = 0.1$
Learning-rate schedule	$l_r^t = \left(\frac{\gamma}{\gamma+t}\right)$
Evaluation $\sigma(G)$	Logistic or Heaviside (configurable)
Right-hand asymptote of $\sigma(G)$	$d = 10$
Effort granularity	`gr_act` $= 2 \Rightarrow 101$ arms on $[0, 1]$
State granularity	`gr_st` $= 3$ (indexing for Q)
Turn duration	$\Delta = 10$

The logistic function in Eq. (3.24) fulfils all the criteria to serve as an evaluation function (Def. 3.6). The resulting reward, expressed as a function of agent i's leisure x_i and the group product G, $\widetilde{u}_i(x_i, \sigma(G))$, remains non-negative across all values.

Figure 3.6 depicts the utilities for each agent in a two-agent team: Agent 1 with $p_1 = 0.5$ and Agent 2 with $p_2 = 0.8$. The logistic evaluation function is employed. The task is additive ($\rho = 1$) with a passing threshold $G^{\mathcal{T}} = 5$ and turn duration $\Delta = 10$. Notice that the most favourable scenario for Agent 1 occurs when their teammate contributes maximally ($g_2 = 10$) while they refrain from any action ($g_1 = 0$). When neither agent contributes ($g_1 = g_2 = 0$), both receive utilities close to zero, reflecting the low evaluation outcome. However, neither agent finds it optimal to contribute maximally, as this would entail forfeiting all leisure time while yielding negligible additional utility.

For the specific case depicted in Figure 3.6, the equilibrium aggregates of individual gifts (in work units) are $\left(g_1^* = 1.45, g_2^* = 4.48\right)$.

Training budget and convergence. Agent expertise vectors are sampled systematically from predefined sets to span homogeneous ($\sigma_p < 0.145$) and heterogeneous ($\sigma_p > 0.145$) team compositions. The same catalogue of team configurations is reused across task types and difficulties for comparability. Unless otherwise stated, each configuration is trained for a fixed budget of `n_episodes` $= 5 \times 10^4$ one-shot episodes. After each episode, agents update the value of the chosen arm only. The learning rate is decayed as per Eq. (3.23). In other words, the agents learn through repeated play to master what is, in essence, a single-turn game.

At evaluation time, each agent plays its greedy action (arg max over its learned Q-row at the pre-decision state). The resulting action profile is executed once in the environment to obtain G and R_i, which are recorded together with the learned Q-tables for reproducibility. In

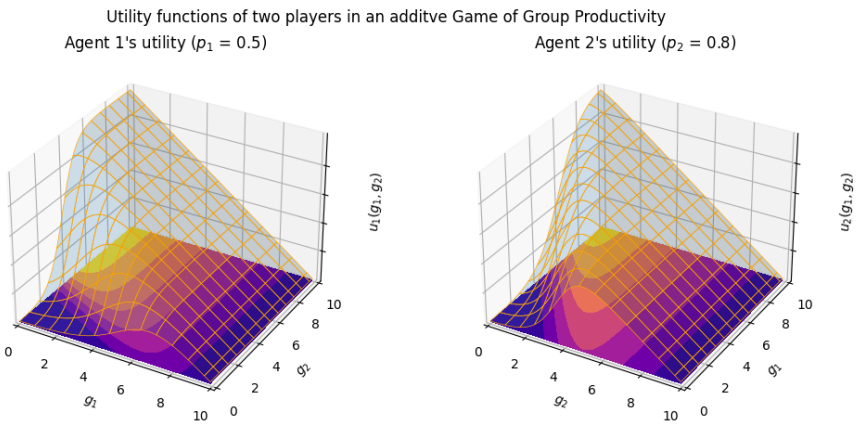


Figure 3.6: **Utilities $\widetilde{u}_i$ of two agents in a team**, shown as functions of the agent's work units gift g_i and the rest of the team's gift G_{-i}. Left: Utility for an agent with expertise $p_1 = 0.5$. Right: Utility for their teammate with expertise $p_2 = 0.8$.

practice, returns stabilise well before the end of the episode budget for the ranges of ρ and $\{p_i\}$ explored here. Increasing `n_episodes` beyond 5×10^4 yields negligible changes in the selected greedy actions for the settings reported.

This setup provides a direct computational instantiation of the analytical model developed earlier, allowing the following experiments to test whether learning dynamics recover or deviate from the predicted equilibria.

3.6.1 *Experiment 1: Do Multi-Agent Multi-Armed Bandit Teams Learn Nash-consistent policies?*

The first experiment addresses the following question: *do independently learning bandits converge to strategies that approximate the Nash Equilibria (NEs) of the underlying Teamwork Games?*. To address this, empirical outcomes from the MA-MAB setting are compared with the closed-form NE aggregates derived earlier in the chapter.

This experiment thus tests whether decentralised learning recovers the analytically predicted equilibria under varying structural and evaluative conditions.

Methods

A total of 140 two-agent experiments are conducted using expertise combinations drawn from $p_i \in [0.1, 0.9]$. Additionally, different task variants are explored, ranging from strongly con-

junctive tasks ($\rho = -100$) to strongly disjunctive tasks ($\rho = 100$) (Table 3.5). Evaluation hardness levels are manipulated through the passing threshold $G^{\mathcal{T}} \in \{3, 5, 7\}$, corresponding to **soft**, **medium**, and **hard** evaluation treatments (Table 3.6).

Table 3.5: **Task types used in the experiments**, characterised by the value of ρ

Strongly Conjunctive	Quasi-Conjunctive	Additive	Quasi-Disjunctive	Strongly Disjunctive
$\rho \in \{-100, -10\}$	$\rho \in \{-3, 0.5\}$	$\rho = 1$	$\rho = 3$	$\rho \in \{10, 100\}$

Table 3.6: **Evaluation conditions used in the experiments**. Passing thresholds $G^{\mathcal{T}} \in \{3, 5, 7\}$ are expressed in terms of *work* units.

Soft	Medium	Hard
$G^{\mathcal{T}} = 3$	$G^{\mathcal{T}} = 5$	$G^{\mathcal{T}} = 7$

Each configuration $(p_1, p_2, \rho, G^{\mathcal{T}})$ in the experimental grid is trained for 5×10^4 one-shot episodes. Agents update only the value of the chosen arm incrementally:

$$Q^{t+1}(\tilde{a}) \leftarrow Q^t(\tilde{a}) + l_r^t\big(R^t - Q^t(\tilde{a})\big)$$

with a diminishing learning rate $l_r^t \in (0, 1]$ that decays with the inverse of time as per Eq. (3.23). Convergence is monitored by the maximum entry-wise change in the Q-table,

$$\max_a \left|Q^{t+1}(a) - Q^t(a)\right| \leq \delta \qquad \text{(for small } \delta\text{)} \tag{3.25}$$

and by tracking the evolution of the greedy action $\operatorname{argmax}_a Q^t(a)$ over episodes (Fig. 3.7). After training, the learned greedy profile $\widetilde{\boldsymbol{a}} = (a_1, a_2)$ induces an empirical group product $\widetilde{G}$. This $\widetilde{G}$ is then compared with the theoretical NE aggregate $\hat{G}$ computed from the closed-form analysis in Section 3.4.2, and the correspondence is visualised through scatter and kernel-density plots (Fig. 3.8).

Results and Discussion

The learned aggregates align closely with the theoretical NEs aggregates across task types and expertise pairings (Fig. 3.8). A linear fit of $\widetilde{G}$ on $\hat{G}$ yields parameters ($\alpha = 0.99$, $\beta = 0.02$) and goodness of fit $\xi^2 = 0.992$—essentially the identity line $y = x$. Convergence diagnostics show stabilisation of Q-values under the criterion in Eq. (3.25) and stable optimal-arm selections over time (Fig. 3.7). Because τ is kept low, smooth best responses concentrate on

near-best replies [Leslie and Collins, 2005]; consequently, the empirical aggregates can be interpreted as near-equilibrium aggregates induced by near-NE strategies. This confirms that decentralised bandit learning, even under independent adaptation, is sufficient to recover the strategic profiles predicted analytically for Teamwork Games.

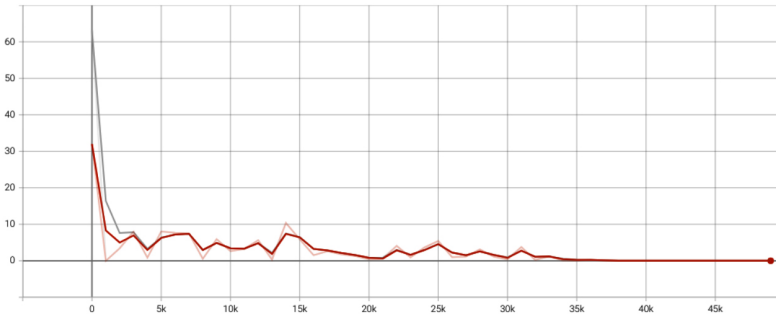


Figure 3.7: **Representation of the convergence of Q-values during training of two agents.** Timesteps are represented on the horizontal axis and the difference $\max_a\{Q_{n+1}(a) - Q_n(a)\}$ is represented on the vertical axis.

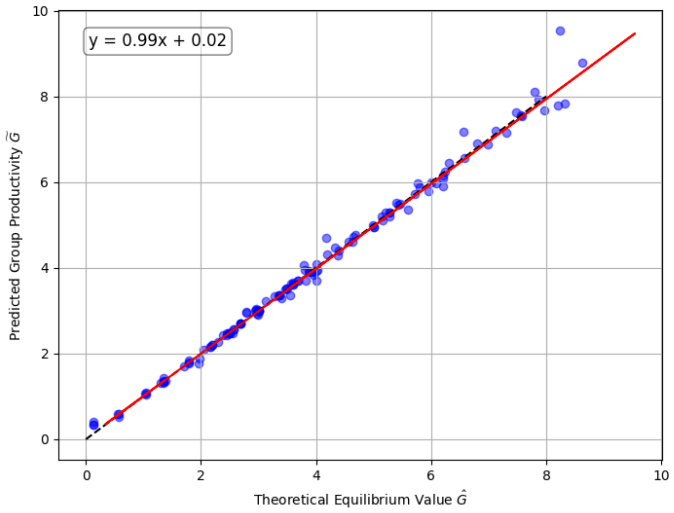


Figure 3.8: **Scatter plot showing the correspondence between theoretical Nash Equilibrium (NE) aggregate productivity values (horizontal axis) and the empirical team productivity (vertical axis) derived from the MA-MAB setting.** The dashed black line represents the identity line $y = x$, while the red line shows a linear regression fit with parameters $\alpha = 0.99, \beta = 0.02$ and a goodness of fit of $\xi^2 = 0.992$. This close alignment underscores that, upon convergence, the aggregate values closely approximate those expected under NE, suggesting that the agents' learned strategies are near-equilibrium policies.

Key Takeaways

In summary:

- In the one-shot MA-MAB setting, independently learning bandits converge to policies whose induced aggregates $\widetilde{G}$ closely match the theoretical NE group products $\hat{G}$ across ρ and (p_1, p_2).
- Low-temperature soft-max action selection and incremental value updates suffice to recover NE-consistent group products, with tight empirical-theoretical correspondence ($\xi^2 \approx 0.992$).

Having established that independent learners can reproduce the Nash-consistent structure of Teamwork Games, the next experiment examines how team composition and task interdependence jointly shape performance

3.6.2 *Experiment 2: Interaction Between Team Composition and Task Type*

The second experiment examines how team productivity varies jointly with team composition—homogeneous *versus* heterogeneous expertise—and task type, using the same two-agent MA-MAB set-up as before.

Methods

All pairs formed by $p_i \in \{0.3, 0.5, 0.7, 0.9\}$ are considered and a two-agent system is trained for each configuration under three task regimes: conjunctive ($\rho = -10$), additive ($\rho = 1$), and disjunctive ($\rho = 10$). Evaluation hardness varies via the passing threshold $G^{\mathcal{T}} \in \{3, 5, 7\}$ (soft/medium/hard). Parameters are fixed as $\beta_1 = \beta_2 = 1$, $\Delta = 10$, and leisure exponent $\alpha = 2$. Each configuration $(p_1, p_2, \rho, G^{\mathcal{T}})$ is trained for 5×10^4 one-shot episodes with Boltzmann action selection ($\tau = 0.1$) and the incremental Q update. After training, the greedy profile is executed and the induced group product G is recorded. Please notice that both $\Delta = 10$ and $\alpha = 2$ are convenient choices adopted to enhance clarity in presenting the results; in principle, any $\Delta, \alpha > 0$ could be used.

Results are visualised as a 3×3 grid of heatmaps (Fig. 3.9): *rows* vary task type (ρ) and *columns* vary evaluation hardness ($G^{\mathcal{T}}$). Within each heatmap, moving *perpendicularly* to the main diagonal contrasts homogeneous with increasingly heterogeneous pairs at fixed mean expertise.

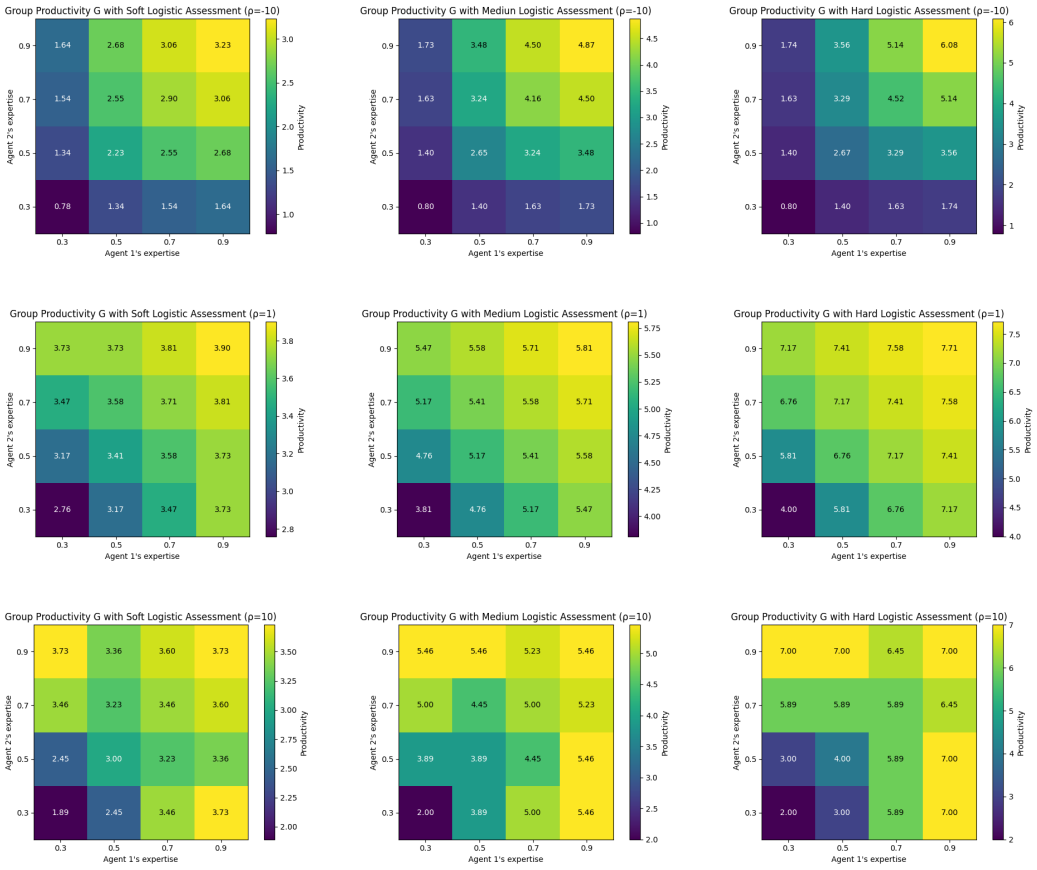


Figure 3.9: **Variation in team productivity across teams and tasks.** Top row: conjunctive ($\rho = -10$). Middle: additive ($\rho = 1$). Bottom: disjunctive ($\rho = 10$). Columns vary evaluation hardness—soft ($G^{\mathcal{T}} = 3$), medium ($G^{\mathcal{T}} = 5$), hard ($G^{\mathcal{T}} = 7$). Dark blue denotes low productivity; yellow denotes high productivity, with the passing threshold marking the colour transition.

Results and Discussion

For this discussion, the focus is on how the group product changes *across rows* in Fig. 3.9 (*i.e.*, as task type varies), holding the column (evaluation treatment) fixed.

Conjunctive tasks (top row, $\rho = -10$). As one moves perpendicularly away from the diagonal of homogeneous teams, group product drops. Under the *medium* evaluation treatment ($G^{\mathcal{T}} = 5$), productivity decreases by **38**%: from $G = 2.65$ for the homogeneous pair $\mathbf{p} = (0.5, 0.5)$ to $G = 1.63$ for the heterogeneous pair $\mathbf{p} = (0.3, 0.7)$. This pattern corroborates Steiner's assertion that **conjunctive tasks are better performed by homogeneous teams** (of

similar expertise) than by heterogeneous ones [Steiner, 1972], at least under the utility specification used here.

Additive tasks (middle row, $\rho = 1$). These consistently yield the highest group products across evaluation treatments, aligning with findings that **additive tasks typically yield team potential exceeding that of the best individual member**. Visually, for a given threshold $G^{\mathcal{T}}$, the group product remains nearly constant perpendicular to the diagonal—for example, $(p_i, p_j) = (0.5, 0.7)$ and $(0.3, 0.9)$ display similar G values. This supports Steiner's hypothesis that **heterogeneity in team composition does not materially affect group product in additive tasks**.

Disjunctive tasks (bottom row, $\rho = 10$). Productivity sits between the conjunctive and additive regimes, but unlike the conjunctive case, *heterogeneity in team composition improves performance*. Under *medium* evaluation ($G^{\mathcal{T}} = 5$), the homogeneous pair $(0.5, 0.5)$ yields $G = 3.89$, whereas the heterogeneous pair $(0.3, 0.7)$ yields $G = 5.00$, an **increase of** 29%. This is consistent with best-shot dynamics and with Steiner's proposition that **disjunctive tasks can be more effectively tackled by teams with diverse expertise** than by uniformly skilled teams.

Taken together, these results confirm that the learning dynamics in *CollabQuest* reproduce classic predictions from the teamwork literature, with conjunctive tasks penalising heterogeneity and disjunctive tasks benefiting from it.

Key Takeaways

In summary:

- In **conjunctive ($\rho = -10$)** tasks, productivity declines as teams become more heterogeneous; homogeneous pairs perform best (*e.g.*, 2.65 $\rightarrow$ 1.63, -38%).
- In **additive ($\rho = 1$)** tasks, productivity is overall the highest. For a fixed evaluation hardness $G^{\mathcal{T}}$, group products are largely invariant to heterogeneity at a given mean expertise (*i.e.*, $(0.5, 0.7) \approx (0.3, 0.9)$).
- For **disjunctive ($\rho = 10$)** tasks, heterogeneity boosts productivity (*e.g.*, 3.89 $\rightarrow$ 5.00, $+29\%$), reflecting best-shot effects.

3.6.3 *Experiment 3: Interaction Between Team Composition and Evaluation Treatment*

Building on the theoretical analysis in Section 3.4, this experiment tests whether increasing the difficulty of the evaluation (via the passing threshold $G^{\mathcal{T}}$, *cf.* Eq. (3.24)) systematically alters empirical group product in the MA-MAB setting.

Methods

To make these comparisons, Fig. 3.9 on page 79 is revisited, focusing on changes in G **column-wise** (*i.e.*, within task type, as $G^{\mathcal{T}}$ increases from soft $\rightarrow$ medium $\rightarrow$ hard).

Results and Discussion

Additive tasks (middle row, $\rho = 1$). As $G^{\mathcal{T}}$ rises, more teams fail the evaluation. In the *soft* evaluation treatment ($G^{\mathcal{T}} = 3$), all teams pass except the pair with two low-expertise agents ($p_1 = p_2 = 0.3$). In the *hard* evaluation treatment ($G^{\mathcal{T}} = 7$), only teams whose *average* expertise reaches at least 60% of the maximum possible expertise pass:

$$\frac{1}{n}\sum_{j=1}^{n} p_j \;\geq\; 60\%$$

For a *fixed team*, productivity generally increases with $G^{\mathcal{T}}$; however, the increments consistently shrink whenever the resulting G falls short of the new threshold. A consistent behavioural pattern emerges: when the task is easy enough, raising $G^{\mathcal{T}}$ increases productivity—most sharply at the point where passing becomes feasible at the next hardness level. This effect naturally stalls once requirements exceed team capacity.

Conjunctive tasks (top row, $\rho = -10$). Teams perform worse overall than in the additive case, but higher thresholds can still elicit larger gifts *when* agents have spare capacity (*i.e.*, when they can increase effort without sacrificing too much leisure utility). In the *soft* evaluation treatment, only the top two teams $\mathbf{p} = (0.7, 0.9)$ and $\mathbf{p} = (0.9, 0.9)$ succeed; for the rest, raising $G^{\mathcal{T}}$ often fails to move G enough because the marginal cost of effort dominates the prospective gain in $\sigma(G)$. In such intrinsically challenging conjunctive settings, improving performance would likely require *increasing* the turn duration Δ or *softening* evaluation by reducing $G^{\mathcal{T}}$.

Disjunctive tasks (bottom row, $\rho = 10$). All teams respond to the initial increase in evaluation hardness (from $G^{\mathcal{T}} = 3$ to $G^{\mathcal{T}} = 5$), and teams with at least one high-expertise member ($p_i \geq 0.7$) also respond to the subsequent escalation (from $G^{\mathcal{T}} = 5$ to $G^{\mathcal{T}} = 7$). Concretely, $(0.7, 0.9)$ increases G from 4.5 to 5.14 (a 14% rise) when moving to the hard condition, whereas $(0.5, 0.7)$ increases by only $\sim$ 1%. Strikingly, pairs with low expertise ($p_{1,2} \leq 0.5$) can *lose* productivity when shifting from medium to hard. Here, raising the demand threshold has a detrimental effect on output—akin to *loss of motivation* in human teams [Steiner, 1972]—and arises endogenously from the agents' reward-maximising behaviour.

Diminishing returns of hardness. Across additive, conjunctive, and disjunctive tasks, G often increases as the evaluation becomes more stringent, yet the *rate* of increase diminishes: gains from soft→medium typically exceed gains from medium→hard. Once the threshold outpaces team capacity, further hardening ceases to improve—and may reduce—productivity.

Key Takeaways

In summary:

- **Thresholding matters:** Raising $G^{\mathcal{T}}$ generally increases effort and output up to the point where passing becomes feasible, but the effect saturates beyond capacity.
- **Additive tasks:** Clear column-wise progression; in the hard condition, only teams with average expertise $\geq$ 60% pass.
- **Conjunctive tasks:** Hardening evaluation helps only when teams have spare capacity; otherwise it reduces productivity.
- **Disjunctive tasks:** Initial hardening lifts productivity, but further increases benefit teams with strong members while reducing output for low-expertise pairs (*motivation-loss* effect).
- **Diminishing returns:** Productivity gains from increasing $G^{\mathcal{T}}$ are largest at lower hardness and taper off thereafter.

Having examined how evaluation hardness interacts with team structure, the next experiment turns to the individual level, analysing the effort patterns that give rise to these aggregate effects.

3.6.4 *Experiment 4: Individual Strategies*

The previous experiments focused on group product G and its relationship with task type and evaluation hardness. This experiment examines how these same factors shape the *individual* strategies selected by agents. A strategy is defined as the percentage a_i of the turn that agent i allocates to the task (with gift $g_i = a_i\, p_i\, \Delta$; *cf.* Eq. (3.1)).

Methods

The learned strategies are reported for two-agent teams under an *additive* task ($\rho = 1$) across three evaluation treatments (soft/medium/hard), keeping the expertise grid $p_i \in \{0.3, 0.5, 0.7, 0.9\}$ (Table 3.7). For compactness, only the upper triangle of each matrix is shown, as the results are symmetric. Analogous tables for *conjunctive* ($\rho < 1$) and *disjunctive* ($\rho > 1$) tasks are provided in Appendix A.7. Table 3.8 summarises the percentage increase in average team gifts for additive, conjunctive, and disjunctive tasks as the passing threshold is raised from soft to medium and from medium to hard.

Results and Discussion

Additive tasks. Within each pair, the agent with the higher expertise level consistently contributes more time than their counterpart. Raising the evaluation hardness does not diminish individual gifts in any observed case, although there are instances of no discernible effect. For example, in the pair $(p_1 = 0.3,\ p_2 = 0.9)$, the less proficient agent contributes nothing under soft or medium evaluation (free-riding) and only engages with a 20% commitment when the passing threshold reaches $G^{\mathcal{T}} = 7$. Still in the additive case, the average gift of an agent with expertise p_i increases across all teams as the passing threshold rises. However, the transition from a soft to a medium evaluation treatment yields a more pronounced effect than the transition from medium to hard. In other words, agents consistently increase their task engagement as evaluation demands intensify, with a more substantial augmentation when moving from soft to medium than from medium to hard.

Conjunctive tasks. For detailed equilibrium strategies in conjunctive and disjunctive tasks, see Appendix A.7. In conjunctive tasks, the weaker member of each pair contributes a greater proportion of time than their counterpart across all evaluation treatments. This pattern mirrors the **Köhler effect** observed in human teams, where a weaker member performs a difficult task better in a cooperative context than alone [Feltz et al., 2011]. Furthermore, increasing the

evaluation level did not decrease individual gifts in any observed case, although there were some instances with no discernible effect.

Disjunctive tasks. In the disjunctive case, hardening the evaluation can *reduce* individual gifts of others when the expertise of the strongest member is sufficiently high (> 0.7). Consider, for instance, the pair $(p_1 = 0.5,\ p_2 = 0.7)$. In that case, the less proficient agent contributes in one of the equilibria under soft evaluation but free-rides as soon as the passing threshold is heightened. This phenomenon resembles **social loafing** in human teams, whereby some members reduce their effort and rely on others' contributions.

Diminishing returns of hardness and cross-task sensitivity. Table 3.8 reports the percentage increase in the average team gift of each agent as the passing threshold is raised from soft to medium and from medium to hard. In all cases but one ($\rho = 10,\ p_i = 0.5$), the percentage increase as the evaluation hardens is nonnegative: the more demanding the evaluation, the more committed the agents. In the lone exception, the weakest agent learns not to contribute as soon as the task becomes maximally difficult (hard treatment) and their counterpart possesses greater expertise, hence the observed negative difference.

These increments tend to be higher overall for additive tasks than for the other task types, although this is not always the case—the agent with expertise $p_i = 0.9$ always learns to lead regardless of task type, increasing their gifts by a similar amount for a given evaluation treatment.

Looking across Table 3.8, one can also discern that **agents with lower expertise are more sensitive to task type**. For instance, when $p_1 = 0.3$ in an additive task, effort nearly doubles as evaluation hardens from $G^{\mathcal{T}} = 3$ to $G^{\mathcal{T}} = 5$. This jump occurs because the agent does not contribute under soft evaluation ($G^{\mathcal{T}} = 3$) when paired with a stronger teammate ($p_2 = 0.7$). Likewise, moving from medium to hard prompts contribution even alongside a very strong partner ($p_2 = 0.9$).

In conjunctive tasks ($\rho = -10$), the same low-expertise agent consistently contributes, leaving little room for improvement.

In disjunctive tasks ($\rho = 10$), the agent refrains from contributing altogether in any equilibrium where counterparts substantially surpass them in expertise (*i.e.*, $p_2 > 0.5$). This equilibrium is robust to changes in evaluation treatment—**hardening the evaluation does not alter this agent's policy**. In general, **high-expertise agents** ($p_i \geq 0.7$) **display more stable policies**

across task types: they increase their gifts as the passing threshold rises, largely unaffected by the task structure itself.

Table 3.7: **Individual team gift across different teams for an additive task under three different evaluation treatments**. For simplicity, only the upper diagonal of the matrices, which are symmetric, is filled.

(a) Soft Evaluation Treatment ($G^{\mathcal{T}} = 3$)

$p_2^{\mathcal{T}}$ \ $p_1^{\mathcal{T}}$	0.30	0.50	0.70	0.90
0.30	(46%, 46%)	(20%, 52%)	(0%, 50%)	(0%, 41%)
0.50		(34%, 34%)	(16%, 40%)	(0%, 41%)
0.70			(27%, 27%)	(23%, 22%)
0.90				(22%, 22%)

(b) Medium Evaluation Treatment ($G^{\mathcal{T}} = 5$)

$p_2^{\mathcal{T}}$ \ $p_1^{\mathcal{T}}$	0.30	0.50	0.70	0.90
0.30	(64%, 64%)	(46%, 68%)	(20%, 66%)	(0%, 61%)
0.50		(52%, 52%)	(34%, 53%)	(16%, 53%)
0.70			(40%, 40%)	(27%, 43%)
0.90				(32%, 32%)

(c) Hard Evaluation Treatment ($G^{\mathcal{T}} = 7$)

$p_2^{\mathcal{T}}$ \ $p_1^{\mathcal{T}}$	0.30	0.50	0.70	0.90
0.30	(67%, 67%)	(64%, 78%)	(46%, 77%)	(20%, 73%)
0.50		(68%, 68%)	(52%, 66%)	(34%, 63%)
0.70			(53%, 53%)	(40%, 53%)
0.90				(43%, 43%)

Key Takeaways

In summary:

- **Expertise drives effort allocation:** In additive tasks, higher-expertise agents consistently contribute more, while low-expertise partners sometimes free-ride until the evaluation hardens enough to compel engagement.
- **Thresholding increases commitment but with diminishing returns:** Across task types, raising the passing threshold generally raises individual effort. The largest increments occur from soft→medium; gains taper from medium→hard.

Table 3.8: **Percentage increase in average dedication** $\Delta_{\%}(\bar{a}_i)$ of each agent to different tasks as the passing threshold rises

(a) Additive Task ($\rho = 1$)

	Soft to Medium	Medium to Hard
$p_i = 0.3$	97%	52%
$p_i = 0.5$	104%	50%
$p_i = 0.7$	68%	39%
$p_i = 0.9$	45%	35%

(b) Conjunctive Task ($\rho = -10$)

	Soft ($G^{\mathcal{T}} = 3$) to Medium ($G^{\mathcal{T}} = 5$)	Medium ($G^{\mathcal{T}} = 5$) to Hard ($G^{\mathcal{T}} = 7$)
$p_i = 0.3$	5%	0%
$p_i = 0.5$	27%	1%
$p_i = 0.7$	46%	11%
$p_i = 0.9$	53%	24%

(c) Disjunctive Task ($\rho = 10$)

	Soft ($G^{\mathcal{T}} = 3$) to Medium ($G^{\mathcal{T}} = 5$)	Medium ($G^{\mathcal{T}} = 5$) to Hard ($G^{\mathcal{T}} = 7$)
$p_i = 0.3$	6%	0%
$p_i = 0.5$	4%	−36%
$p_i = 0.7$	42%	18%
$p_i = 0.9$	49%	28%

- **Conjunctive tasks elicit a Köhler effect:** The weaker team member consistently contributes more effort than their counterpart, and hardening evaluation does not reduce gifts.
- **Disjunctive tasks risk social loafing:** When a strong member is present, weaker partners may disengage entirely as evaluation hardens, exploiting the stronger agent's efforts.
- **Differential sensitivity to task type:** Low-expertise agents are highly sensitive to whether the task is additive, conjunctive, or disjunctive, whereas high-expertise agents ($p_i \geq 0.7$) follow a steadier pattern, consistently increasing their gifts as thresholds rise.

The final experiment examines whether these patterns persist when the continuous evaluation function is replaced by a binary pass-fail criterion.

3.6.5 *Experiment 5: Binary Pass-Fail Evaluation*

Finally, to test the versatility of the MA-MAB framework beyond smoothly graded evaluations, a *discontinuous* reward regime that mirrors a binary pass-fail evaluation is examined. Let $\mathbb{H}\colon \mathbb{R} \to \{0,1\}$ denote the Heaviside step function:

$$\mathbb{H}(x) \;=\; \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

and define the binary per-agent reward as

$$R_i(G, \boldsymbol{a}) \;=\; \Delta^{\alpha}(1 - a_i)^{\alpha}\, d\, \mathbb{H}(G - G^{\mathcal{T}}) \tag{3.26}$$

where d serves as the right-hand asymptote in Eq. (3.24) and $G^{\mathcal{T}}$ is the passing threshold. To maintain continuity with previous sections, $d = 10$ is used.

Methods

Two-agent teams formed from all ordered pairs of expertise $p_i \in \{0.3, 0.5, 0.7, 0.9\}$ are evaluated (as in Experiment 3.6.2). The analysis focuses on the **additive** task regime ($\rho = 1$) under the **medium** evaluation treatment ($G^{\mathcal{T}} = 5$). For each configuration (p_1, p_2, ρ=1, $G^{\mathcal{T}}$=5), the pair is trained for 5×10^4 one-step episodes using the Heaviside evaluation. To quantify robustness, training is repeated *three* times per configuration, yielding group products $\{\widetilde{G}_1^{\mathbf{p}}, \widetilde{G}_2^{\mathbf{p}}, \widetilde{G}_3^{\mathbf{p}}\}$. The mean $\overline{G}$ is reported in Figure 3.11 and the *percentage dispersion* of each triple is verified to be below 2%:

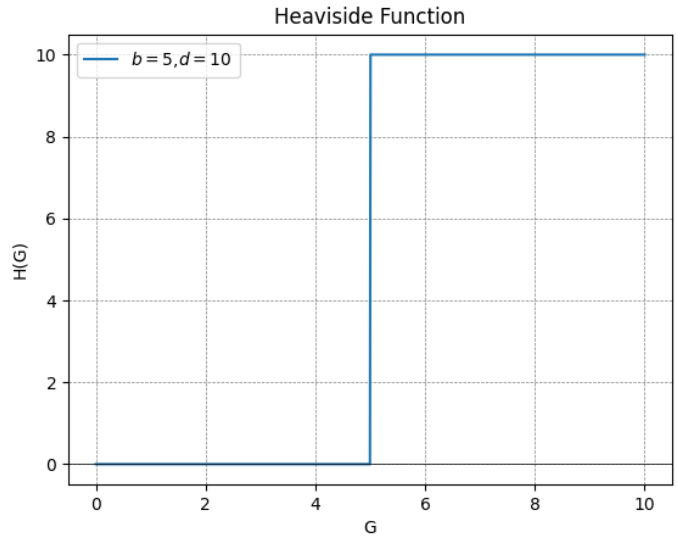


Figure 3.10: Heaviside evaluation with parameters $G^{\mathcal{T}} = 3$ and $d = 10$.

$$\text{Dispersion}(G) = \frac{\max\{\widetilde{G}_1^{\mathfrak{p}}, \widetilde{G}_2^{\mathfrak{p}}, \widetilde{G}_3^{\mathfrak{p}}\} - \min\{\widetilde{G}_1^{\mathfrak{p}}, \widetilde{G}_2^{\mathfrak{p}}, \widetilde{G}_3^{\mathfrak{p}}\}}{\overline{G}} \times 100\%$$

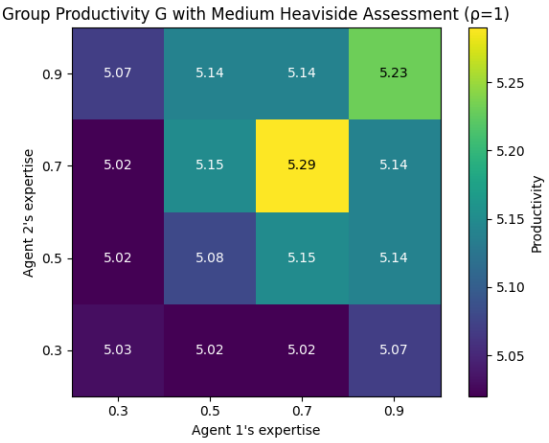


Figure 3.11: **Group product** $\overline{G}$ for the additive task ($\rho = 1$) under medium evaluation ($G^{\mathcal{T}} = 5$) with a Heaviside evaluation. Darker shades indicate group products below the threshold; yellow indicates group products above it.

Results and Discussion

Table 3.9 reports the percentage dispersion of $\{\widetilde{G}_1^{\mathbf{p}}, \widetilde{G}_2^{\mathbf{p}}, \widetilde{G}_3^{\mathbf{p}}\}$ for each team. For compactness, only the upper triangle is shown (the table is symmetric). In all cases, dispersion is $< 2\%$.

Table 3.9: **Percentage dispersion of group products across three independent trainings per team**. All values are $< 2\%$.

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	0.00%	0.40%	0.60%	1.19%
0.50		1.00%	1.78%	0.20%
0.70			0.20%	0.18%
0.90				0.00%

Under the binary pass-fail regime, the group product G is highly stable across repeats (dispersion $< 2\%$) even though the *converged strategies* may differ slightly across trainings. A pronounced qualitative change emerges for **highly heterogeneous pairs**. In particular, for $(p_i, p_j) \in \{(0.3, 0.7), (0.3, 0.9), (0.5, 0.9)\}$ the *less skilled agent* systematically *withdraws* (contributes 0%), leaving the stronger teammate to shoulder the full provision. This behaviour is consistent across repeats (again $< 2\%$ dispersion on G).

This contrasts with the *continuous* evaluation at the same difficulty (Table 3.7b), where full withdrawal at $G^{\mathcal{T}} = 5$ was observed *only* for the maximally heterogeneous pair $(0.3, 0.9)$. The step-like incentive therefore **polarises** strategies relative to the logistic case: when expertise gaps are large (≥ 0.4), the weaker agent prefers to free-ride while the stronger agent carries the effort to reach the threshold. Despite this polarisation, teams consistently *meet* the assignment criterion in Figure 3.11, illustrating that binary evaluation can stabilise "best-shot-like" role specialisation in additive tasks.

Key Takeaways

In short, when applying step-like evaluation functions on Teamwork Games, the main observations are:

- **Discontinuity induces polarisation.** Replacing the smooth evaluation with a Heaviside step causes weaker members of heterogeneous pairs to withdraw, while stronger teammates increase effort to clear the threshold.

- **Stable group products despite policy variability.** Even with slight variability in learned strategies across seeds, final group product G is highly stable (dispersion $< 2\%$).
- **Contrast with continuous evaluation.** Under the logistic evaluation at $G^{\mathcal{T}} = 5$, full withdrawal was confined to the maximally heterogeneous pair; with Heaviside, it extends to pairs with expertise gaps of ≥ 0.4.

Altogether, the first four experiments show that decentralised bandit learning, even under independent adaptation, is sufficient to recover the strategic profiles predicted analytically for Teamwork Games. While Experiments 2-4 clarify how productivity and individual effort are shaped by structural variables such as task type, team composition and evaluation hardness, Experiment 5 extends these findings to discontinuous incentives, showing that most qualitative patterns persist. This reinforces the thesis that the *structure of the teamwork environment*—rather than fine-tuning of the reward function—drives the behavioural regularities uncovered throughout this chapter.

These results close the loop between formal analysis and learning-based simulation, and inform the design of empirical studies such as *The Island* (Appendix C), where analogous incentive schemes can be implemented with human participants. In particular, the structural similarities between logistic and step-like evaluations offer a principled basis for choosing Heaviside-like scoring rules—experimentally more tractable—without disrupting the strategic incentives at the core of Teamwork Games.

With these foundations in place, Chapter 4 moves beyond one-shot interaction and introduces a dynamic model of teamwork. It examines temporal alignment: how agents anticipate one another under sequential dependence, and how temporal coupling reshapes the learned strategies established in the present chapter.



4 *Multi-Agent Sequential Teamwork: The Dynamics of Cooperation*

> Whether the players trust each other or not is less important in the long run than whether the conditions are ripe for them to build a stable pattern of cooperation.
>
> *Robert Axelrod*

4.1 Introduction: Anna's Choice 95
4.2 Teamwork Across Time: Dynamics and Determinants of Group Performance 97
4.3 Sequential Teamwork Games 99
4.4 TEMPO: A Markov Game Environment for Sequential Teamwork 107
4.5 Experiments 112

Chapter 3 began formalising psychological teamwork models as computational frameworks by representing teamwork as a one-time *aggregative game*. This allowed us to characterise the structural alignment mechanisms underpinning voluntary human teamwork—from the role specialisation induced by heterogeneity to the fundamentally different strategic incentives created by additive, conjunctive, and disjunctive tasks.

Yet real-world teamwork unfolds over time. One-shot formalisations, while analytically illuminating, cannot capture how agents adjust their effort in response to previous actions, evolving beliefs about teammates, or temporally extended evaluation criteria. The strategic patterns uncovered in Chapter 3 therefore raise a natural question: *Do these static patterns persist, dissolve, or transform when interaction becomes sequential?*

This chapter answers that question by extending the psychological formalisation developed in Chapter 3 into a temporally extended setting grounded in the framework of Markov Games [Shapley, 1953a, Littman, 1994] and the psychological IMOI model of team performance [Ilgen et al., 2005]. This formulation accommodates the temporal contingencies of real collaboration—how individuals pace their effort, anticipate others' behaviour, and react to partial or delayed feedback across an episode.

Chapter Contributions. This chapter advances the formal modelling in multi-agent teams of psychological theories of temporally extended teamwork through the following contributions:

(C1) **Formalisation of Sequential Teamwork Dynamics:** It introduces *TEMPO*, a Markov Game environment for studying cooperative tasks under temporal pressure, heterogeneous team composition, and partial observability.

(C2) **Unified parameterisation of Task Interdependence:** It provides a single, tunable formalism capable of expressing additive, conjunctive, and disjunctive tasks, enabling systematic comparisons across qualitatively different teamwork settings.

(C3) **Empirical Identification of Sequential Social Dilemmas:** Using empirical game-theoretic analysis, the chapter identifies when and how the induced Markov Game satisfies the criteria for a Sequential Social Dilemma (SSD), extending SSD analysis beyond classical repeated-matrix settings.

(C4) **Psychologically Grounded Environment Design:** The framework incorporates constructs from theories of group performance and team effectiveness, embedding interdependence, resource asymmetry, and evaluative feedback directly into the game dynamics.

(C5) **Analysis of Expertise-Effort Relationships:** Through agents trained via independent Proximal Policy Optimization, the chapter characterises how emergent contribution patterns depend on average expertise and task structure, revealing systematic interactions between heterogeneity and interdependence.

(C6) **Scalable Foundation for Social Simulation:** The environment is designed to support further operationalisations of decision-making and group behaviour, offering a flexible and extensible platform for future empirical and theoretical work.

Taken together, these contributions frame teamwork from a perspective that complements much of the existing MARL and MAS literature on coordination. Indeed, the chapter treats MARL as a computational framework for studying the structure and dynamics of teamwork itself, grounded in theories from organisational and social psychology.

This serves two complementary goals. First, it provides a controlled setting in which psychologically grounded theories of teamwork can be operationalised, formalised, and systematically evaluated. Second, it uses artificial agents as generative instruments for exploring emergent behavioural patterns, organisational failure modes, and hypotheses about cooperation in both human and hybrid teams.

The contribution is therefore bidirectional: psychological theories help structure the design of Sequential Teamwork Games, while the resulting multi-agent environments provide a computational setting in which the behavioural consequences of those theories can be explored, stress-tested, and refined.

The following section instantiates the kinds of situations this framework is designed to model with an illustrative vignette.

4.1 Introduction: Anna's Choice

It's been five years since Anna, our student from Chapter 3, graduated with honours. She is now a computer engineer at a tech company and has just been assigned to a new development team. Although the team is newly formed, she is a senior in the company and has previously worked with most of its members. Their task is clear: to complete a new functionality for the

company's flagship app in time for a high-stakes demo. As their boss explained, this event could open the door for the company to go public, a once-in-a-lifetime opportunity.

There are three intense months of work ahead, but Anna is ready to sacrifice her free time if necessary. As a minor shareholder, she stands to gain significantly from a public offering. She is not alone in her commitment: two other team members are also shareholders, and Anna knows from experience that they are reliable and highly motivated.

Still, one thing concerns her. In a last-minute decision to cut costs, the department head has added a junior developer to the team. From what Anna has gathered, he is technically competent but lacks experience. Worse, he was recently denied a raise, a decision that might weigh on his motivation. He will be responsible for the testing module, a critical component without which the entire functionality cannot be validated or released. No matter how well the other modules are built, if testing fails or falls behind, the demo cannot move forward.

As Anna contemplates the project's demands, a troubling question arises: *What if the junior developer is not willing to go the extra mile?* Given his treatment by the company, such a reaction would be understandable. Yet, this prompts a deeper, rational consideration: *How much effort should she herself expend?* Not out of fairness or loyalty, just out of reason.

A static model, as that in Chapter 3, would suggest that Anna should match the effort level of the least committed teammate. However, that chapter treated teamwork as a one-time decision. Seen that way, it is impossible to capture how Anna and her teammates react to the team workflow. In practice, engagement and effort levels evolve, shaped by prior contributions, perceptions of commitment, ongoing interactions, and expectations about future performance [Nijstad, 2009]. While one might consider extending the previous model via repeated *Teamwork Games* to simulate temporal dynamics, this approach remains limited. Despite being valuable for modelling phenomena such as trust or retaliation in iterated one-shot interactions, repeated games do not embody genuine environmental evolution; they replicate the same strategic interaction, incapable of capturing the environmental evolution and interdependent decision-making that characterise real-world teamwork.

A sequential model instead would allow for dynamic adaptation, where each action could influence not just immediate outcomes but the future state of the task, the team, and the individual's incentives.

4.2 *Teamwork Across Time: Dynamics and Determinants of Group Performance*

To build a more comprehensive model of temporal cooperation in teams, we must look beyond abstract game structures and toward the empirical insights of social psychology, which offer guidance on how coordination, pacing, and interdependence unfold over time. Chapter 3 examined Steiner's task taxonomy and used his theoretical framework as a foundation for designing and analysing *Teamwork Games*. The resulting insights supported a structural model of one-shot strategic contributions to shared tasks—an essential first step, but insufficient for capturing how teamwork evolves in temporally extended settings.

This section extends the review of group research by exploring more recent theories that emphasise the *dynamic* nature of teamwork and the increasing relevance given to contextual variables external to the group. These theoretical advancements seek to integrate such dynamism with the foundational concepts that sustain our understanding of team performance.

Of particular relevance is Ilgen's conceptualisation of teams as dynamic systems that evolve over time [Ilgen et al., 2005]. In this view, outputs and processes at one time point (t) can become inputs at a subsequent point ($t + 1$). This temporal dynamism is encapsulated in the Input-Mediator-Output-Input (IMOI) model of team performance [Ilgen et al., 2005], which underpins the multi-agent simulation framework that will be presented in Section 4.3: Just like Teamwork Games formalised the main psychological primitives of Steiner's theory of group performance, the IMOI model of team performance will inform Sequential Teamwork Games.

As Chapter 3, this one will focus on additive, conjunctive, and disjunctive **unitary** tasks—those that cannot be subdivided into smaller, independent components.

Throughout this chapter, group performance[1] or effectiveness will be assessed through the lens of **productivity**, as it bears a direct mapping to the multi-agent model to be introduced.

[1] An important remark to make at this point is the exact meaning of team *performance* (also called team *effectiveness*) in this context. A review of research on groups and teams in organisational settings by Guzzo and Dickson [1996] highlights a lack of consensus in the field regarding what it means. In response, the authors advocate a broad definition of the term, pointing out that team performance can be indicated by **i**) the group-produced outputs, **ii**) the consequences that the group has for its members, and **iii**) the enhancement of the team's ability to perform effectively in the future. In a similar vein, Campion et al. [1993] acknowledge that group effectiveness is defined both in terms of productivity and affective outcomes such as employee satisfaction.

4.2.1 *The IMOI Model of Team Performance*

Ilgen et al. [2005] observe a clear shift in team research from static views of performance to perspectives that treat teams as dynamic, adaptive systems. This shift redirected attention from fixed input-process-output (Input-Process-Output (I-P-O)) structures toward models that incorporate feedback and temporal change, most notably the Input-Mediator-Output-Input (IMOI) framework.

In IMOI models, team effectiveness is shaped by two types of mediators: **team processes** and **emergent states**. Emergent states capture dynamic properties such as trust or cohesion that arise from interaction and subsequently influence behaviour. Figure 4.1 illustrates the cyclical relationships between inputs, processes, emergent states, and performance that define the IMOI perspective.

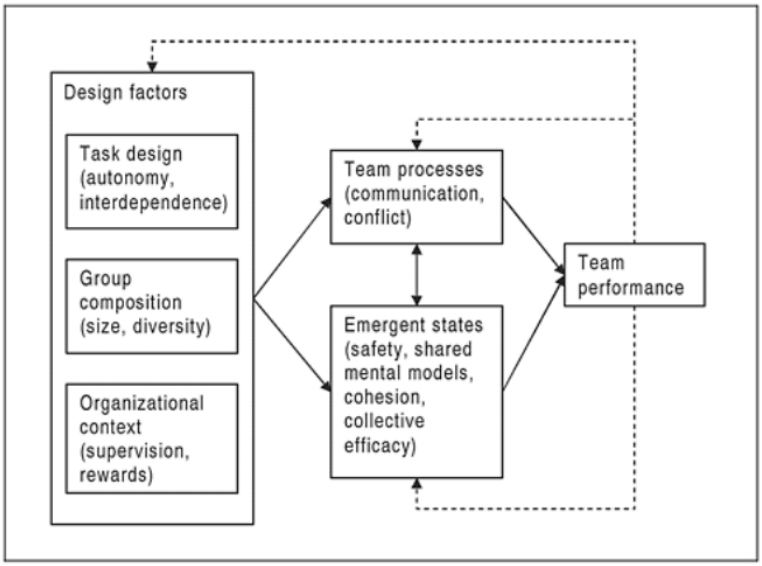


Figure 4.1: **The Input-Mediator-Output-Input model of team performance [Ilgen et al., 2005]**. Design factors like task type, group composition and incentive schemes are inputs to teamwork. This gives rise to team processes and emergent states. As a result of these mediators, the team produces an output at each timestep, which in turn is fed back into the team as input in the next turn. Adapted from Nijstad [2009].

Rico et al. [2011] summarise the main components of the IMOI model and their influence on team effectiveness:

- **Inputs** include both external resources (*e.g.*, organisational rewards) and internal characteristics such as team composition, member skills, and task structure. External reward structures matter: shared rewards, for instance, tend to increase work-group confidence and performance [Hyatt and Ruddy, 1997]. As established in Chapter 3, tasks embed specific patterns of **interdependence**. Within the IMOI framework, these patterns function as inputs shaping team dynamics: *task interdependence* determines how much members must rely on one another during execution, while *goal* (or outcome) *interdependence* captures the degree to which their incentives are aligned. Both forms are consistently associated with increases in team effectiveness [Rico et al., 2011].
- **Mediators—team processes** and **emergent states**—capture the psychosocial mechanisms through which inputs translate into performance. Emergent states are dynamic properties arising from interaction that subsequently shape future behaviour. For example, communication is a team process, whereas the trust it builds (or erodes) constitutes an emergent state. While many mediators have been identified in the literature, this chapter focuses on those that can be meaningfully represented within the multi-agent framework of Section 4.3. Section 4.5.3 analyses *implicit coordination* (a process) and *transactive memory systems* (an emergent state).
- **Team outcomes**—including objective performance, member satisfaction, and innovation—may feed back into the system as inputs in subsequent periods, thereby influencing future interaction patterns (Figure 4.1). In this chapter, the focus is on productivity-related outcomes, as these map most directly onto the model developed below.

Together, these insights highlight the need for models that capture the temporal interplay of team inputs, processes, emergent states, and outcomes. The next section introduces a multi-agent framework that operationalises this perspective.

4.3 *Sequential Teamwork Games*

The next step is the development of a multi-agent reinforcement learning environment in which agents interact, adapt, and learn within a temporal extension of the setting developed in Chapter 3. Rather than replicating specific organisational contexts, the framework abstracts key constructs from the group performance literature—including **task and goal interdepen-**

dence, **effort allocation**, and **team resources**—and embeds them in a computationally tractable Markov Game formulation.

The section first situates the framework within related work, then incrementally outlines the methodological approach, covering agent architecture, environment design, and core parameters. Section 4.5 presents a series of experiments illustrating the dynamics that emerge in teams of four agents under varying reward structures, task types, and agent abilities. It also provides an empirical game-theoretic analysis that identifies the conditions under which the proposed Markov Game satisfies the definition of a Sequential Social Dilemma (SSD).

4.3.1 Context and Related Work

Early Models. Temporally extended models of teamwork are not new to Multi-Agent Systems (MAS). This line of research built upon foundational efforts to engineer coordination through system design, most notably the *Agent-Oriented Programming* (AOP) framework proposed by Shoham [1993]. By defining agents through mental categories such as beliefs and commitments, Shoham established a logic-based approach to distributed computation that treated social interaction as a programmable feature [Shoham, 1993, Shoham and Tennenholtz, 1995]. Building on these foundations, Tambe [1997] proposed a dynamic framework for supporting task achievement in rule-based multi-agent settings. Because their goal was to build effective teams for real-world applications, agents were cooperative by design. The resulting architecture, *STEAM*, drew on two influential theories of teamwork—joint intentions [Levesque et al., 1990, Cohen and Levesque, 1991a] and the *SharedPlan* framework [Grosz and Sidner, 1990, Grosz and Kraus, 1996, Grosz, 1996]—in order to coordinate agents' contributions over time.

A related line of work is that in Dunin-Keplicz and Verbrugge [2010], which developed multi-agent frameworks intended to cover the full life-cycle of teamwork, from team formation and task evolution to communication and eventual dissolution. Their use of BDI (Beliefs, Desires, Intentions) agents allowed for flexible behaviour and, at least in principle, some degree of uncooperative action. However, the settings they considered still relied on the assumption of shared motivational attitudes, which effectively guaranteed cooperation.

In both Tambe [1997] and Dunin-Keplicz and Verbrugge [2010], then, team failure stemmed primarily from communication or coordination breakdown—not from strategic incentives. By

contrast, the approach developed in this chapter builds on the descriptive (rather than normative or prescriptive [Shoham et al., 2007]) perspective of Chapter 3, and extends it to the temporal domain in order to model *unenforced* teamwork. In this setting, the state of the public good (*i.e.*, the task) **changes over time** as a function of teammates' actions, the type of task, and the structural interdependence that links them. Under these conditions, deception, unequal contributions, or even team failure are not model pathologies but natural outcomes of a mixed-motive scenario.

Multi-Agent Reinforcement Learning. Markov Games (Chapter 2) model how environment states and rewards evolve as a function of the agents' joint actions, which makes them a natural choice for studying dynamic multi-agent interaction. A substantial body of recent MARL work uses this formalism to explore emergent social behaviour [Albrecht et al., 2024].

Two influential examples are Perolat et al. [2017] and Leibo et al. [2017]. The first uses independent reinforcement learning agents to recreate the spatial and temporal dynamics of a Common-Pool Resource (CPR) appropriation dilemma. Beyond introducing metrics for analysing emergent behaviour, the authors show that agents eventually learn exclusionary strategies that stabilise the population below the environment's carrying capacity, in line with economic predictions.

Leibo et al. [2017] build on this approach with two new sequential environments, *Gathering* and *Wolfpack*. Their contributions include a formal definition of Sequential Social Dilemmas (SSDs), a method for mapping sequential interactions to matrix-game representations, and a demonstration that static analyses can be misleading when temporal structure is ignored. One of their key observations is that cooperation and defection emerge as *continuous* features of learned policies, shaped by the spatial and temporal distribution of rewards. More recent work extends this line of research by quantifying self-interest in Markov Social Dilemmas and analysing when cooperation can emerge [Willis et al., 2025].

Another stream of work explicitly incorporates social preferences or theory of mind into MARL. Hughes et al. [2018] adapt the inequity-aversion model of Fehr and Schmidt [1999] to decentralised, temporally extended settings and define SSDs for more than two agents. Jaques et al. [2018, 2019] and Ndousse et al. [2021] draw on theories of social learning [Bandura and Walters, 1977] to model how agents learn from and influence each other. In a related vein, McKee et al. [2020] draw on interdependence theory from social psychology and implement

preferences over group outcome distributions using the construct of Social Value Orientation (SVO) [Murphy et al., 2023].

Taken together, these works treat reinforcement learning agents not only as engineering tools but also as models of socially situated decision-makers, opening the door to productive exchanges between computational modelling and the behavioural sciences.

It is in this interdisciplinary context that *Sequential Teamwork Games* are developed. The framework extends the structural logic of Teamwork Games into a dynamic setting, embedding sequential decision-making, interdependence, and heterogeneous abilities within a Markov Game formulation. Unlike spatial environments such as *Gathering* and *Wolfpack*, which revolve around navigation and spatial competition, Sequential Teamwork Games focus on effort allocation and CES-based interdependence, offering a closer conceptual match to organisational teamwork.

Before introducing the formal model, the next subsection provides an intuitive motivation for Sequential Teamwork Games.

Modelling Sequential Teamwork: The Intuition

To motivate the key elements of a *Sequential Teamwork Game*, let us briefly revisit the opening scenario of this chapter. Anna and her teammates must contribute to a shared project under a deadline, and their progress depends not only on their own decisions but also on one another's contributions. In practice, this means that the "state" of the task changes over time as work accumulates or stalls.

Three features of this situation stand out. First, teamwork is inherently *dynamic*: decisions made at one moment influence what teammates can or should do later. This temporal aspect cannot be captured by a repeated one-shot interaction, since the task itself evolves and constrains future behaviour.

Second, team members are mutually dependent. As in Chapter 3, a CES production function provides a simple way to represent different patterns of structural interdependence—from highly substitutable to strongly bottlenecked tasks. In a dynamic setting, these patterns interact with the timing of contributions, making coordination more complex than in the static case.

Third, team members differ in both motivation and ability. In Anna's team, some agents have clear incentives to put in more effort, and others differ in their technical proficiency. These asymmetries matter because they shape not only momentary decisions but also the roles that

may emerge over time. As discussed in Section 4.2, such asymmetries relate to well-established psychological constructs like expertise, coordination, and transactive memory.

When these three elements—temporality, interdependence, and heterogeneity—are combined, a rich range of behavioural patterns can arise. For example, agents may alternate their effort, specialise, compensate for the weakest contributor, or gradually withdraw. Some of these dynamics resemble those observed in the static model from Chapter 3, but the temporal extension introduces new possibilities: roles may stabilise, coordination routines may form, and asymmetric contribution patterns may persist.

To study these dynamics, this chapter introduces *TEamwork under Multi-agent Policy Optimisation (TEMPO)*, a multi-agent reinforcement learning framework designed to model teamwork that unfolds over multiple timesteps. The framework serves a dual purpose. On the one hand, it extends standard MARL environments with psychologically grounded representations of teamwork, including interdependence, heterogeneous expertise, and temporally evolving coordination demands. On the other hand, it provides a controlled computational setting for studying teamwork itself: how different organisational structures give rise to cooperation, free-riding, role differentiation, implicit coordination, and other emergent behavioural patterns. The aim is therefore not to predict individual trajectories, but to analyse systematic behavioural regularities across runs as a function of task structure and team composition.

Building on this intuitive motivation, the next subsection formalises Sequential Teamwork Games by introducing the underlying Markov Game.

4.3.2 *Formal Model*

Despite growing interest in modelling social preferences and temporally extended dilemmas, most existing MARL environments remain constrained by symmetry assumptions and by a heavy reliance on physically grounded domains. Similarly, prior efforts in MAS to model teamwork often assume homogeneous agents, rigid cooperation structures, and fully observable environments, reducing the richness and realism of multi-agent interaction [Andrejczuk et al., 2016]. In response to these limitations, an abstract, decision-focused environment that allows for the endogenous emergence and maintenance of cooperation or defection during teamwork is introduced. Modelling effort allocation as an unembodied decision-making problem not only enables a more realistic exploration of social dynamics but also allows the framework to extend beyond teamwork to a wide variety of social dilemmas and interaction topologies.

To make the presentation clear, the model is introduced in stages. The previous subsection outlined the intuition that motivates Sequential Teamwork Games. The remainder of this section focuses on the formal elements of the Markov Game (Section 4.3.2) and on how these components are instantiated in *TEMPO* (Section 4.4). The learning setup is described in Section 4.4.2.

Sequential Teamwork Games: Formal Components

The goal of this subsection is to formalise the core structural and temporal components that underpin *Sequential Teamwork Games*. These elements constitute the building blocks from which the full Markov Game will be instantiated in Section 4.4. Rather than specifying the environment itself, the focus here is on defining the quantities—gifts, task dynamics, interdependence structure, and evaluation criteria—that shape strategic interaction over time.

The following components formalise how a collective task is progressively completed through temporally sustained group effort. To support the study of realistic team dynamics—where agents continuously adapt to one another and to the evolving demands of the task—the framework incorporates heterogeneous team composition, CES-based task interdependence, mixed-motive incentives, and partial observability. The focus is not on explicit prosocial constructs (*e.g.*, Social Value Orientation (SVO) [McKee et al., 2020]), but on how cooperative behaviour *emerges, stabilises, or decays* from the environment's structural and, critically, temporal properties.

The collective task is completed across an episode of H turns by a set of agents $\mathcal{I} = \{1, \ldots, n\}$. As in Teamwork Games (Chapter 3), agency lies in the time allocated to the task. At each turn t, player i selects the **percentage of time** devoted to teamwork, $a_i^t \in \mathcal{A}_i \subseteq [0, 1]$. Actions are taken in parallel and yield a joint action vector $\boldsymbol{a}^t \in \mathcal{A} = \prod_{i=1}^{n} \mathcal{A}_i$. Each agent's *expertise* represents their ability—a **conversion factor of effort into productivity**—denoted by $p_i \in [0, 1]$, which remains constant throughout the episode.

The effective contribution or *gift* of agent i at time t is then:

$$g_i^t = a_i^t \, p_i \, \Delta \tag{4.1}$$

and the vector of simultaneous gifts is denoted by $\boldsymbol{g}^t$.

Task type. At each timestep, individual gifts are combined through a CES production function, yielding the instantaneous *group product*. As in Chapter 3, different values of the elasticity parameter ρ recover Steiner's unitary task types:

- $\rho \to -\infty$: Conjunctive tasks (progress determined by the minimum gift),
- $\rho = 1$: Additive tasks (arithmetic mean),
- $\rho \to +\infty$: Disjunctive tasks (maximum gift).

The *task state* evolves deterministically as:

$$G^{t+1} = G^t + \left(\sum_{i=1}^{n} \beta_i \, (g_i^t)^\rho \right)^{1/\rho}, \qquad t = 0, \ldots, H-1, \qquad G^0 \stackrel{\text{def}}{=} 0. \tag{4.2}$$

Notation: To distinguish between momentary and accumulated quantities, we refer to the CES aggregation term as the *group product*, and to G^t as the *task state*.

Role weights. In contrast to Teamwork Games, role weights are normalised in the sequential setting, $\sum_i \beta_i = 1$. This reflects a shift in analytical focus: while Chapter 3 examined *absolute* group output under structural heterogeneity, the present chapter investigates *relative* effort patterns as the task unfolds. Normalisation ensures that the task state G^t operates as a consistent, type-independent indicator of collective progress rather than as a literal productivity total. Accordingly, β_i is set to $\frac{1}{n}$ throughout, yielding a Hölder generalised mean over the vector of gifts and enabling direct comparison of effort dynamics across task types. As in Teamwork Games, intermediate values of ρ provide a continuous spectrum of task interdependence.

Figure 4.2 illustrates the evolution of the task state as successive gifts are aggregated.

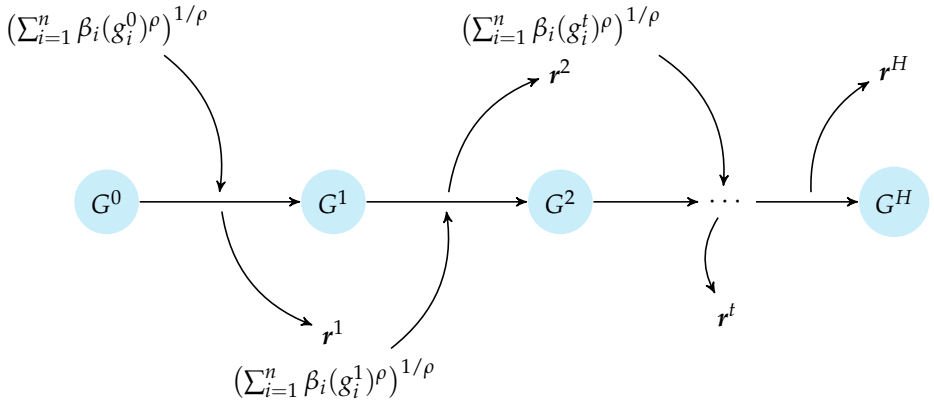


Figure 4.2: **Evolution of the task state over time, with rewards.** At time t, agents act ($\boldsymbol{a}^t$), the state updates to G^{t+1} (top), and the joint reward $\boldsymbol{r}^{t+1}$ is delivered (bottom).

Episodic tasks. With these elements in place, an episodic task is defined as follows:

Definition 4.1 (Episodic Task). Given a set of agents $\mathcal{I}$ that must jointly complete a shared task across an episode of $H \geq 1$ timesteps, an episodic task $\mathcal{T}$ is a tuple:

$$\mathcal{T} \stackrel{\text{def}}{=} (G^{\mathcal{T}}, \Delta, \Theta^{\mathcal{T}}, H), \tag{4.3}$$

where $(G^{\mathcal{T}}, \Delta, \Theta^{\mathcal{T}})$ encodes the same parameters that defined a *unitary team task* in Chapter 3: $G^{\mathcal{T}}$ is the passing threshold, Δ the timestep duration, and $\Theta^{\mathcal{T}} = (\rho, \beta_1, \ldots, \beta_n)$ the structural properties of the task.

The task is completed if and only if the accumulated task state satisfies:

$$G^H \geq G^{\mathcal{T}}. \tag{4.4}$$

Rewards. Agents derive utility from two goods: **i)** their personal leisure time, and **ii)** the outcome of the final task evaluation. Individual differences in how agents trade off these goods are represented by a preference exponent $\alpha_i \in \mathbb{R}^+$, capturing the subjective utility of leisure. Leisure corresponds to the fraction of a timestep not devoted to teamwork:

$$x_i^t = (1 - a_i^t)\Delta,$$

with utility $(x_i^t)^{\alpha_i}$.

The final evaluation $\sigma(\cdot)$ is a logistic function of G^H:

$$\sigma(G^H) = \frac{d}{1 + \mathrm{e}^{-\kappa(G^H - G^{\mathcal{T}})}}, \tag{4.5}$$

which maps task states into evaluations within the interval $[0, d]$. The steepness parameter κ controls how quickly evaluation rises with G^H, and the parameter $G^{\mathcal{T}}$ represents the passing point: when $G^H = G^{\mathcal{T}}$, the evaluation equals $d/2$. This smooth parameterisation reflects that small deviations around the threshold may still yield partial credit, consistent with many real-world evaluation systems.

Consequently, reward for agent i at time $t+1$ is:

$$r_i(x_i^t, G^{t+1}) = \begin{cases} (x_i^t)^{\alpha_i} & \text{if } t < H-1, \\ (x_i^t)^{\alpha_i} + \sigma(G^{t+1}) & \text{if } t = H-1. \end{cases} \tag{4.6}$$

In the next section, these components are embedded within the formal structure of Markov Games, providing the foundation for a fully specified MARL environment capable of modelling sequential team dynamics. Certain features—including agent communication, evolving

preferences, and adaptive expertise—are deliberately omitted in this first formulation, allowing for a focused investigation of structural interdependence, partial observability, and delayed evaluation as drivers of emergent effort dynamics. These simplifications serve as a principled baseline for future extensions incorporating richer social mechanisms.

4.4 TEMPO: A Markov Game Environment for Sequential Teamwork

Building on the formal components introduced in Section 4.3.2, this section presents ***TEamwork under Multi-agent Policy Optimisation (TEMPO)***—a multi-agent environment that operationalises Sequential Teamwork Games within a fully specified Markov Game. In the same spirit as *CollabQuest* in Chapter 3, *TEMPO* provides a concrete instantiation of the theoretical framework, enabling the systematic study of cooperative behaviour under well-defined structural and motivational conditions.

Where *CollabQuest* captured one-shot teamwork through a multi-agent multi-armed bandit, *TEMPO* extends this modelling approach to the sequential domain. It does so by incorporating three core features highlighted in the group performance literature: **i)** temporally unfolding task production based on Steiner's typology, **ii)** heterogeneous agent abilities and preferences under decentralised learning, and **iii)** a dual-motivational reward structure that balances immediate personal benefit with delayed collective evaluation. Together, these elements yield a flexible yet analytically tractable environment for examining how cooperation emerges, stabilises, and degrades over time.

The remainder of this section specifies *TEMPO* formally. It describes its state space, observation and action structures, transition dynamics, and reward formulation, following the notation established earlier. This lays the foundation for the experimental analyses in Section 4.5, investigating how task type, interdependence, and team composition shape dynamic effort allocation in sequential teamwork.

4.4.1 Markov Game.

Let $\mathcal{M} = \langle \mathcal{S}, T, \mathcal{A}, R \rangle$ denote a Markov Game in which n agents are trained simultaneously and independently in a temporally extended multi-agent environment of teamwork.

A core feature of the environment is partial observability. In real-world group work, individuals rarely have full access to teammates' intentions, motivations, or abilities. Even in face-to-face settings, what can be observed is the gift itself, not the underlying effort, willingness, or expertise that produced it. To reflect these realistic constraints of collaboration, agents receive only local information about the evolving task. Specifically, at each timestep t, agent i receives a local observation $o_i^t \in \mathbb{R}^{N+4}$, structured as shown in Table 4.1.

Table 4.1: **Structure of the agent's local observation vector** o_i^t at timestep t.

Component	**Type**	**Description**
G^t	Scalar	Cumulative task progress up to timestep t.
$\delta_t = H - t$	Integer	Remaining timesteps in the episode.
$\delta_G^t = G^{\mathcal{T}} - G^t$	Scalar	Remaining gap to the evaluation threshold.
g_i^{t-1}	Scalar	Agent i's own gift at timestep $t-1$.
$\{g_j^{t-1}\}_{j \neq i}$	Vector	Gifts made by the other $n-1$ agents at timestep $t-1$.
ρ	Scalar (param.)	Task-type parameter governing substitutability in the CES function.

Based on its local observation, agent i must decide how much of the current timestep to allocate to the shared task. Conceptually, this allocation lies in the continuous interval $[0, 1]$. In practice, however, a discretised action space with fixed resolution (step size 0.01) is adopted. This choice reflects both modelling considerations and practical aspects of learning:

1. The behavioural distinctions of interest concern coarse levels of engagement (*e.g.*, low, moderate, high), for which a finely tuned continuous action space is unnecessary.

2. Discrete actions simplify and stabilise training under Proximal Policy Optimisation in multi-agent settings, avoiding the added parametrisation and regularisation required for Gaussian policies.

3. Since the goal of this work is to study emergent cooperation, coordination, and free-riding—rather than marginal productivity effects—discretisation offers sufficient expressive power at substantially lower computational cost.

Accordingly, each agent chooses an action a_i^t from the discretised set $\mathcal{A}_i = \{0.0, 0.01, \ldots, 1.0\} \subset [0, 1]$, representing the proportion of the timestep allocated to the task. All agents act simultaneously, forming the joint action $\boldsymbol{a}^t = (a_1^t, \ldots, a_n^t)$. The environment, which occupies an

unobserved state $s^t \in \mathcal{S}$, then transitions deterministically to $s^{t+1} = T(s^t, \boldsymbol{a}^t)$. A component of this state is the task state, updated according to Eq. (4.2).

$$G^{t+1} = G^t + \left(\sum_{i=1}^{n} \beta_i \, (g_i^t)^\rho \right)^{1/\rho}, \qquad t = 0, \ldots, H-1, \qquad G^0 \stackrel{\text{def}}{=} 0.$$

The agent-level reward is given by Eq. (4.6). Following the convention from Chapter 3 and the notation of Sutton and Barto [2018], realised quantities are written in uppercase, and functions in lowercase. The realised reward is therefore $R_i^{t+1} \stackrel{\text{def}}{=} r_i(x_i^t, G^{t+1})$:

$$R_i^{t+1} \stackrel{\text{def}}{=} \begin{cases} (x_i^t)^{\alpha_i} & \text{if } t < H-1, \\ (x_i^t)^{\alpha_i} \; + \; \sigma(G^{t+1}) & \text{if } t = H-1. \end{cases} \tag{4.7}$$

Rewards are reported in raw units—per-step leisure $x_i \in [0, \Delta]$ and terminal evaluation $\sigma \in [0, d]$—and normalised online during training. Combining leisure and final evaluation allows agents to receive temporally distributed feedback while preserving an incentive aligned with eventual team success. This mirrors real-world settings in which individuals value both immediate personal benefit (*e.g.*, rest) and delayed collective outcomes (*e.g.*, task completion or group recognition). In short, the summative reward balances interpretability, expressiveness, and learnability.

The parameter d is set to $d = 1$, and the passing threshold is chosen from a range $G^{\mathcal{T}} \in (0, \hat{G}_{\max}]$ (Table 4.2), where the upper limit $\hat{G}_{\max}$ is a *population-level benchmark*: the output that a uniform team of *average-skilled* agents ($p_i = 0.6$ for all i) would generate if each member devoted all their time to the task at every timestep, *i.e.*, $a_i^t = 1$ for all (i, t). Formally,

$$\hat{G}_{\max} = \left(\sum_{i=1}^{n} \beta_i \; (0.6 \cdot 1.0 \cdot \Delta)^\rho \right)^{\frac{1}{\rho}} \cdot H, \tag{4.8}$$

where the weights are normalised to $\beta_i = 1/n$. The reference point $p = 0.6$ serves as a normative benchmark rather than an empirical average: human ability distributions in occupational and academic domains are well known to be heavy-tailed [Yang et al., 2024, Bornmann, 2024]. Nonetheless, a fixed external benchmark enables systematic comparison across team compositions and task types.[2]

Figure 4.3 depicts the agent-environment interaction loop.

[2] An alternative would be to define $\hat{G}_{\max}$ relative to each team's composition. While methodologically defensible, this would obscure relationships between expertise and effort allocation, as all teams would face tasks of equal relative difficulty. A fixed benchmark thus provides more informative contrasts.

Table 4.2: **Difficulty levels** expressed as fixed fractions of the benchmark $\hat{G}_{\max}$ (Eq. (4.8)).

Evaluation level	**Passing threshold** $G^{\mathcal{T}}$	**Interpretation**
Easy	$0.3\,\hat{G}_{\max}$	Generally achievable with modest effort
Medium	$0.5\,\hat{G}_{\max}$	Reasonable expectations for an average team
Hard	$0.7\,\hat{G}_{\max}$	Demanding; requires high effort or high expertise

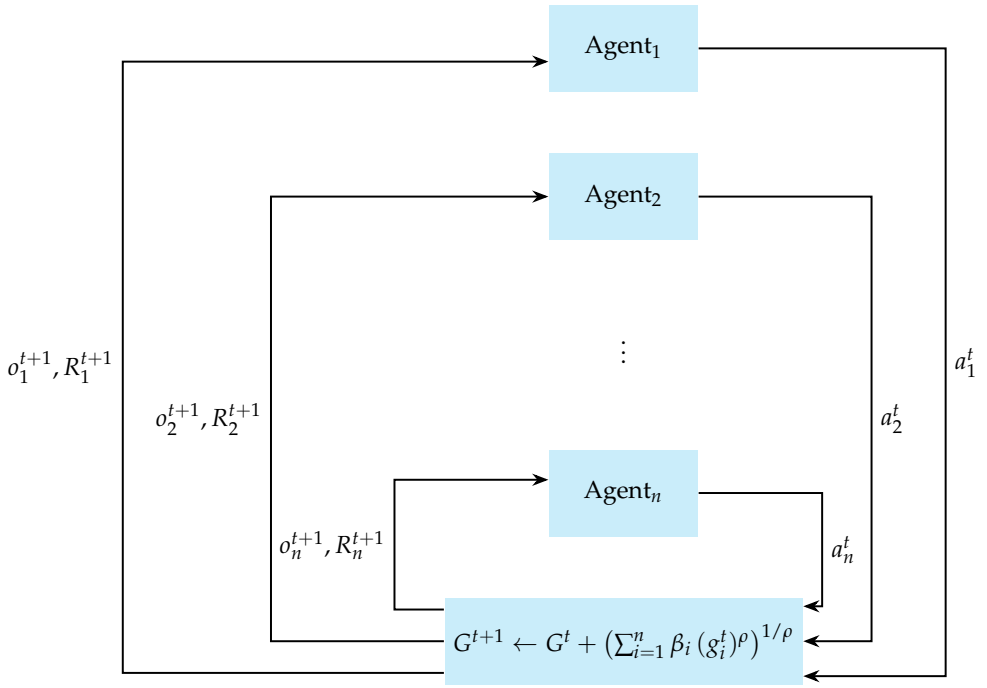


Figure 4.3: **Agent-environment interaction in a Sequential Teamwork Game.** At each timestep t, agents independently select effort allocations a_i^t, which are transformed into gifts and aggregated via a CES function to update the shared task state G^{t+1}. The environment then returns local observations o_i^{t+1} and individual rewards R_i^{t+1}, reflecting both personal leisure and group progress. This architecture supports temporally extended teamwork under partial observability and individualised incentives.

Each agent aims to maximise its expected discounted return:

$$\mathcal{R}_i = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{t=0}^{H-1} \gamma^t R_i^{t+1}\right]. \tag{4.9}$$

In the learning setup described next (Section 4.4.2), each agent's policy π_i is trained using independent Proximal Policy Optimization (iPPO) [Schulman et al., 2017]. The subsequent experimental section builds on this setup to examine the effort dynamics that emerge across task types and team compositions.

4.4.2 *Learning Setup*

This section summarises the training pipeline used to learn decentralised policies in Sequential Teamwork Games. Full implementation details, including environment wrappers, vectorisation, normalisation and logging utilities, are provided in Appendix B.1.

Environment. Training takes place in a synchronous multi-agent environment implemented on top of `PettingZoo`'s `ParallelEnv`. Each episode spans H timesteps, during which n agents repeatedly allocate effort toward the shared task. To accelerate experience collection, multiple environment replicas are executed in parallel, and observations and rewards are normalised online during training. To prevent agents from over-fitting to fixed roles, each episode begins by *randomly permuting* the intrinsic attributes (p, α) across agent identities. Since these variables are not included in the observation space, policies cannot condition on them directly. This shuffling prevents identity-ability coupling during learning and ensures that the resulting policies remain *exchangeable*, a requirement for the empirical game-theoretic analyses that follow. Episodes are collected in alternating phases of rollout and policy optimisation, with advantages computed using Generalised Advantage Estimation (GAE) (see Chapter 2 for further reference).

Agents. Each agent learns independently via independent Proximal Policy Optimization (iPPO) [Schulman et al., 2017], without parameter sharing or centralised training. Policies and value functions are implemented as separate multilayer perceptrons with two hidden layers (128 units, Leaky ReLU activations). The actor outputs a categorical distribution over discretised effort levels, while the critic outputs a scalar state-value estimate.

Parameter updates follow standard PPO-style policy-gradient optimisation. For a batch of M sampled trajectories, the policy-gradient update is:

$$\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}_j + \alpha \frac{1}{M} \sum_{m=1}^{M} \sum_{t=0}^{H-1} \nabla_{\boldsymbol{\theta}_j} \log \pi(a^{t,m} \mid s^{t,m}; \boldsymbol{\theta}_j)\, \hat{A}^{t,m} \tag{4.10}$$

where M is the number of sampled trajectories, t indexes timesteps, H is the episode length and $\hat{A}^{t,m}$ denotes the GAE-based advantage estimate. This update promotes actions that yield higher-than-expected returns relative to the value baseline, allowing agents to improve policies solely from local interaction data. Note that Eq. (4.10) expresses the underlying policy-gradient update common to Proximal Policy Optimization (PPO) , but omits PPO-specific

components such as importance-sampling ratios and clipped objectives (see Chapter 2 for the full derivation).

During evaluation, agents act greedily by selecting the mode of the learned categorical distribution.

4.5 *Experiments*

This section uses *TEMPO* to examine how the structural incentives formalised in Chapter 3 are expressed once teamwork unfolds over time. The central finding is that sequential dynamics do not erase the incentive geometry of Teamwork Games; rather, they filter it through expertise, state visitation, and learned effort allocation. Social dilemmas emerge most clearly in intermediate-skill teams, cooperation switches on once success becomes attainable, and psychologically meaningful mediators such as implicit coordination and role specialisation arise endogenously from decentralised learning.

The experiments are organised around three questions:

- **Experiment 1: When do Sequential Teamwork Games become SSDs?**
  The first experiment shows that SSDs arise primarily when teams have enough expertise for effort to matter, but not so much that success becomes trivial. Empirical game-theoretic analysis is used to estimate Schelling diagrams across task types, expertise levels, and team compositions.

- **Experiment 2: How do task type and team composition shape emergent cooperation?**
  The second experiment shows that cooperation exhibits a switching pattern: agents remain inactive when success is unattainable, then sharply increase effort once the task threshold can be reached. The location and shape of this transition depend on task interdependence, expertise, heterogeneity, leisure preference, and task difficulty.

- **Experiment 3: To what extent do psychologically grounded teamwork mediators emerge in *TEMPO*?**
  The third experiment shows that trained agents display behavioural signatures compatible with implicit coordination and role specialisation. Conjunctive tasks induce the

strongest anticipation and dynamic adjustment, whereas disjunctive tasks induce the clearest temporal division of labour.

General Training Setup

All experiments share a common training pipeline, described below.

Normalisation, Networks, and PPO Setup. Observations and rewards are normalised during training following the approach of Gymnasium [2023]. Each agent uses an actor-critic architecture composed of multilayer perceptrons with configurable depth (`h_depth`) and width (`h_width`); the default configuration employs two hidden layers of 128 units with Leaky ReLU activation.

Agents are trained using Proximal Policy Optimization (PPO) with clipped objectives, entropy regularisation, and GAE-based advantage estimation. Each update performs `n_epochs` **full passes** over the collected batch (no minibatching). Episodes contain $H = 10$ timesteps. Although short compared with ecological MARL environments such as Leibo et al. [2017], this horizon is intentional: episodes in *TEMPO* represent compact teamwork cycles consisting of a small number of interdependent effort-allocation decisions. In this setting, the discount factor $\gamma = 0.9$ plays a comparatively minor role because rewards are dominated by terminal task evaluations.

Hyperparameter Optimisation. Three key hyperparameters are optimised via Bayesian search using Optuna: the actor learning rate (`l_rate`), the number of PPO epochs per update (`n_epochs`), and the number of steps per environment per update (`n_steps`) (Table 4.3). All other hyperparameters remain fixed across experiments (Table 4.4).

Table 4.3: **Optimised hyperparameters** using Optuna.

Parameter	**Value range**
Actor learning rate	$[1 \times 10^{-5}, 3 \times 10^{-3}]$
PPO epochs per update	[3, 10]
Steps per environment per update	[20, 200]

Because some configurations tended to under-explore the action space during early training, hyperparameter optimisation for every setup was performed in two sequential stages:

Table 4.4: **Fixed hyperparameters used across all training runs.**

Parameter	Value
Critic learning rate	$3 \times$ `l_rate`
GAE parameter	$\lambda = 0.95$
Discount factor	$\gamma = 0.9$
Clipping parameter	0.2
Entropy coefficient	10^{-3}
Target KL divergence	0.02
Hidden layer width	128
Hidden layer depth	2
Episode length	$H = 10$

- **Phase 1: Entropy-oriented Optimisation.** Five optimisation trials search over `l_rate`, `n_steps`, and `n_epochs` to maximise initial policy entropy and avoid cold starts.
- **Phase 2: Reward-oriented Optimisation.** Another five trials refine hyperparameters to maximise mean episode return, warm-started from the best entropy-phase parameters. The resulting hyperparameters are then used for training and evaluation in the corresponding configuration.

Reproducibility Note: Random seeds were initialised at the start of every run. Hyperparameter optimisation was performed using Optuna, with each configuration stored in a dedicated SQLite study. When a study already existed, subsequent runs reused the stored hyperparameters and skipped redundant optimisation. Trained parameters were likewise reused when available, while all training logs and evaluation outputs were preserved for post-processing.

Team Configurations and Training Budget

To ensure systematic coverage of composition effects, a fixed set of 38 expertise vectors is used (Appendix B.4): 19 homogeneous teams (low variance, $\sigma_p < 0.145$) and 19 heterogeneous teams (high variance, $\sigma_p > 0.145$). These compositions are reused across all task types and experimental conditions.

Each configuration is trained for `tot_steps` $= 1.5 \times 10^6$ environment steps per seed, using five seeds. With four agents, this corresponds to 6×10^6 agent actions per seed. After training,

each agent is evaluated over 100 rollouts per seed. Training curves are monitored to ensure convergence.

Figure 4.4 shows representative learning dynamics for a heterogeneous team ($p = \{0.2, 0.8, 0.8, 1.0\}$) on an additive task ($\rho = 1$). Returns rise sharply during the first 1-2 $\times 10^5$ steps and stabilise around 1.0-1.2 $\times 10^6$, remaining steady thereafter. Although agents exhibit similar learning dynamics, asymptotic returns systematically reflect individual expertise, supporting the adequacy of the training budget.

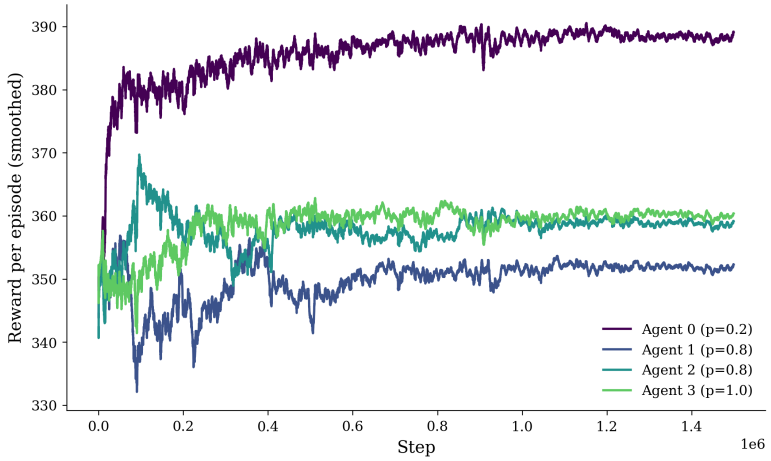


Figure 4.4: **Per-agent training curves** (additive task, $\rho = 1$). Learning curves for each agent in a four-agent team with expertise vector $p = (0.2, 0.8, 0.8, 1.0)$. Each curve reports the moving average of the episodic reward obtained by an agent during training. Rewards are averaged across parallel environments to reduce variance.

All experiments are fully reproducible using a single configuration script.[3]

4.5.1 *Experiment 1: Sequential Teamwork Games and Sequential Social Dilemmas*

This experiment tests when Sequential Teamwork Games constitute Sequential Social Dilemmas (SSDs). The main result is that SSDs concentrate in intermediate-expertise regimes, where effort is impactful but success is not guaranteed. To establish this, we adapt empirical game-theoretic analysis to *TEMPO*.

[3]The experimentation pipeline is automated using Slurm array jobs; post-processing scripts aggregate results and compute behavioural metrics.

Background

To account for spatial and temporal extensions of traditional social dilemmas, Leibo et al. [2017] introduced SSDs. An SSD is a partially observable Markov game $\mathcal{M}$ together with two disjoint sets of policies: Π_C (cooperative) and Π_D (defecting). Let $s_0 \in \mathcal{S}$ denote an initial state, and let $\pi_C \in \Pi_C$, $\pi_D \in \Pi_D$. The long-term expected payoff to agent i under joint policy $\boldsymbol{\pi} = (\pi_1, \pi_2)$ from state s_0 is:

$$v_{i,\boldsymbol{\pi}}(s_0) = \mathbb{E}\left[\sum_{t=0}^{H} \gamma^t r_i(s^t, \boldsymbol{a}^t) \mid \boldsymbol{a}^t \sim \boldsymbol{\pi}, s^t \sim T(s^{t-1}, \boldsymbol{a}^{t-1})\right] \tag{4.11}$$

From these, one can define the empirical payoff matrix as:

$$R(s_0) \stackrel{\text{def}}{=} v_{i,[\pi_C,\pi_C]}(s_0), \tag{4.12}$$

$$P(s_0) \stackrel{\text{def}}{=} v_{i,[\pi_D,\pi_D]}(s_0), \tag{4.13}$$

$$S(s_0) \stackrel{\text{def}}{=} v_{i,[\pi_C,\pi_D]}(s_0), \tag{4.14}$$

$$T(s_0) \stackrel{\text{def}}{=} v_{i,[\pi_D,\pi_C]}(s_0) \tag{4.15}$$

$$\tag{4.16}$$

A Markov game $\mathcal{M}$ qualifies as an SSD if there exists an initial state s_0 for which (R, P, S, T) satisfy the classical social dilemma inequalities (see Chapter 2.1). Specifically, Leibo et al. [2017] require:

(1) Mutual cooperation preferred to mutual defection:	$R > P$
(2) Mutual cooperation preferred to being exploited:	$R > S$
(3) Mutual cooperation preferred to probabilistic unilateralism:	$2R > T + S$
and either	
(4) Temptation to defect:	$T > R$
or	
(5) Fear of being exploited:	$P > S$

Since cooperation in sequential settings is typically graded rather than binary, Π_C and Π_D are often defined by thresholding a continuous behavioural metric (*e.g.*, aggression or contribution level). In practice, Leibo et al. [2017] operationalised this distinction by training two

sets of agents under contrasting environmental conditions: high-resource, low-conflict settings yielding cooperative policies (Π_C), and low-resource, high-conflict settings yielding defecting policies (Π_D). Empirical payoff matrices were estimated by evaluating all four pairings (π_C, π_C), (π_D, π_D), (π_C, π_D) and (π_D, π_C) over multiple episodes, repeating the process until estimates converged.

While this definition applies to two-agent scenarios, many social dilemmas involve larger groups. To address this, Hughes et al. [2018] extended the concept of SSDs to n-agent environments using Schelling diagrams [Schelling, 1973], a compact representation of multi-agent incentives. For a focal agent, let m denote the number of the other $n-1$ agents who cooperate. A Schelling diagram plots the expected payoff for cooperating, $R_C(m)$, and defecting, $R_D(m)$, as functions of m. According to Hughes et al. [2018], an n-agent setting constitutes an SSD if:

(1) Total cooperation is preferred to total defection:	$R_C(n-1) > R_D(0)$
(2) Total cooperation is preferred to being exploited:	$R_C(n-1) > R_C(0)$
Either:	
(3) Fear: Defecting is preferred to being exploited	$R_D(i) > R_C(i)$ for **small** i
or	
(4) Greed: Exploiting is preferred to cooperating	$R_D(i) > R_C(i)$ for **large** i

Estimating these curves follows the same empirical game-theoretic procedure: agents are sampled from Π_C and Π_D to form mixed populations, simulated repeatedly, and average payoffs are computed for each configuration until convergence.

The following subsection adapts and extends Hughes et al. [2018]'s methodology to empirically determine whether Sequential Teamwork Games in *TEMPO* exhibit the incentive structure characteristic of SSDs.

Methods

To balance performance requirements and team resources, evaluation is restricted to tasks of **medium difficulty** (see Table 4.2) and agents trained with a fixed leisure preference exponent of $\alpha = 1.0$. This configuration ensures interpretability of the resulting cooperation dynamics. Extensions to other—and heterogeneous—values of α and evaluation hardness are left for future work.

Three experimental dimensions are systematically varied, as summarised in Table 4.5, yielding $5 \times 4 \times 2 = 40$ unique environment configurations.[4] **All simulations use teams of $n = 4$ agents**, matching the environment's default configuration.

Table 4.5: **Experimental variables used in Experiment 1.**

Factor	**Levels**	**Description**
Task type (ρ)	-2.0 (conjunctive) 0.5 (weakly conjunctive) 1.0 (additive) 1.5 (weakly disjunctive) 2.0 (disjunctive)	Degree of agent substitutability
Average team expertise ($\overline{p}$)	Low $(0.21 - 0.35)$ Medium $(0.36 - 0.55)$ High $(0.56 - 0.8)$ Very high $(0.8 - 1.00)$	Binned by mean agent skill
Team composition	Homogeneous ($\sigma_p < 0.145$) Heterogeneous ($\sigma_p > 0.145$)	Based on expertise dispersion†

†: σ_p denotes the within-team standard deviation of the agents' expertise ($n = 4$ agents per team):

$$\sigma_p = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \overline{p})^2}.$$

Teams with $\sigma_p < 0.145$ were classified as homogeneous, and those with $\sigma_p > 0.145$ as heterogeneous.

Policy Classification. For each environment configuration, trained agent policies are classified as *cooperative* or *defective* based on their average gifting g_i during evaluation. The procedure is as follows:

1. **Filter by configuration.** Attention is restricted to policies trained on tasks of medium difficulty, with $\alpha = 1.0$, $\gamma = 0.9$, episode length $H = 10$, and fixed network architecture (two hidden layers of width 128). Within each experimental condition, policies are further grouped by task type, team composition (homogeneous vs. heterogeneous), and average team expertise $\overline{p}$ (Table 4.5).

[4] Teams with extreme average expertise values reported in Appendix B.4 (close to 0.1 and 1.0) were excluded, as such values can only occur in fully homogeneous teams.

2. **Estimate individual gifting.** For each trained policy and each agent i in the corresponding team:

   (a) The original team of four agents is re-instantiated in the environment.

   (b) The team is evaluated for 100 episodes under greedy action selection (deterministic policies).

   (c) At each timestep, the *gift* of agent i is computed as

   $$g_i^t = a_i^t \, p_i \, \Delta$$

   where a_i^t is the discretised effort allocation, Δ the timestep duration, and p_i the agent's expertise.

   (d) For each episode, gifts are averaged over timesteps, and then averaged across episodes to obtain a scalar summary $\bar{g}_i$ for that agent.

3. **Classify agents by relative contribution.** Within each configuration, the set of $\bar{g}_i$ values is aggregated across all agents from all trained teams. Agents whose $\bar{g}_i$ falls in the bottom 20% of this distribution are labelled *defectors*, whereas those in the top 20% are labelled *cooperators*. Agents in the intermediate range are left unclassified and are not used in the subsequent empirical game-theoretic analysis.

This contribution-based classification ensures that cooperation is defined **relative to peers facing similar structural and motivational conditions** (same task type, difficulty, and approximate expertise level). Without conditioning on these contextual factors, low-skilled agents might be systematically misclassified as defectors (due to low absolute output), and high-skilled agents as cooperators, regardless of their actual effort.

Empirical SSD Calculation. Once cooperative (Π_C) and defective (Π_D) policy sets are obtained, the empirical incentive structure of the environment is estimated using Schelling diagrams. Following the approach of Hughes et al. [2018], teams of exactly 4 agents are sampled and, for a given focal agent, we estimate:

- $R_C(m)$: the payoff to a cooperating focal agent when $m \in \{0, 1, 2, 3\}$ of the other agents cooperate,
- $R_D(m)$: the payoff to a defecting focal agent under the same conditions

For each value of m and for each focal-agent role (cooperator or defector), teams are generated by sampling from Π_C and Π_D and simulate sequential teamwork until the focal agent's estimated return satisfies a confidence-interval-based convergence criterion (see Appendix B.2). Although raw (unscaled) rewards are logged during training, the SSD analysis uses the *normalised* reward signal produced by the environment's `NormalizeReward` wrapper, matching the signal used during training. Because this transformation is monotonic and stable after training, it preserves the qualitative incentive structure (*e.g.*, relative ordering of cooperative and defective payoffs), affecting only the scale of the plotted curves. The resulting payoff curves $R_C(m)$ and $R_D(m)$ are then examined to assess whether each configuration exhibits the incentive structure characteristic of an SSD.

With these procedures in place, the empirical incentive structure of each environment configuration can be reliably estimated. The resulting Schelling diagrams characterise how cooperative and defective policies interact under systematically controlled conditions, providing a basis for assessing whether a given configuration constitutes a sequential social dilemma.

Results and Discussion

Having established the classification procedure and empirical game-theoretic methodology, we present the results obtained across the 40 environment configurations. These analyses reveal how task type, team composition, and average expertise jointly shape the incentive structures in Sequential Teamwork Games.

Throughout, let $m \in \{0, \ldots, n-1\}$ denote the number of *other* cooperators faced by a focal agent (the focal agent's own choice determines whether the cooperative or defective payoff curve applies). To avoid conflating ability with effort, contributions are *stratified by expertise*. Labels therefore reflect *relative behavioural profiles*: a "defector" may still exert non-zero effort, and "full cooperation" does not guarantee task success at low expertise.

By construction, Schelling diagrams are generated by assembling *ad hoc* teams matching a target composition m: for each focal agent, m cooperative and $n-1-m$ defective policies are sampled from the trained cohort. This procedure can produce combinations of high-expertise cooperators that occurred relatively infrequently during training, particularly in **heterogeneous** teams, even though the underlying environment parameters remain identical. This *composition-by-construction effect* is not a shortcoming of the learned policies; rather, it reflects the fact that the evaluation protocol intentionally probes regions of the payoff landscape that

the training distribution visited less often. Consequently, cooperative curves may appear flatter at large m (for example, a weaker increase in $R_C(n-1)$ relative to $R_D(n-1)$) whenever multi-cooperator configurations were under-represented during training.

To illustrate the resulting incentive structure, consider additive tasks performed by heterogeneous teams with medium average expertise (0.36-0.55). Figure 4.5 shows the corresponding Schelling diagram, which satisfies the defining properties of an Sequential Social Dilemma (Section 4.5.1):

1. $R_C(n-1) > R_D(0)$: mutual cooperation yields a higher marginal payoff than universal defection. In the diagram, the right end of the blue curve lies above the left end of the gold curve.

2. $R_C(n-1) > R_C(0)$: cooperating is more attractive when supported by other cooperators than when acting alone.

3. For sufficiently large m, $R_D(m) > R_C(m)$: the gold curve overtakes the blue, indicating a *greed*-driven temptation to exploit cooperating teammates.

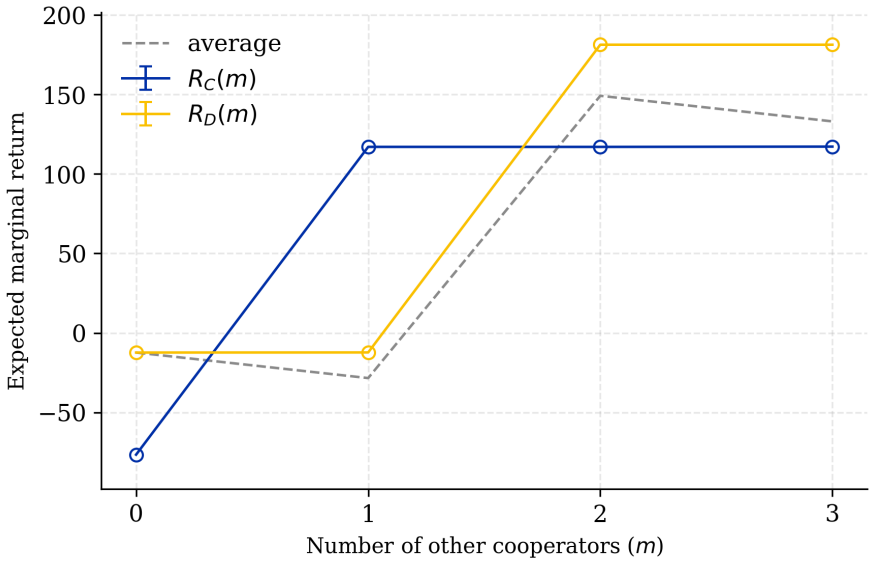


Figure 4.5: **Schelling diagram for medium-difficulty, additive tasks**, performed by **medium-expertise** agents in **heterogeneous** teams. The grey-dotted line shows the team's average payoff from the perspective of a focal defector when m others cooperate.

Thus, for additive medium-difficulty tasks with medium-skilled heterogeneous teams, sequential teamwork **constitutes a social dilemma**. Cooperation is attractive when at least one

teammate contributes, yet as the number of cooperators increases, defection becomes increasingly profitable—producing a classical greed incentive. Moreover, when $m = 0$, the left ends of the curves show that defection strictly dominates cooperation, making universal inaction a (Pareto-inefficient) Nash equilibrium.

The diagram also highlights the absence of a stable fully cooperative equilibrium. When $m = n-1$, a focal agent compares $R_C(n-1)$ and $R_D(n-1)$; since the latter is larger, a rational agent would defect, preventing full cooperation despite its organisational desirability.

The dotted grey line represents the **team's average payoff** when exactly m agents cooperate, assuming the focal agent defects. As shown, average returns peak around $m = 2$, illustrating that the all-defect equilibrium is strategically stable but socially inefficient. A compressed overview of which configurations qualify as SSDs under additive tasks is shown in Figure 4.6. Each point corresponds to an empirical analysis; points are shown in blue when the criteria for an SSD are satisfied and in yellow otherwise. Complete summaries for all remaining configurations are provided in Appendix B.3.[5]

Empirical findings derived from similar diagrams across all task types are summarised in Table 4.6. A closer look at the diagram reveals *when* sequential teamwork constitutes an SSD and *how* the temporal dimension of SSDs underpins the structurally-induced incentives in Teamwork Games.

Table 4.6: **Conditions under which different task types give rise to SSD, along with the dominant incentive**. Only settings that *do* yield an SSD are listed. Superscripts indicate the mapping between diversity and incentive: [1]Homogeneous → Fear, [2]Heterogeneous → Greed.

Task Type	Expertise Level	Diversity	Incentive
Conjunctive	Medium—High	Homogeneous	Fear
Weakly Conjunctive	Medium	Homogeneous	Fear
	High—Very High	Heterogeneous	Greed
Additive	Medium	Homogeneous[1] & Heterogeneous[2]	Fear[1] & Greed[2]
	High—Very High	Homogeneous & Heterogeneous	Greed
Weakly Disjunctive	Medium—Very High	Homogeneous	Greed
	Medium—High	Heterogeneous	Greed
Disjunctive	Medium	Homogeneous & Heterogeneous	Greed

[5]For additional illustrations of how the five task types relate to their emergent cooperation patterns, an animation is available at https://github.com/aloaberasturi/TEMPO.

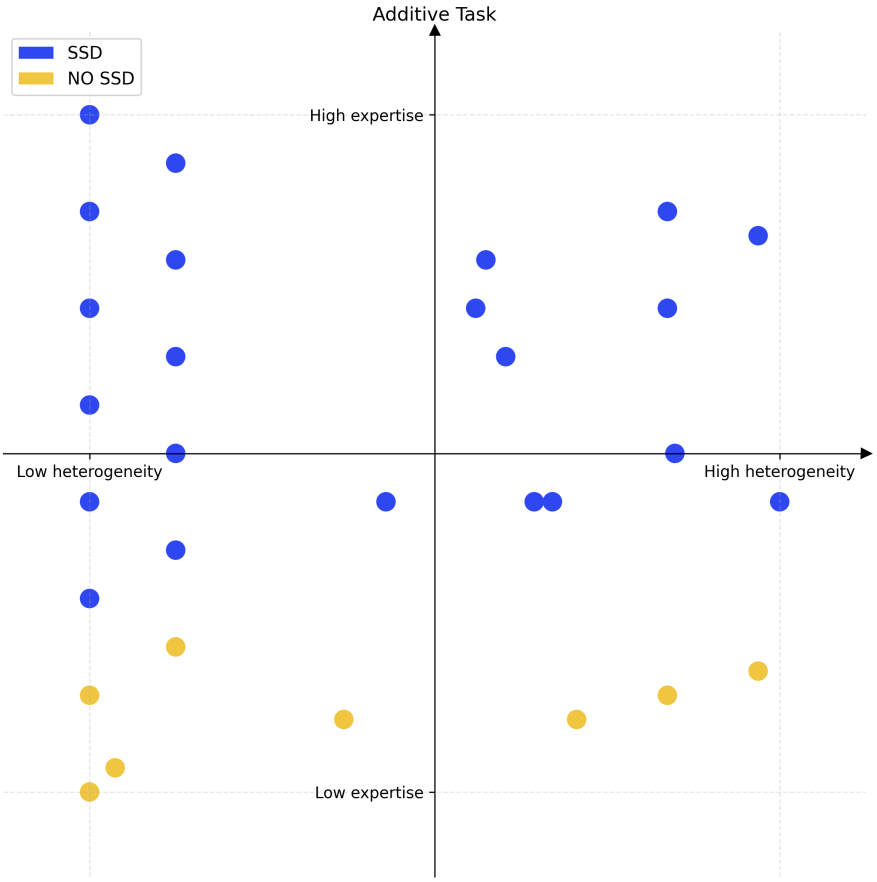


Figure 4.6: **Empirical mapping of Sequential Social Dilemmas (SSDs) under the additive task regime with medium difficulty.** Each point corresponds to a specific team configuration, defined by average expertise (vertical axis) and expertise heterogeneity (horizontal axis). Dark markers indicate configurations where the empirical payoff structure satisfies the conditions for an SSD, as determined by empirical Schelling diagram analysis. Beige markers correspond to configurations that do not exhibit SSD characteristics.

When do Sequential Teamwork Games become SSDs? Taken together, the empirical Schelling diagrams show that the underlying incentive geometry identified in Chapter 3—and importantly in Appendix A.5—is largely preserved in the sequential setting, but whether these incentives become behaviourally relevant depends systematically on team expertise and heterogeneity.

Across task types, three broad regimes emerge. First, **under-powered teams** with low average expertise rarely exhibit a meaningful conflict between individual and collective incentives. Because even sustained high effort has little impact on the task state, unilateral cooperation

produces negligible improvements in team success. In additive and weakly conjunctive tasks, this leads either to approximate indifference between cooperating and defecting, or to a strict preference for defection in homogeneous low-skill teams: leisure is preserved while the probability of success remains essentially unchanged. In conjunctive tasks, where the minimum contribution binds, heterogeneous teams perform particularly poorly: the weakest member becomes a hard bottleneck and even very skilled teammates cannot compensate, so additional cooperation provides little collective benefit. As a result, teams converge towards widespread defection, not because free-riding is especially attractive, but because cooperation is largely ineffective.

Second, **moderately skilled teams** concentrate the conditions under which Sequential Teamwork Games constitute SSDs. Here, effort has a sufficient marginal impact on the task state for structural incentives to meaningfully shape behaviour. Homogeneous conjunctive tasks exhibit clear fear-based dilemmas: cooperation is collectively optimal but individually risky, reproducing a Stag-Hunt structure in the sequential setting. Additive tasks tend to display mixed motives: depending on composition, the Schelling curves reveal either fear (when joint failure is still common) or greed (when free-riding on cooperative teammates becomes attractive). Weakly disjunctive and disjunctive tasks show the cleanest greed incentives. In these regimes, heterogeneity is beneficial for task completion but socially problematic: a few highly skilled agents can carry the team, creating strong opportunities for others to free-ride. The resulting incentive structure closely resembles a Volunteer's Dilemma in sequential form.

Third, in **highly skilled teams**, many of the static tensions are attenuated or disappear. When average expertise is high, the task rapidly enters the high-evaluation region even if agents do not contribute at every timestep. In disjunctive and weakly disjunctive tasks, the difference between cooperating alone and cooperating alongside other high-ability team members becomes negligible, so $R_C(0) \approx R_C(n-1)$ and the greed component effectively vanishes in the empirical Schelling diagrams. In heterogeneous high-skill teams, the individually optimal response is often simply to cooperate: the marginal cost of effort is small compared with the near-certain success of the task. In highly skilled homogeneous teams, by contrast, cooperation and defection tend to become payoff-equivalent: once success is almost guaranteed, leisure and evaluation decouple, and no strong social dilemma remains.

Heterogeneity, therefore, modulates incentives in a task-dependent way. In **conjunctive** tasks, it is primarily detrimental: variance in expertise deepens bottlenecks and pushes teams into regions where defection dominates, eliminating the dilemma. In **disjunctive** and weakly

disjunctive tasks, it is mainly beneficial for completion but amplifies greed: a small number of highly productive agents can secure success for everyone, making free-riding attractive for their teammates at intermediate skill levels. In **additive** tasks, heterogeneity plays a more modest role; what matters most is average expertise, with the same task structure yielding fear-based, greed-based, or no dilemma as the team moves from under-powered to over-qualified.

In short, SSDs emerge most frequently when efforts are close to perfect substitutes. Substitutive production technologies create the structural conditions for greed: the team can compensate for an individual's defection, so free-riding becomes attractive. Complementary technologies, on the other hand, either nullify unilateral impact (eliminating the dilemma) or generate fear-based coordination problems when joint success is attainable.

Social Dilemmas vs. Sequential Social Dilemmas. The one-shot model developed in Chapter 3 identifies the *structural incentive anatomy* induced by each task type. This anatomy is determined by the CES production technology and the evaluation rule, and it specifies the *set of strategic motives* that can arise:

- fear-driven incentives in conjunctive tasks, where unilateral effort is ineffective;
- greed-driven incentives in disjunctive tasks, where free-riding on a volunteer is attractive;
- and mixtures of fear and greed in additive tasks under softened evaluation.

These motives are intrinsic to the task and exist regardless of whether interaction is static or sequential. Sequential Social Dilemmas (SSDs) do not change this underlying anatomy. What they reveal instead is *which components of that incentive structure become strategically relevant* once policies evolve in a temporally extended environment. In a Markov game, agents do not choose a single effort level but a full policy conditioned on the evolving task state. Episodes trace trajectories through regions of low, medium, or high collective progress, and the marginal effectiveness of effort varies sharply across these regions. Different teams spend different proportions of time in distinct incentive regimes, even though the underlying dilemma remains unchanged. Formally, each point in a Schelling diagram,

$$R_C(m), \qquad R_D(m),$$

is the expected *time-aggregated return* to a focal cooperator or defector when exactly m of the remaining agents also cooperate. These values implicitly encode how temporal dynamics

shape incentives. High-expertise or highly cooperative teams drive the task quickly into high-performance regions, spending little time where additional effort has low marginal value. Low-expertise or strongly defecting teams, by contrast, remain trapped in low-progress regions where unilateral cooperation is unproductive. The Schelling curves thus reflect how the structural incentives are *filtered through the state-visitation patterns* induced by the policies being evaluated.

This perspective explains both the convergences and divergences between the static and sequential analyses. In disjunctive tasks, for example, the one-shot model contains a clear greed component: defecting strictly dominates cooperating once someone else reaches the threshold. In the sequential setting, however, highly skilled teams drive the task rapidly into the high-evaluation region, so the practical difference between cooperating and free-riding becomes small. As a result, the empirical $R_C(m)$ curve becomes comparatively flat. The underlying temptation to free-ride remains structurally present, but its impact on realised behaviour weakens because successful teams spend little time in states where additional effort has high marginal value.

Conjunctive tasks with low expertise illustrate the opposite pattern. In the static model, such teams may face indifference between cooperating and defecting: neither agent can raise the minimum above threshold, so both actions yield zero evaluation. In the sequential Markov game, however, defection becomes strictly more attractive, since

$$R_D \approx x_i^{\alpha}, \qquad R_C \approx 0$$

because cooperation repeatedly sacrifices leisure without meaningfully improving the task state. Sequential interaction therefore strengthens the incentive to defect in poorly resourced conjunctive teams: agents remain trapped in low-progress states where effort has negligible payoff.

Overall, Sequential Teamwork Games complement the structural logic of the one-shot model with insights into the *dynamic expression* of social dilemmas. They clarify which components of the underlying incentive anatomy become behaviourally operative for a given team, and how this depends on expertise, heterogeneity, and the cumulative marginal effectiveness of cooperation over time.

Key Takeaways

In summary:

- The underlying dilemma type—fear, greed, or absence of incentives—is determined by the task's CES geometry, but its behavioural expression depends on expertise and heterogeneity. Sequential dynamics only reveal these incentives to the extent that teams visit regions where effort has non-zero marginal impact.
- SSDs emerge primarily in intermediate-ability teams and in tasks with near-substitutable contributions (additive, disjunctive). Complementary tasks either suppress greed entirely or produce fear-based coordination problems at high ability, while low-ability or highly heterogeneous teams face no genuine dilemmas because cooperation is futile.
- High-ability teams tend to neutralise the dilemma: rapid progress either makes free-riding unprofitable or cheapens cooperation. Thus, the sequential setting filters structural incentives through state visitation, determining which motives actually matter.

These insights complete the structural analysis. Experiment 2 now turns to *emergent behaviour*: how teams of different compositions deploy effort across time.

4.5.2 *Experiment 2: Emerging Behaviour in Sequential Teamwork Games*

This experiment analyses the behavioural expression of cooperation. The main result is a threshold-like pattern: teams free-ride when success is unattainable, but switch into active contribution once the task becomes feasible. To characterise this transition, the expertise that team members bring is treated as the environment's **resource**, and its availability is manipulated to study emergent behavioural patterns.

Background

Prior MARL environments, such as *Gathering* [Leibo et al., 2017], provided external and physically grounded resources (*e.g.*, apples) to pose a dilemma. The authors manipulated the apple respawn rate, and the cost of aggression to inhibit and induce mixed-motive settings. In another environment (*Wolfpack*), cooperation was promoted by group-capture synergy (bonus and radius). These settings naturally led to questions such as *How do aggressiveness and coordination in the learned policies respond to the tunable parameters in the environment?*

In turn, *TEMPO* models resource abundance as average team expertise $\overline{p}$, and uses the CES parameter ρ to vary the *technology of joint production* (task type), which is conceptually close to *Wolfpack*'s group-capture synergy dial (see Table 4.7 for reference).

Table 4.7: **Conceptual mapping of environment dials.**

Environment	Resource scarcity	Exclusion	Synergy
Gathering	Respawn time N_{apple}	Timeout N_{tagged}	—
Wolfpack	—	—	$r_{\text{team}}/r_{\text{lone}}$ & capture radius
TEMPO	Avg. expertise $\overline{p}$	—	Task type ρ

Notes. In *Gathering*, N_{apple} controls resource abundance (scarcity) and N_{tagged} the cost of conflict. In *Wolfpack*, cooperation is shaped by the group-capture bonus and capture radius; resource abundance and conflict are not directly manipulated. *TEMPO* endogenises scarcity via expertise $\overline{p}$ and varies technological interdependence with ρ. Conflict is not directly manipulated.

Another insightful question explored in Leibo et al. [2017] concerned how properties *intrinsic to the agents* influenced their learned behaviour. In response, they analysed the moderator effect of the discount parameter γ, the batch size, and the hidden-layer width on **i)** the relationship between aggressiveness and scarcity in *Gathering*, and **ii)** the balance between lone-wolf capture rate and group benefit in *Wolfpack*.

We follow this line of analysis by examining how average gifting responds to team expertise $\overline{p}$, and how this relationship is modulated by agents' preference for leisure (α).

Methods

To systematically analyse how different variables shape cooperative behaviour, we conducted a one-factor-at-a-time grid of experiments varying task type (ρ), team homogeneity, task difficulty (passing threshold), and leisure preference (α). Each experimental block manipulated a single variable while holding the others constant, enabling a clean attribution of behavioural changes to the factor of interest.

Experimental grid. When varying difficulty, the leisure preference was fixed at $\alpha = 1.0$; when varying α, difficulty was fixed at the medium level. Table 4.8 summarises the manipulated and controlled factors.

The full grid comprises 1,140 unique configurations: 5 task types, 2 homogeneity regimes, 2 manipulated variables, 3 levels each, and 19 expertise vectors; each trained with 5 random seeds and using 5 parallel environments. In total, this results in 28,500 independent training processes.[6]

[6]Definitions of difficulty levels are given in Table 4.2.

Table 4.8: **One-factor-at-a-time experimental grid used for Experiment 2.**

Varied factor	Levels	Held constant
Difficulty	{`easy`, `medium`, `hard`}	$\alpha = 1.0$
α	{0.5, 1.0, 1.5}	difficulty = `medium`

Training and evaluation. All models were trained under stochastic sampling with 5 parallel environments. Evaluation switched to deterministic rollouts using `num_envs` $= 1$, running 100 episodes per training seed. This ensured that behavioural summaries reflected learned policies rather than sampling noise.

Metrics. For each configuration, three complementary metrics were recorded:

- **Team success rate:** proportion of evaluation episodes in which the team exceeded the passing threshold $G^{\mathcal{T}}$.
- **Normalised effort:** the proportion of timesteps during which agents selected non-zero contribution levels.[7]
- **Action trajectories:** per-agent action sequences, averaged over time and seeds, including variability bands.

These statistics were aggregated across seeds to produce effort-expertise curves and heatmaps for each manipulated variable and homogeneity regime. These behavioural datasets allow us to trace where cooperation activates, how effort is distributed, and how task structure modulates this transition.

Results and Discussion

This experiment aims to characterise the conditions under which cooperation emerges in Sequential Teamwork Games, and to understand how task structure, expertise, and motivational parameters shape that emergence. Within-episode action patterns are first studied, followed by aggregate cooperation across task types, effort dispersion and the impact of heterogeneity. We conclude with moderator analyses focusing on agents' leisure preference (α) and the evaluation hardness ($G^{\mathcal{T}}$).

[7]Effort is defined as the percentage of time devoted to teamwork. This measure decouples cooperation from expertise, ensuring that it reflects *willingness* to cooperate rather than skill.

Action trajectories. Figure 4.7 illustrates typical action trajectories for heterogeneous four-agent teams performing additive, medium-difficulty tasks. The trajectories provide a qualitative view of how division of labour unfolds in practice. Despite noticeable variability across agents, the team reaches a 100% success rate. This qualitative pattern is broadly consistent with the incentive structure predicted in Chapter 3: in additive tasks, high-expertise agents have the greatest marginal impact on team output, making compensatory effort and asymmetric workload allocation strategically advantageous.The relevance of Figure 4.7 lies in showing that these theoretically expected coordination patterns emerge endogenously from decentralised learning, without explicit coordination protocols or role assignment.

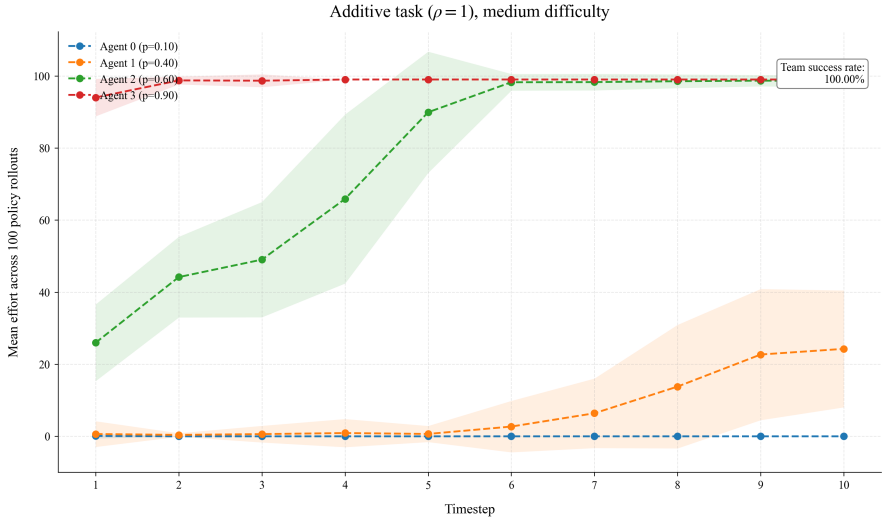


Figure 4.7: **Average action trajectories for four agents performing an additive task ($\rho = 1$) of medium difficulty.** The team is heterogeneous ($\sigma_p > 0.145$).

A clear division of labour emerges. The least skilled member ($p_0 = 0.1$) disengages entirely, contributing no effort with negligible variability. The most capable agent ($p_3 = 0.9$) works almost continuously from the start, effectively anchoring the team's progress. Intermediate agents adopt mixed strategies: one delays effort until late in the episode, while another ramps up gradually. These patterns mirror two well-documented phenomena in (human) substitutable tasks:

- *Social compensation*: high-expertise agents contribute early to ensure task progress [Steiner, 1972, Williams and Karau, 1991].

- *Implicit coordination*: intermediate agents avoid redundant effort and adjust dynamically as others act [Fiore and Salas, 2004, Rico et al., 2008].

Variability bands indicate that high- and low-expertise policies stabilise rapidly, whereas intermediate agents face greater strategic ambiguity. Although identities of early and late contributors may swap across seeds, the qualitative structure persists: agents spontaneously coordinate to meet the team threshold $G^{\mathcal{T}}$.

General patterns of cooperation. To obtain a global view of cooperative behaviour, Figure 4.8 plots average team effort across task types and expertise levels for homogeneous teams at medium difficulty.

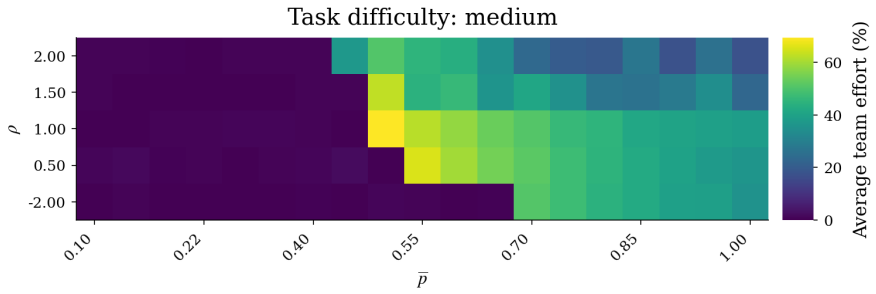


Figure 4.8: **Average team effort (percentage of time) across task type (ρ) and average expertise $\overline{p}$.** Homogeneous teams, medium-difficulty tasks, $\gamma = 0.9$, $\alpha_i = 1$ for all i.

Across all task types, two behavioural regimes appear. At low expertise, agents contribute very little effort: because contributions have negligible marginal impact on success, sustained cooperation becomes strategically ineffective. In these regimes, low effort reflects disengagement or **futility** rather than classical free-riding on successful teammates. This changes abruptly once success becomes attainable. Beyond this activation point, effort increases sharply, revealing a consistent "switching" from inaction to cooperation.

The position of this boundary depends on both ρ and $\overline{p}$. More skilled teams require less effort to succeed, and teams with insufficient expertise remain inactive regardless of task type. Beyond this shared backbone, task types show clear differences: additive tasks elicit the highest overall effort; conjunctive tasks require even engagement once activated; and disjunctive tasks demand the least.

These differences are quantified through pairwise contrasts in Table 4.9. For instance, weakly conjunctive tasks require substantially more effort than weakly disjunctive ones ($\Delta =$

+28.3%, $p < 0.001$), and conjunctive tasks exceed disjunctive tasks by +41.1% ($p < 0.001$). These findings illustrate the central role of ρ as a structural determinant of cooperation.

	$\rho = 2$	$\rho = 1.5$	$\rho = 1$	$\rho = 0.5$	$\rho = -2$
$\rho = 2$	—				
$\rho = 1.5$	+21.9% **	—			
$\rho = 1$	+64.4% ***	+34.9% ***	—		
$\rho = 0.5$	+54.5% ***	+28.3% ***	−4.9% $^{\text{ns}}$	—	
$\rho = -2$	+41.1% ***	+15.7% **	−14.2% ***	−9.8% **	—

Table 4.9: **Pairwise global comparisons of average team effort across task types (restricted to active regimes and homogeneous teams)**. Each entry shows the percentage difference (Δ) of the row relative to the column. Statistical significance is denoted by stars: $^{*}p$-value < 0.05, $^{**}p$-value < 0.01, $^{***}p$-value < 0.001. Non-significant differences are denoted by an "ns" superscript.

Effort dispersion. Task structure also affects how effort is distributed within a team. As shown in Table 4.10, conjunctive tasks exhibit the lowest dispersion: all agents must contribute similarly to avoid bottlenecks. Additive tasks produce moderate dispersion, reflecting mixed strategies combining free-riding and compensatory effort. Disjunctive and weakly disjunctive tasks generate the highest dispersion, with a few high-contributing "volunteers" carrying largely inactive teammates.

	$\rho = 2$	$\rho = 1.5$	$\rho = 1$	$\rho = 0.5$	$\rho = -2$
$\rho = 2$	—				
$\rho = 1.5$	−8.7% **	—			
$\rho = 1$	−65.0% ***	−61.7% ***	—		
$\rho = 0.5$	−83.4% ***	−81.9% ***	−52.6% ***	—	
$\rho = -2$	−88.4% ***	−87.3% **	−66.9% ***	−30% **	—

Table 4.10: **Pairwise global comparisons of team effort dispersion across task types (restricted to active regimes and homogeneous teams)**. Each entry shows the percentage difference (Δ) of the row relative to the column. Statistical significance is denoted by stars: $^{*}p$-value < 0.05, $^{**}p$-value < 0.01, $^{***}p$-value < 0.001. Non-significant differences are denoted by an 'ns" superscript.

Effect of heterogeneity. Introducing heterogeneity ($\sigma_p > 0.145$) produces strong interactions between team composition, task type, and expertise. Three main effects are observed:

- **Conjunctive tasks: collapse of cooperation.** A persistent weakest link makes success unattainable in most settings. Agents learn to coordinate downwards and conserve leisure, resulting in near-zero effort.
- **Disjunctive tasks: stable high-effort volunteers.** A single highly skilled member can often guarantee success, leading that agent to sustain high effort while others free-ride. Effort declines more slowly than in homogeneous teams.
- **Additive tasks: robustness to heterogeneity.** Activation occurs at lower expertise levels than in the homogeneous case, but once active, teams converge to the minimum effort needed for success. This mirrors classic findings that additive tasks tolerate internal variance [Steiner, 1972].

Full heterogeneous results are provided in Appendix B.5.

Moderators of Cooperativeness. To develop a fuller understanding of the forces that inhibit or promote cooperation in Sequential Teamwork Games, the moderator effect of two interpretable dials was studied: **i)** an *agent-level* parameter, the preference for leisure α_i, and **ii)** an *environmental* parameter, the passing threshold $G^{\mathcal{T}}$. Although the analyses below focus primarily on homogeneous teams, all contrasts were replicated for heterogeneous ones. These supplementary runs did not reveal qualitatively new patterns, nor did they complicate the emerging narrative. For this reason, the corresponding observations are integrated into the discussion rather than being devoted to independent subsections. As before, cooperation is defined as the **percentage of timesteps during which agents contribute to the task**, thereby decoupling cooperative willingness from expertise.

For each moderator, differences in cooperation across conditions were evaluated through three complementary analyses:

1. *Global effort shifts:* for an intuitive comparison, effort-expertise curves were aligned on a common domain and their average vertical separation (bin-wise mean difference) was computed.[8]

[8]These global differences should be interpreted descriptively. Because the effective sample size corresponds to the number of active expertise bins (those with non-zero effort), the resulting statistical power is often limited. Consequently, substantial vertical gaps may not achieve significance at the global level. Significance is therefore assessed solely from the per-expertise contrasts (item 3), with global differences serving as a compact indication of effect magnitude and direction.

2. *Activation expertise ($\overline{p}_{\text{act}}$):* Because the curves can shift horizontally, for each condition, the minimum average expertise level at which cooperation consistently exceeded a given threshold was identified. Pairwise differences in $\overline{p}_{\text{act}}$ across seeds then captured horizontal shifts.

3. *Per-expertise contrasts:* Welch's t-tests were conducted independently for each active bin in order to detect local differences in mean effort.

Preference for leisure (α). In the reward function from Eq. (4.7), α_i controls each agent's valuation of leisure. While it shares a conceptual similarity with the discount factor γ—both can increase short-termism—they act through distinct mechanisms: α_i alters the *within-step* utility of effort, whereas γ modifies the *temporal aggregation* of future payoffs. Preliminary runs revealed that α exerts a markedly stronger behavioural influence. Combined with the short episode horizon ($H = 10$), and the fact that α reflects an interpretable motivational preference (as opposed to the more algorithmic character of γ), this made it the natural dial for capturing agents' propensity towards immediate reinforcement.

Figure 4.9 shows how α modulates cooperation as a function of expertise. The results correspond to medium-difficulty tasks, with homogeneous teams, a fixed discount factor ($\gamma = 0.9$), and a uniform leisure preference across agents ($\alpha_i = \alpha \in \{0.5, 1.0, 1.5\}$). To avoid figure clutter, only the three paradigmatic task structures—conjunctive, additive, and disjunctive—are displayed. Analogous plots for heterogeneous teams are provided in Appendix B.6.

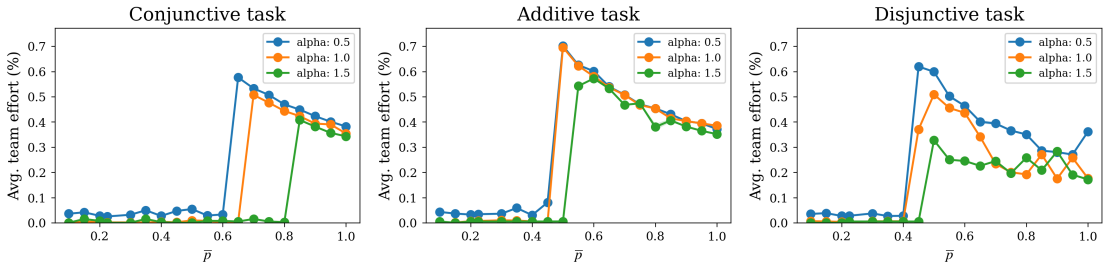


Figure 4.9: **Effect of the preference for leisure (α) on effort-expertise curves** across task types (homogeneous teams, medium difficulty, uniform α across agents).

Across task types, a coherent pattern emerges: higher preference for leisure **systematically depresses cooperation**. This effect manifests in both *vertical* reductions in mean effort and

horizontal delays in the onset of cooperative behaviour. Tables 4.11 and 4.12 summarise these global tendencies; inference is based on local per-expertise contrasts.

1. **Conjunctive tasks.** Conjunctive settings are acutely sensitive to motivational costs. Increasing α leads to substantial declines in effort (up to 11% when comparing $\alpha = 0.5$ to $\alpha = 1.5$) and shifts the activation threshold rightward by as much as 0.22 in expertise. The weakest-link structure magnifies the effect of within-step leisure valuation: as soon as leisure becomes more rewarding, agents reduce effort sharply.

2. **Additive tasks.** The influence of α is milder but highly consistent. Vertical differences remain below 8%, yet per-expertise contrasts are robust. Higher α reliably delays activation by roughly 0.05 in $\overline{p}_{\text{act}}$. Even when contributions are perfect substitutes, a stronger preference for immediate leisure discourages early engagement.

3. **Disjunctive tasks.** Here the moderator pattern is non-uniform. At low and intermediate expertise, higher α substantially reduces effort, with nearly all local contrasts significant. Once the team contains a clearly capable member, however, the disjunctive structure reduces sensitivity to α: an "expert carrier" can sustain performance, compressing the differences across α levels. Horizontal shifts are modest (approximately 0.05 in $\overline{p}_{\text{act}}$). Overall, α mainly depresses cooperation *before* a competent volunteer emerges.

Table 4.11: **Global vertical shifts in average cooperation** (percentage of time on task) across α levels.

Comparison	Conjunctive	Additive	Disjunctive
$\alpha = 0.5$ vs. 1.0	+5.6%	0.0%	+44.0%
$\alpha = 0.5$ vs. 1.5	+11.2%	+7.4%	+65.6%
$\alpha = 1.0$ vs. 1.5	+5.0%	+6.9%	+24.0%

Table 4.12: **Horizontal shifts in activation thresholds** $\Delta\overline{p}_{\text{act}}$ across α levels. Negative values denote earlier activation for the left condition.

Comparison	Conjunctive	Additive	Disjunctive
$\alpha = 0.5$ vs. 1.0	−0.05	−0.05	0.00
$\alpha = 0.5$ vs. 1.5	−0.20	−0.10	−0.05
$\alpha = 1.0$ vs. 1.5	−0.15	−0.05	−0.05

Passing Threshold ($G^{\mathcal{T}}$). Figure 4.10 shows how increasing the passing threshold modulates cooperation across teams of varying expertise. Results correspond to homogeneous teams with $\gamma = 0.9$ and uniform leisure preferences ($\alpha_i = 1$). As before, the three structurally distinct task types are displayed here. Panels for heterogeneous teams are deferred to Appendix B.6.

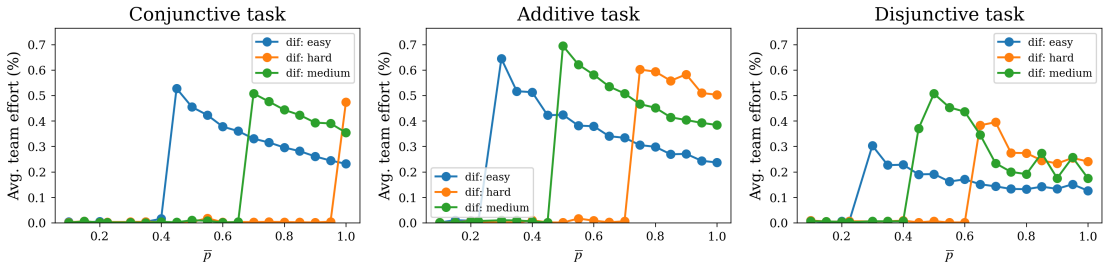


Figure 4.10: **Effect of evaluation hardness, via the passing threshold $G^{\mathcal{T}}$, on effort-expertise curves** (homogeneous teams). Passing thresholds are defined relative to $\hat{G}_{\max}$, the benchmark output of a homogeneous team of average-skilled agents ($p_i = 0.6$ for all i) fully contributing at every timestep.

Increasing evaluation hardness has a pronounced effect on the *timing* of cooperative behaviour: it reliably **delays activation**. Moving from easy to medium difficulty shifts $\overline{p}_{\text{act}}$ by approximately 0.23 across task types; moving from medium to hard adds a further ≈ 0.26. Within each difficulty level, once activated, the cooperation curve declines with expertise, reflecting the reduced effort required as teams become more capable. Because the curves preserve a broadly similar shape across difficulty levels, the horizontal shift induces a concomitant *vertical* increase in effort: at comparable points in the activation window, harder evaluations elicit higher contributions.

Tables 4.13 and 4.14 summarise these global effects. As before, statistical significance is determined through per-expertise contrasts.

1. **Conjunctive.** Teams facing medium-difficulty tasks expend roughly 34% more effort than those facing easy tasks; hard tasks add a further $\sim$33%.

2. **Additive.** The pattern mirrors the conjunctive case: vertical shifts are consistent, and medium-to-hard increases are of comparable magnitude.

3. **Disjunctive.** Vertical shifts remain present but are less stable, averaging around 40%. As with α, disjunctive tasks display reduced sensitivity at high expertise. Once a competent volunteer can guarantee success, increasing $G^{\mathcal{T}}$ has diminishing impact on group effort.

All per-expertise contrasts within the active range are strongly significant across task types.

Table 4.13: **Global vertical shifts in average cooperation across difficulty levels** (relative differences in mean effort).

Comparison	Conjunctive	Additive	Disjunctive
medium vs. *easy*	+34.4%	+35.9%	+44.7%
hard vs. *medium*	+33.8%	+33.4%	+27.4%
hard vs. *easy*	+51.1%	+51.6%	+50.1%

Table 4.14: **Horizontal shifts in activation thresholds** $\Delta\overline{p}_{\text{act}}$ across difficulty levels. Positive values denote later activation for the left condition.

Comparison	Conjunctive	Additive	Disjunctive
medium vs. *easy*	+0.25	+0.20	+0.26
hard vs. *medium*	+0.33	+0.25	+0.20
hard vs. *easy*	+0.55	+0.45	+0.46

Key Takeaways

In summary:

- **Switching behaviour in average cooperation.** General patterns of cooperation in Sequential Teamwork Games suggest that as soon as the task's passing threshold $G^{\mathcal{T}}$ becomes attainable, the average policy transitions abruptly from free-riding to active engagement.
- **Task type (ρ) as a structural driver.** Moreover, when agents can substitute for one another (additive), cooperation tends to be distributed within the team, increasing its overall value. When contributions are strict complements (conjunctive), teammates converge downwards, matching the weakest link. Under extreme substitutability (disjunctive), one or a few agents sustain effort while others free-ride. Considering the interaction between task type, average expertise, and composition: heterogeneity harms cooperativeness in conjunctive tasks, favours it in disjunctive tasks, and does not show a marked impact in additive tasks.
- **Leisure preference (α) discourages cooperation.** Higher α raises the marginal subjective value of leisure and systematically reduces cooperation through two levers: lowering time contributions and delaying activation. The effect is strongest in conjunctive

environments; additive tasks show milder but robust effects; in disjunctive tasks, the discouraging impact appears chiefly when immediacy preference is highest.

- **Passing threshold ($G^{\mathcal{T}}$) shifts activation.** Increasing $G^{\mathcal{T}}$ robustly delays the onset of cooperative behaviour across task types.

This closes a thorough examination of how team composition, task difficulty, and relevant model parameters interact to shape average effort. The next and final experiment shifts the focus to individual agents, delving into the psychological construct of implicit coordination, operationalising it through plausible behavioural proxies and applying it to the trained policies studied in this chapter.

4.5.3 Experiment 3: Teamwork Mediators in Sequential Teamwork Games

At the outset of this chapter, the aim was to design a model of teamwork grounded in well-established theories from social psychology while accounting for its temporal dynamics. To this end, the chapter draws on the IMOI framework [Ilgen et al., 2005], which highlights the cyclical nature of team processes: outputs at one stage become inputs for the next. A central element of this framework is the notion of **mediators**, that is, psychosocial mechanisms that help team members integrate resources, coordinate actions, and overcome motivational or organisational barriers.

The Sequential Teamwork Game model introduced earlier captures an abstract form of interaction in which coordination is not imposed *a priori*. Agents receive only partial observations of the task and of their teammates' behaviour. At each timestep, they decide how much effort to allocate, guided solely by their own reward functions, applying decentralised learning with decentralised execution. This choice deliberately prevents agents from building explicit models of one another: they learn independently and treat the remainder of the team as part of the environment. Importantly, while agents do not have access to others' identities or abilities, they do observe last-step effort bursts (Table 4.1), which provide behavioural traces but not explicit partner models.

This final experiment asks whether the learned policies exhibit behavioural traces of teamwork mediators. The main result is a dissociation: implicit coordination is strongest in conjunctive tasks, whereas temporal role specialisation is strongest in disjunctive tasks.

Background

When human teamwork unfolds over time, **team processes** emerge. A key class of these are **action processes**, which include behaviours aimed at monitoring tasks and making progress toward team goals [Marks et al., 2001]. Within the social and organisational psychology literature, **coordination** has been a central theme: it refers to behavioural patterns that align the actions, knowledge, and objectives of interdependent members in pursuit of a shared outcome [Rico et al., 2011, Arrow et al., 2000]. Coordination may be *explicit* or *implicit*. Explicit coordination relies on planning (*e.g.*, deadlines, schedules) and communication (oral or written exchanges).

Rico et al. [2008] argue that explicit coordination, although essential, provides too static a view of how teams actually function. They propose **implicit coordination** as a more dynamic alternative. Implicit coordination occurs when teammates *anticipate* task demands or others' needs without being asked and then *adjust* their behaviour accordingly [Rico et al., 2008, Espinosa et al., 2004, Fiore and Salas, 2004]. Anticipation requires private predictions about **i)** the demands of the task and **ii)** the likely behaviour or needs of other members. Dynamic adjustment involves ongoing mutual adaptation. Because it unfolds without communication, implicit coordination may appear automatic or even unconscious. Since agents in *TEMPO* cannot communicate and have no access to others' skills, this framework is well-suited to testing the spontaneous emergence of implicit coordination.

Implicit coordination is closely linked to another teamwork mediator: **Transactive Memory Systems (TMSs)**. A TMS is an emergent property of the team capturing a distributed meta-knowledge of the members' expertise (the "who knows what"). These representations are dynamically constructed and content-specific.

In the Sequential Teamwork Games defined in Section 4.3.2, each agent observes $o_i^t \in \mathbb{R}^{N+4}$ as detailed in Table 4.1. These local observations make some—but not all—components of a TMS explicit. All agents know the time remaining, the task type, and the deviation from the passing threshold. Crucially, they receive no explicit information about expertise (their own or others'). As a result, the "who knows what" element is absent by design. If such a knowledge structure is to arise, it must be an emergent outcome of interaction rather than a built-in feature.

Methods

Building on this framework, this section describes how two teamwork mediators—implicit coordination and TMS—are operationalised in *TEMPO*. Novel proxies for each construct are introduced and their values measured. All analyses are conducted post-hoc on policies obtained after training. Importantly, although these proxies identify behavioural regularities compatible with implicit coordination and TMS, they do not evidence internal partner models or theory of mind. They are therefore interpreted as emergent *functional* signatures rather than as evidence of cognition.

Proxies of implicit coordination. Implicit coordination comprises two components: **i)** anticipation and **ii)** dynamic adjustment. Anticipation captures the extent to which the team *proactively increases* output when a future shortfall is forecast and *decreases* output when success is already secured. Dynamic adjustment reflects compensatory effort following a sudden decrease in team productivity, consistent with *backing-up behaviour*.

Basic definitions. Let

$$E^t = \sum_{i=1}^{n} a_i^t$$

denote the team's aggregate effort at timestep t, where $a_i^t \in [0,1]$ is the effort allocated by agent i. Recall that cumulative productivity evolves as $G^{t+1} = G^t + Y^t$ with team output Y^t being

$$Y^t = \left(\sum_{i=1}^{n} \beta_i \, (g_i^t)^\rho \right)^{\frac{1}{\rho}}$$

The team output Y^t is used to project the expected final outcome under inertial effort[9]:

$$\mu^t = (1-\lambda)\, Y^t + \lambda\, \mu^{t-1}, \qquad \mu^0 \stackrel{\text{def}}{=} Y^0, \tag{4.17}$$

$$\widehat{G}^{H,t} = G^t + (H-t)\, \mu^t \tag{4.18}$$

Anticipation. Anticipation is measured through the extent to which teams adjust their effort in response to predicted shortfalls. Let $\widehat{G}^{H,t}$ denote the projected final output under inertial effort and define the (ex-ante) *need* as the remaining gap to the passing threshold:

$$N^t = G^{\mathcal{T}} - \widehat{G}^{H,t}$$

[9] $\lambda \in [0.3, 0.7]$ in robustness checks; results are qualitatively stable.

Changes in predicted need and collective effort are then given by

$$\Delta N^t = N^t - N^{t-1}, \qquad \Delta E^{t+1} = E^{t+1} - E^t$$

To avoid polluting the regressions with trends driven by proximity to the horizon, each variable is centred by timestep:

$$\widetilde{\Delta N}^t = \Delta N^t - \overline{\Delta N}^t, \qquad \widetilde{\Delta E}^{t+1} = \Delta E^{t+1} - \overline{\Delta E}^{t+1}, \qquad \widetilde{E}^t = E^t - \overline{E}^t$$

Team anticipation is then estimated via

$$\widetilde{\Delta E}^{t+1} = \theta_0 + \theta_1 \widetilde{\Delta N}^t + \gamma_1 \widetilde{E}^t + u^{t+1}$$

where u^{t+1} captures unexplained fluctuations. A positive and significant θ_1 indicates that increases (decreases) in anticipated need are followed by systematic increases (decreases) in collective effort, consistent with team-level anticipation.

Whereas anticipation captures proactive behaviour, dynamic adjustment focuses on reactive compensation following unexpected drops in productivity. **Dynamic adjustment.** Dynamic adjustment captures how teams compensate for negative shocks to output. Shocks are identified as

$$\Delta_-^t = \max\{0,\, Y^{t-1} - Y^t\}, \qquad \Delta_+^t = \max\{0,\, Y^t - Y^{t-1}\}$$

Because the terminal evaluation is logistic,

$$\sigma(G^H) = \frac{1}{1 + e^{-\kappa\,(G^H - G^T)}}$$

the same change in output may have different implications depending on the region of the assessment curve the team ends at. Therefore, a marginal utility weight is defined as

$$w^t = \sigma'\big(\widehat{G}^{H,t}\big) = \kappa\,\sigma\big(\widehat{G}^{H,t}\big)\big[1 - \sigma\big(\widehat{G}^{H,t}\big)\big]. \tag{4.19}$$

which scales shocks by their impact on the probability of success:

$$\Delta_{-,*}^t = w^t\,\Delta_-^t, \qquad \Delta_{+,*}^t = w^t\,\Delta_+^t$$

Given a negative shock at time t_m, the dynamic adjustment index is defined as the fraction of the shock that is compensated within a recovery window of length τ:

$$D_*^m \;=\; \frac{\min\Big\{\sum_{k=1}^{\tau} \Delta_{+,*}^{t_m+k},\;\; \Delta_{-,*}^{t_m}\Big\}}{\Delta_{-,*}^{t_m}}$$

so that $D_*^m \in [0,1]$, where 0 indicates no compensation, 0.5 partial compensation, and 1 full compensation. Notice that index m enumerates shocks, not timesteps. Aggregating across shocks with effective-size weights yields

$$\bar{D} \;=\; \frac{\sum_m \Delta_{-,*}^{t_m} D_*^m}{\sum_m \Delta_{-,*}^{t_m}}$$

This bounded index can be interpreted directly as the average fraction of shocks compensated by the team.

Proxies of Transactive Memory System (TMS). A TMS combines individual knowledge with team-level awareness of "who knows what". TMS is operationalised via an index defined on *effort allocations* rather than realised output:

Role specialisation. Let $a_i^t \in [0,1]$ be the fraction of timestep t that agent i allocates to the task, with a team of n agents over horizon H. If one picks a random unit of effort from the whole episode, the probability of it coming from agent i at time t is

$$\mathbb{P}(i,t) = \frac{a_i^t}{\sum_{i',t'} a_{i'}^{t'}}$$

with marginals $\mathbb{P}(i)$ and $\mathbb{P}(t)$. With this distribution, the Mutual Information (MI) index can be computed, which answers the question "*how much knowing the timestep t reduces uncertainty about which agent provided effort?*":

$$\mathrm{MI}(I;W) = \sum_{i=1}^{n}\sum_{t=1}^{H} \mathbb{P}(i,t)\log\frac{\mathbb{P}(i,t)}{\mathbb{P}(i)\mathbb{P}(t)}$$

In this section the normalised version—the Normalised Mutual Information (NMI)—is used, which allows comparisons across task types and runs:

$$\mathrm{NMI}(I;W) = \frac{\mathrm{MI}(I;W)}{\min\{\mathrm{H}(I),\mathrm{H}(W)\}} \in [0,1]$$

Here, $H(\cdot)$ denotes Shannon entropy, unrelated to the episode horizon H. High NMI[10] indicates that each agent's effort is sharply localized in time, with little overlap; values near 0 reflect that time and agent identity are independent: everyone contributes similarly across time. In this temporally extended environment, role differentiation naturally manifests as temporal specialisation, making MI a suitable signal of emergent "who does what and when".

The next section summarises the behavioural traces symptomatic of implicit coordination and role specialisation found across trained policies in *TEMPO*.

[10] If $\sum_{i',t'} a_{i'}^{t'} = 0$, Normalised Mutual Information (NMI) is set to 0 by convention.

Results and Discussion

The above proxies of implicit coordination and roles specialisation were evaluated on homogeneous teams, setting $\alpha_i = 1$ for all i, and the analysis was restricted to medium-difficulty tasks. For each task type, anticipation, dynamic adjustment and NMI indices were first computed at the episode level. These values were then averaged within each seed, ensuring that no episodes were crossed between seeds. Finally, the seed-level results were aggregated by computing averages across seeds. The box plots presented below display these cross-seed distributions, thereby reflecting variability due to random initialisation rather than episode-level noise.

Implicit Coordination The resulting anticipation (Fig. 4.11) and dynamic adjustment (Fig. 4.12) indices suggest that implicit coordination was preponderant in conjunctive tasks (*vs.* additive and disjunctive). Specifically, anticipation was highest in conjunctive tasks ($\theta_1 = .277$), followed by additive tasks ($\theta_1 = .208$). Anticipation in disjunctive tasks was near zero ($\theta_1 = .03$). Across task types, the regression models underlying the anticipation proxy achieved moderate but consistent fits ($R^2 \approx 0.2-0.3$ for additive and conjunctive tasks, and $R^2 \approx 0.44-0.46$ for disjunctive tasks), indicating that while the simplified specification does not capture all variability, it explains a stable portion of the dynamics across seeds.
As per the dynamic adjustment index, the maximum recovery window length allowed was $\tau = 3$. Teams demonstrated highly effective compensation through dynamic adjustment for negative shocks in strongly conjunctive tasks ($\bar{D} = .9$), moderate adjustment in additive tasks, and the weakest adjustment in strongly disjunctive tasks.

Taken together, these results align well with the broader argument of the chapter. In conjunctive tasks, where free-riding by anyone is severely penalised with team failure, agents develop implicit coordination to secure positive rewards. Specifically, they learn policies that anticipate when the team will fall short, and that reassess and readjust throughout episodes. Additive tasks, where gifts are substitutable but teammates still need to coordinate to avoid effort waste, induce moderate implicit coordination. Finally, disjunctive tasks can be (and usually are) carried out by a single capable teammate. Once someone *volunteers* to work, disjunctive tasks become individual work where coordination is no longer necessary. Even when several agents split work, they do so by alternating bursts of effort across timesteps—but seldom concurrently. In short, conjunctive tasks induce the highest levels of interdependency, followed by additive tasks. In disjunctive tasks, the only coordination needed is to designate

a capable leader to carry out the work. Interestingly, theoretical and empirical work in organisational psychology Rico et al. [2008] suggests that task interdependence moderates the relationship between implicit coordination and team performance. In particular, situations that make team tasks more interdependent especially benefit from the utilisation of implicit coordination mechanisms. The results of the above simulations are consistent with this thesis.

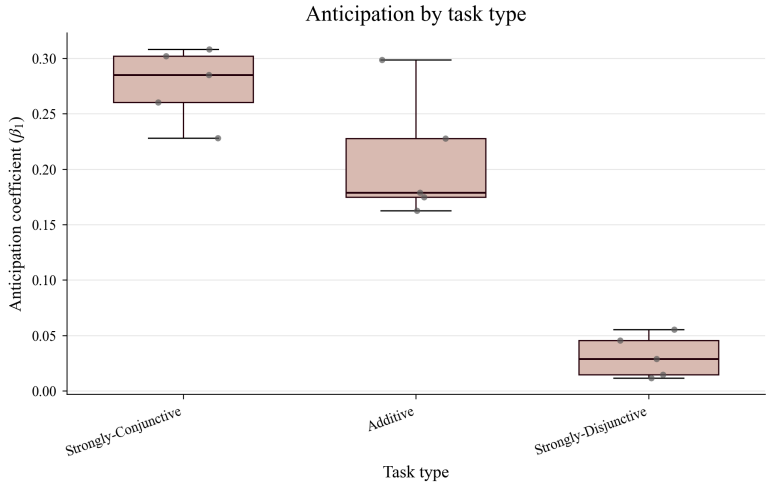


Figure 4.11: **Distribution of anticipation coefficients (θ_1) across task types**. Each box summarises seed-level averages over multiple rollouts.

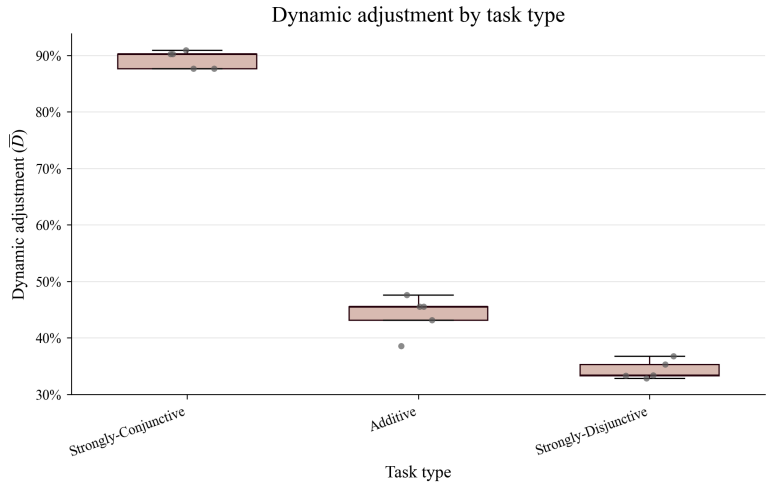


Figure 4.12: **Distribution of dynamic adjustment indices ($\overline{D}$) across task types**. Values closer to ~ 1 indicate more effective compensation of negative productivity shocks.

Transactive Memory System (TMS) Post-hoc computation of the Normalised Mutual Information (NMI) index on the trained policies unveils a more complete picture of the mediating mechanisms behind cooperation. As mentioned before, NMI measures how distributed work is across agents and time, making it a proxy for the team's awareness of who should step in. The boxplot in Figure 4.13 reveals a clear pattern: disjunctive environments induce the highest levels of this meta-knowledge (NMI $\sim$ 53% across seeds), followed by additive tasks (NMI $\sim$ 10%). By contrast, NMI in conjunctive tasks was negligible (NMI $<$ 1%), consistent with the high concurrency observed in these environments. This is reflected in their *active fraction* (the proportion of total possible effort actually invested across agents and timesteps), which reaches about 35% in conjunctive tasks, compared with only 17% in disjunctive teams. In other words, conjunctive teams work largely in parallel, additive teams coordinate moderately by staggering gifts, and disjunctive teams develop the strongest temporal differentiation of roles.

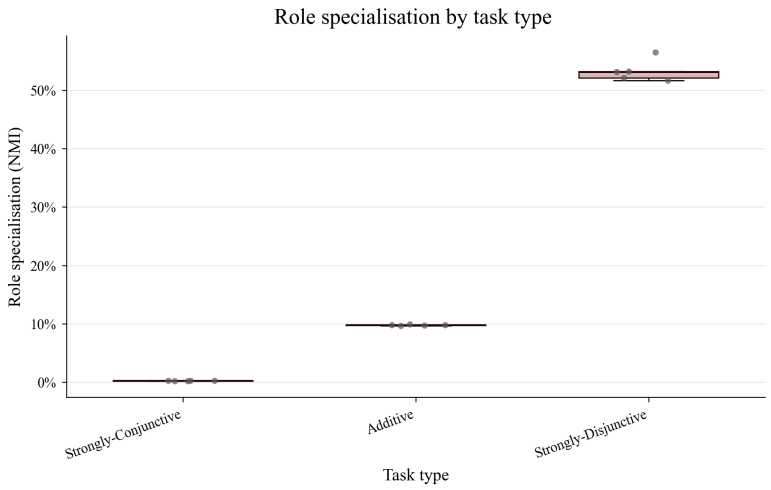


Figure 4.13: **Distribution of role specialisation (NMI) across task types**. Higher values reflect stronger alignment between agents and temporal roles.

Key Takeaways

To summarise:

- **Emergent division of labour and coordination timing.** Individual trajectories suggest an implicit division of labour moderated by task type. In additive settings, higher-expertise agents learn to carry the workload (social compensation) while weaker links completely

free-ride and mid-skilled agents avoid effort concurrence; in disjunctive tasks, one or two agents shoulder the task while the rest loaf, echoing free-riding [Smith, 2022]. In conjunctive tasks, weaker members work hard to avoid being responsible for a weak group performance (Köhler effect). Policies also display implicit coordination: agents avoid redundant early effort and converge on full commitment as the episode unfolds.

- **Task interdependence shapes the emergence of implicit coordination.** The more interdependence a task imposes, the stronger the traces of implicit coordination that teams display. In conjunctive tasks—where everyone must contribute for success and members depend most on one another—both indices of implicit coordination (anticipation and dynamic adjustment) score highest. Additive tasks, which involve a moderate degree of interdependence—balancing the need to avoid wasted effort with the need to secure collective success—come second. Finally, in disjunctive tasks, two conditions must co-occur: **i)** at least one capable member must be present to carry the workload, and **ii)** the rest of the team must acknowledge them as the *best shot* and accept their output as the team's. This structure entails some interdependence during the selection stage, but almost none once work begins: the designated agent acts alone. Accordingly, the indices for implicit coordination decrease as interdependence weakens, in line with findings from social psychology Rico et al. [2008].

- **Traces of Transactive Memory Systems are strongest in disjunctive tasks.** By design, the proxies of implicit coordination did not capture the critical moment of interdependence in disjunctive tasks: the selection of a spokesperson or leader. Here, NMI as a proxy for TMS offers a complementary lens. It reflects how much agents appear to "know who knows what" and how they distribute effort accordingly. On this measure, conjunctive tasks score lowest—since success hinges on the weakest member, knowledge of who knows more or less adds little—while disjunctive tasks score highest, consistent with their reliance on a single best-shot contributor.

The findings above suggest that Sequential Teamwork Games naturally reproduce key psychosocial mechanisms observed in human groups: implicit coordination emerges most clearly when tasks require tight interdependence, whereas TMSs arise most strongly when success depends on recognising and leveraging individual differences.

More broadly, the results of this chapter illustrate how Sequential Teamwork Games can function as computational testbeds for studying teamwork beyond artificial agents alone. The

goal is not to claim direct psychological realism or to predict individual human trajectories. Rather, the framework provides a controlled setting in which structural properties of teamwork—such as interdependence, substitutability, expertise heterogeneity, and evaluation pressure—can be systematically manipulated and linked to emergent behavioural regularities.

From this perspective, *TEMPO* complements empirical and organisational research by enabling the exploration of how different teamwork structures may promote free-riding, coordination, compensation, role differentiation, or disengagement. The resulting patterns should therefore be interpreted not as literal models of human cognition, but as mechanistic hypotheses about how incentive structures and temporal dynamics shape collective behaviour in human and hybrid teams.



5 *Agents, Firms and Identity*

> The subjects did not even know which other persons were in their group. Group formation did not seem to reflect but to *cause* attraction between people: People liked others not as *individuals* but as members of the same *group*.
>
> *J. C. Turner & S. A. Haslam*

5.1 Introduction: Times of Change 152
5.2 Self-Categorisation Theory and Groups at Work 154
5.3 Existing Multi-Agent and Identity-Based Models in Economics 157
5.4 An Agent-Based Model of Firm Identity and Effort Contribution 160
5.5 Comparative Statics 171
5.6 Simulation Design and Implementation 183
5.7 Experiments 185

Thus far, the models presented in this dissertation (Chapters 3 and 4) have shared a common motivation: given a team description, a task specification, and a collective goal, both propose multi-agent models of unenforced teamwork cooperation.

Chapter 3 framed teamwork as a mixed-motive, one-shot game—the Teamwork Game—clarifying the *structural* drivers of cooperation and defection through a multi-agent multi-armed bandit environment predicting boundedly rational behaviour (*CollabQuest*). Chapter 4 extended this analysis to the temporal domain, examining how coordination, pacing, and interdependence unfold over time when teamwork is modelled as a Markov game played by independent PPO learners (*TEMPO*).

Both CollabQuest and TEMPO focus on the **team** as the primary unit of analysis. In these settings, the team is a *closed system* in which relationships of **interdependence** are induced by the task characteristics and the incentive scheme that rewards (or penalises) team performance. Agents decide only how much they wish to contribute to their group, either as a one-shot decision or sequentially.

The key departure from these previous chapters is that now the team is replaced with the **firm**—*i.e.* company—as the object of study, and is analysed as an *open system*. Agents can now make decisions at two levels: **i)** whether they wish to remain in their firm or move to another, and **ii)** once a firm is chosen, how much they want to contribute to it sequentially. These decisions are taken in the pursuit of maximal utility.

Unlike the previous settings, the utility formulation allows for a gradable shift in the locus of optimisation—ranging from purely individualistic to fully social motives. Specifically, the activation of *firm identity* will be formalised in a manner consistent with how it has been defined in Self-Categorisation Theory (SCT) [Turner, 1985, 1987, Turner and Haslam, 2001].

Chapter Contributions. This chapter advances the field of Agent Based Modelling (ABM) by integrating grounded psycho-social mechanisms into the decision-making processes of utility-maximising agents. In doing so, it shows—both analytically and computationally—that the activation of firm identity can fundamentally reshape cooperation patterns, welfare outcomes, and the emergent structure of the economy. The chapter thus completes the computational theory of cooperation developed throughout this dissertation, with a specific focus on the *social* alignment mechanisms that support or inhibit cooperation within teamwork.

The theoretical model to be presented bridges individual cognition, social identity theory, and aggregate economic regularities, showing that identity-driven motivations can be mod-

elled as endogenous components of agents' preferences without sacrificing analytical tractability. In particular, the chapter makes the following contributions:

(C1) **Mathematical formalisation of self-categorisation dynamics.** The chapter provides a formal, micro-founded representation of how social identity processes influence economic behaviour. It includes:

- *Formalisation of the Meta-Contrast Ratio (MCR):* a mathematical definition of the MCR that operationalises the comparison of intra- and inter-group similarities, consistent with the postulates of Self-Categorisation Theory (SCT) [Turner, 1985, 1987, Turner and Haslam, 2001].
- *Continuous salience function:* the specification of a smooth mapping linking the MCR to the salience—*i.e.* activation—of a given social identity, thereby allowing salience to vary endogenously across agents and contexts.
- *Salience-weighted utility formulation:* the introduction of a hybrid log-utility function that integrates individualistic and social frames of evaluation. This embeds identity salience directly into the optimisation problem, enabling endogenous shifts between egoistic and collectivistic behaviour.

(C2) **Analytical results on salience, firm size, and cooperation.** Through comparative statics, the chapter derives the joint effect of variations in identity salience and firm size on agents' optimal effort. It formally proves that higher salience of social identity may enhance cooperative behaviour, while also identifying the boundary conditions under which this effect may reverse. Similarly, it demonstrates that, in most cases, increasing firm headcount inhibits individual contributions, thereby linking micro-level motivation to organisational scale.

(C3) **Analytical decomposition of welfare and firm-size effects.** The model provides a closed-form decomposition of how equilibrium utility varies with firm size and salience. It distinguishes between the productivity and identity components of welfare, identifying the parameter regimes in which each channel becomes dominant.

(C4) **Multi-scale empirical study through large-scale simulations.** An extensive Multi-Agent Systems (MAS) experiment with a hundred thousand agents confirms the analytical predictions across micro, meso, and macro scales. The simulations reproduce theoretical

comparative statics, reveal the coexistence of behavioural regimes (free-riders vs. strong contributors), and show how endogenous salience shapes firm-level and aggregate regularities.

(C5) **Characterisation of emergent structural and statistical regularities.** Finally, the simulations show that the model reproduces key empirical regularities of decentralised economies, including the heavy-tailed distribution of firm sizes and stable macro patterns emerging from individual heterogeneity and identity dynamics. This provides a unified account of how identity-based motives can scale to produce aggregate economic structure.

The motivating question is how firm identity, understood as a salient social category, shapes agents' cooperative behaviour and firm-level outcomes in an economy of utility-maximising agents. The core modelling insight is that social identity can be represented as a scalar field, assigning each agent a context-dependent level of identity salience. This salience emerges endogenously from the contrast between the focal group and relevant out-groups. In addressing this, the chapter introduces a multi-agent model in economics in which the salience of social identity is *endogenously generated* from contextual inter- and intra-group contrast, as formalised by the meta-contrast ratio. This salience is then embedded directly into agents' optimisation through a salience-weighted utility formulation consistent with the postulates of Self-Categorisation Theory.

5.1 *Introduction: Times of Change*

As is customary in this dissertation, Anna's story will once again serve to motivate the purpose of the present chapter. In Chapter 4, we left Anna on the verge of becoming a minor shareholder in the company where she worked. Good news: the app was completed on time and with great success, and Anna was duly rewarded after months of intense effort.

Five more years have now passed, meaning that Anna has accumulated ten years of professional experience—eight of them within the same firm. Somehow, this turn of the decade has prompted her to look back on the past years and reflect on what lies ahead. In reviewing her future prospects and revisiting the ambitions she had set for herself when she was a recent

graduate, she realises that her job satisfaction has gradually declined. The projects that once excited her no longer inspire much enthusiasm.

Anna is aware that part of this disenchantment stems from a takeover bid and the subsequent merger of her mid-sized software firm with a multinational tech corporation. This change has had both direct and indirect effects on her situation. First, she was required to sell her shares, and thus she is no longer a shareholder. Second, the merger brought about a complete reorganisation of the firm, including several layoffs. Among those dismissed were most of her department colleagues—the ones she had been closest to. They have now been replaced by employees from the merged company who already knew one another. In fact, her new teammates share an unspoken language and private jokes. Individually, they are all polite and friendly, and their interactions with her are perfectly cordial. Yet, when they are together, they revert to their shared codes and symbols and, although Anna knows they mean no harm, she cannot help but feel displaced—an outsider looking in.

Despite her efforts, the situation has not improved: one-on-one, her relationships remain pleasant, but she simply cannot get in. And to be frank, she finds herself caring less and less. With the recent payout from the sale of her shares and the experience she has gained, Anna feels that she may be at the ideal moment to pivot and change her career path.

On one hand, the idea of founding her own firm has always appealed to her. The prospect of working on challenges that truly inspire her—without having to answer to anyone—is undeniably attractive. Yet entrepreneurship comes with the risk of loss, and she is not sure she is risk-prone enough to stake the capital she has worked so hard to accumulate.

On the other hand, there are a couple of companies that have approached her in recent years and that, as far as she knows, share to some extent the values that have guided her work. She has friends there and people willing to recommend her, so the transition should not be too difficult. Deep down, Anna feels that in a place where she is surrounded by people more similar to herself, she would find it easier to perform.

This chapter presents a multi-agent system that abstracts the essential elements of such situations, modelling the processes of work contribution and labour mobility through an integration of computational economics and social-psychological theory. Before formalising these mechanisms, the theoretical foundations that inform the model must first be introduced. This is the purpose of the next section.

5.2 Self-Categorisation Theory and Groups at Work

Among the various frameworks that explain the emergence of socially aligned behaviour, Self-Categorisation Theory (SCT) provides perhaps the most comprehensive account of how individuals' perceptions of self and group membership shape cooperation and identity-based motivation. At its core, SCT proposes that individuals do not always act purely as isolated agents. Depending on the social context, they may instead perceive themselves primarily as members of a group, causing their motivations, evaluations, and behaviour to shift accordingly. Accordingly, SCT is fundamentally a theory of the psychological group. The adjective *psychological* underscores the group's nature as a *reference group*: a group with which its members subjectively identify in order to establish social comparisons, to act, and to acquire *values and norms* Turner [1987].

In its original formulation, the theory was articulated as a series of seventeen hypotheses and a number of assumptions concerning the functioning of the social *self-concept* (*i.e.*, the concept of self based on comparison with other people and relevant to social interaction) [Turner, 1984, 1985, 1987]. This section draws extensively on Turner [1987] and Turner and Haslam [2001] to summarise a subset of these assumptions and hypotheses, focusing specifically on those directly related to the multi-agent model to be introduced in this chapter.

Two ideas are especially important for the present chapter. First, SCT assumes that individuals possess multiple possible representations of the self. Second, which representation becomes active depends on the social context.

In SCT, the self-concept is defined as the cognitive component of the *self*; and is composed of multiple elements. Each of these elements constitutes a cognitive representation of self available to the individual. In other words, an individual possesses multiple self-concepts, each activated in specific situations depending on contextual factors and individual predispositions. When a self-concept is activated, it is said to become *salient*. A crucial postulate of the theory is that cognitive representations take the form of *self-categorisations*. That is, human beings classify their own self and certain classes of stimuli—understood here as socially relevant cues such as traits, behaviours or observable attributes—as belonging to the same category (*i.e.*, as interchangeable) in contrast to other classes of stimuli. These categorisations are hierarchically organised, with more abstract categories encompassing more specific ones. At a minimum, self-categorisation operates at three levels:

1. The **superordinate level**, in which the self is perceived as a member of the human species.

2. The **intermediate level**, where individuals categorise themselves as members of an *in-group*, distinguishing themselves from *out-groups* based on perceived similarities and differences.

3. The **personal (or individual) level**, where self-perception is based on similarities and differences with other in-group members, highlighting one's uniqueness.

The Meta-Contrast Ratio (MCR) Principle. The process of categorisation is driven by stimulus comparison and follows the Meta-Contrast Ratio (MCR) principle: within a given context, a set of stimuli is more likely to be grouped together as identical when intra-group similarity is high and contrast with external stimuli is maximised. Consequently:

- Personal self-categorisations emerge from comparisons between the self and other in-group members (*i.e.*, interpersonal comparisons occur within the in-group).
- In-group/out-group categorisations arise from comparisons with other human beings (*i.e.*, intergroup comparisons occur within the human species).
- Human self-categorisations are based on comparisons with other species, operating at a more inclusive level of identity.

In the computational model developed later, this comparative mechanism will determine the endogenous salience of firm identity.

Consequently, self-perception varies along a *continuum*, ranging from perceiving oneself as a unique individual (where intrapersonal identity is maximised and the perceived difference between oneself and in-group members is greatest) to perceiving oneself primarily as a member of an in-group category (where similarity with in-group members is maximised and contrast with out-group individuals is strongest). Typically, individuals occupy an intermediate position on this spectrum. Factors that increase the salience of in-group/out-group categorisations tend to *accentuate perceived intra-class similarity and inter-class differences*, resulting in the *depersonalisation* of self-perception—that is, a shift from personal to group-based self-definition—along the stereotypical dimensions defining in-group membership.

Depersonalisation constitutes a fundamental psychological process **underlying all group phenomena**, including social cooperation, altruism, and the emergence of shared norms. In other words, group behaviour reflects a shift in self-concept whereby individuals move from perceiving themselves as unique persons to self-categorising within an in-group/out-group

framework. This transition leads to a diminished sense of individuality and an increased perception of oneself as an **interchangeable** member of the in-group. In depersonalising the self, salient social identity transforms differing individual self-interests into social we-group interests. These ideas become especially relevant in situations involving social dilemmas, where individual and collective incentives may diverge.

SCT provides a useful framework for interpreting certain—otherwise puzzling—results in game theory as manifestations of shifts in players' motivational structures induced by varying levels of social-identity salience. For instance, it is well established that the only Nash equilibrium in the one-shot Prisoner's Dilemma does not yield the highest possible payoff for the players. Classical game theory therefore predicts that rational players will converge on mutual defection rather than cooperation, despite the latter being collectively preferable.

Empirical studies on one-shot social dilemmas broadly support the existence of this tension, showing that cooperation rates are often substantially lower than the socially optimal outcome would require [Dawes, 1980, Axelrod and Hamilton, 1981, Eiser, 1978]. For example, Eiser [1978] reported that cooperation occurred in only 30% of cases in experimental Prisoner's Dilemma settings. At the same time, the persistence of non-negligible cooperation suggests that additional psychological and social mechanisms may shape behaviour beyond purely individual payoff maximisation.

Indeed, early research on social dilemmas soon identified other mechanisms associated with increased cooperation. For instance, Dawes [1980] observed that increasing the perceived similarity between players or making their shared group identity more salient led to a higher incidence of cooperative behaviour. A clue to a possible explanation had already been suggested in Kelley and Thibaut [1978] and Kelley [1979] through the notion of *psychological transformations*. According to Kelley [1979], when faced with dilemma situations, individuals transform the objective or "given" payoff matrices into *effective* matrices through a series of psychological processes, which, he argues, provide a better account of their actual behaviour. De Cremer and Van Vugt [1999] explored two competing explanations for the observed positive effect of social identity on cooperation in social dilemmas: **i)** the *goal-transformation hypothesis*, which posits—similarly to the thesis in Kelley and Thibaut [1978]—that increased group identification raises the perceived value of collective goods, and **ii)** the *goal-amplification hypothesis*, which suggests that stronger group identity enhances trust in others' cooperation. Experimental results supported the goal-transformation hypothesis, indicating that **even self-**

interested individuals can be encouraged to cooperate by increasing the salience of their group membership.

In a similar vein, research by Breckler and Greenwald [1986] and Karau and Williams [2001] demonstrates that strengthening identification with an in-group can motivate individuals to exert greater effort, driven by a desire to validate their *collective self*. Likewise, Kramer et al. [2014] notes that when social identity is salient, individuals are more inclined to align with collective goals, reshaping their decision-making processes in social dilemmas.

The strong link between these findings and the domain of social dilemmas and game theory allowed them to resonate beyond social and organisational psychology, influencing the field of economics through the rise of behavioural economics [Karau and Williams, 2001, Ostrom, 2003b]. This broader research programme also influenced computational and evolutionary approaches to cooperation, including work on adaptive social networks, structured populations, and emergent group behaviour in multi-agent systems [Santos et al., 2006a,b]. The following section delves into the intersection between Agent Based Modelling and behavioural economics, reviewing prior efforts to integrate ideas from SCT into economic theory.

5.3 *Existing Multi-Agent and Identity-Based Models in Economics*

The use of Multi-Agent Systems (MAS) for the study of social phenomena is not new to economics. As detailed in a recent and comprehensive review [Axtell and Farmer, 2025], multi-agent simulation enables economists to explore the bottom-up processes through which micro-level behaviour gives rise to macroeconomic phenomena. In line with its applications to other social sciences and ecology, the typical pipeline involves: **i)** constructing a computational model that imposes constraints on individual agent behaviour (*e.g.*, maximising a specific utility function, or lowering prices by some n% when inventory exceeds target levels); and **ii)** observing the emergent phenomena that appear at the aggregate scale, allowing the system to evolve under its own dynamics without solving equations for the governing macro-social structure[1]. In the economics jargon, such methodology is known as Agent Based Modelling (ABM).

[1]Complementary, top-down applications of MAS—where macro-level targets guide the search for micro-behaviour—have also been explored in the literature [Sierra et al., 2003].

A renowned family of models adopting this approach is that developed by Axtell [Axtell et al., 1999, Axtell, 2002, 2016a, 2018]. These models study firms that self-organise within populations of heterogeneous, boundedly rational agents who interact locally and under partial information. According to the author, their purpose is twofold: first, to test existing theories of the firm by assessing their generality through systematic experimentation in a computational "laboratory" where out-of-equilibrium dynamics can be examined; and second, to contribute to MAS research by grounding agent behaviour in microeconomic theory. The resulting models reproduce well-known empirical regularities such as the right-skewed size distribution of firms, and remain robust to alternative preference specifications, including deviations from Cobb-Douglas [Axtell et al., 1999].

For this reason, Axtell's job-mobility and firm-formation framework provides the economic backbone of the model developed in this chapter. The proposed model extends this framework by preserving its core structure—team production, equal sharing of firm output, income–leisure trade-offs, boundedly rational local search, and endogenous firm dynamics—while augmenting agents' utility with an endogenous social-identity component grounded in SCT.

This extension is motivated by a central limitation of purely pecuniary models of team production: the economic literature on employee compensation has repeatedly shown evidence of non-economic motivations underlying workers' behaviour [Prendergast, 1999, Akerlof and Kranton, 2005]. Nobel Memorial Prize in Economic Sciences laureate George A. Akerlof, together with Rachel E. Kranton, pointed to *identity* as one such missing source of motivation in standard economic models of the organisation [Akerlof and Kranton, 2000a]. Building on sociological and psychological theories of identity (and specifically citing Turner [1987]), they proposed a series of economic models of increasing complexity, where work incentives depended on both pecuniary and non-pecuniary components, offering an alternative framework for analysing effort provision and the internal dynamics of firms.

For instance, in Akerlof and Kranton [2005] the authors propose a *prototype* principal-agent model with identity, in which the principal (employer) offers a contract with wages w_H^c, w_L^c to an agent (worker). The worker must choose an effort level $e \in [e_B, e_A]$ that probabilistically affects revenues, and receives the corresponding wage. The novelty of the approach lies in allowing the agent's decision to depend not only on pecuniary considerations but also on identity-related motives. Each agent belongs to a fixed category C, from which she derives intrinsic utility. Membership prescribes a normative effort level $e^\star(C)$, and deviation from that prescription entails a loss of utility.

Applying standard participation and incentive constraints, Akerlof and Kranton [2005] show that identity reduces the need for monetary rewards. In this sense, identity functions as a form of *motivational capital*: if the firm can convert an outsider into an insider, the resulting change in identity increases the worker's utility at any given wage, allowing the firm to offer lower pay for the same overall level of utility.

The theoretical relevance and qualitative significance of this work for bridging economics and psychology are undeniable, as it laid the first stone for a new research stream that, in many cases, explains behaviour more accurately than classical economic models: *Identity Economics*. In light of the brief recapitulation of the main concepts of SCT, however, it becomes apparent that certain inconsistencies arise between the formalisation by Akerlof and Kranton [2005] of key psycho-social constructs—and their interplay within the model—and the way these constructs are conceived in the original theory. These can be distilled into the following list:

1. First, the model lacks an equivalent to the process of self-categorisation: categories are predefined and presented to the individual with constant intensity, regardless of the context. However, the central thesis in Turner [1985] is that the degree to which an individual identifies with a group depends on the situational context. Social identity is neither omnipresent nor expressed in the same way across situations; rather, it is characterised by its dynamic and relational nature. In other words, the salience of a given self-categorisation depends on the contrast perceived by the individual between members of the in-group and those of the out-group.

2. Second, by not allowing for a dynamic representation of social-identity salience, the model cannot capture the variation in employees' identification along a continuum of group salience. Once an individual is assigned to a category, the identity utility affects total utility in the same way for all members. The only means through which the identity-related utility (or more precisely, the identity-related loss) is modulated in the model is via a weight parameter, which is constant and common to all members of the same category.

3. Third, just as the categories are defined exogenously, the authors also impose the social norm of behaviour (the *prescribed behaviour*). By contrast, Self-Categorisation Theory posits that the social norm represents the behavioural translation of the group prototype. In other words, rather than being imposed, the social norm emerges *bottom-up* from the processes of categorisation.

4. Consequently, this leads to what is perhaps the most substantial and consequential divergence for the model's predictions: the (mistaken) assumption made by Akerlof and Kranton [2005] that high identification with the group necessarily entails high contribution. In their own words: "When workers identify with the firm, they act as its fiduciaries and in its interest", which leads them to conclude that non-pecuniary motivations and wages are substitutes. While this may indeed hold in firms whose social norm prescribes cooperation or high effort, it need not be the case in general. What SCT actually posits is that **the greater the salience of group categorisation, the greater the adherence to the social norm**—whatever that norm may be.

Another important approach, *team reasoning* [Bacharach, 1999], models the shift from individual to collective frames of evaluation. While aligned with the central insight of SCT, it formalises the cognitive transition itself rather than its underlying dynamics or contextual salience, treating the activation of team reasoning as a probabilistic event rather than as an emergent function of social context.

The next section builds on these behavioural economics insights to introduce a new formalisation of group identity within Axtell's framework. The model endogenises self-categorisation through the Meta-Contrast Ratio (MCR), represents salience on a continuous scale, and allows the social norm to *emerge* from the group prototype. This yields a utility structure capable of capturing a smooth transition between individualistic and social motives.

5.4 An Agent-Based Model of Firm Identity and Effort Contribution

This section introduces the proposed extension of Axtell's job-mobility and firm-formation model. It first motivates the key elements that a general utility function must include to remain compatible with SCT. It then formalises the mechanisms of social-identity salience through the definition of the MCR and the associated salience function. Finally, it embeds these mechanisms into a specific utility formulation grounded in Axtell et al. [1999].

5.4.1 Setup and Notation

Standing assumptions. The present model represents job mobility and effort contribution throughout several decision-making rounds or timesteps. To simplify notation, an arbitrary timestep t is fixed and time indices are suppressed. All variables in this section refer to this t imestep unless stated otherwise. The time budget is normalised to $\Delta = 1$, so optimal efforts lie in $e_i^\star \in [0,1]$. Unless otherwise stated, comparative analyses keep fixed the parameters (a, b, β), the preference parameters α_i, and the firm-level statistic $\widetilde{\alpha}$.

Population and firms. Let $\mathcal{A} = \{1, \ldots, n\}$ be the set of agents and let $\mathcal{F}$ denote the (endogenous) set of firms. Each agent $i \in \mathcal{A}$ is employed by exactly one firm. Formally, there exists an assignment function

$$\varphi : \mathcal{A} \to \mathcal{F}$$

where $\varphi(i)$ is the firm employing agent i. For any firm $f \in \mathcal{F}$, let

$$\mathcal{A}_f \stackrel{\text{def}}{=} \{i \in \mathcal{A} : \varphi(i) = f\}$$

denote the set of its members, and write $n_f = |\mathcal{A}_f|$ for its size. By construction, the family $\{\mathcal{A}_f\}_{f \in \mathcal{F}}$ forms a partition of $\mathcal{A}$: every agent belongs to one and only one firm.

Each agent $i \in \mathcal{A}$ chooses an effort level $e_i \in [0,1]$, devoting the remaining time $1 - e_i$ to leisure, and has a preference parameter for income $\alpha_i \in (0,1)$. Similarly, each firm $f \in \mathcal{F}$ is characterised by a behavioural statistic $\widetilde{\alpha}_f$ that captures the group's characteristic valuation of income relative to leisure. In line with the interpretation of social identity as a shift toward the group prototype, $\widetilde{\alpha}_f$ is defined as the within-firm median of the individual preference parameters:

$$\widetilde{\alpha}_f \stackrel{\text{def}}{=} \text{med}\{\alpha_i : i \in \mathcal{A}_f\}$$

The median provides a robust summary of the firm's normative orientation—insensitive to outliers—and serves as the anchor for the social frame of evaluation introduced later in the utility specification.

Production technology and income sharing. For any firm $f \in \mathcal{F}$, total effort is

$$E_f = \sum_{i \in \mathcal{A}_f} e_i$$

and output is generated according to

$$O(E_f) = a_f E_f + b_f E_f^{\beta_f}$$

where a_f, b_f, and β_f are the firm's technological parameters. For notational simplicity, we sometimes omit the firm subscript when focusing on a single firm. Consistent with the original framework in Axtell [2016b], firm output is shared equally among all members. Beyond preserving comparability with the baseline model, equal sharing captures a stylised form of team production in which individual contributions are difficult to disentangle and rewards depend on collective performance. This assumption is also theoretically aligned with the present chapter's focus on common fate and group-based evaluation, since all members become jointly exposed to the firm's aggregate outcome. Each agent $i \in \mathcal{A}_f$ thus receives:

$$y_i = \frac{O(E_f)}{n_f}, \qquad n_f \stackrel{\text{def}}{=} |\mathcal{A}_f|$$

Alternative compensation schemes beyond simple equal sharing—including seniority-based rewards and individual bonuses—have also been explored within Axtell-type models [Axtell et al., 1999]. The present chapter focuses on equal sharing in order to isolate the behavioural effects of endogenous identity salience under collective incentives. Extending the framework to heterogeneous compensation schemes constitutes a natural direction for future work.

Since within-firm incomes are identical under equal sharing, the firm's median income coincides with this common value:

$$\widetilde{y}_f = \text{med}\{y_i : i \in \mathcal{A}_f\} = \frac{O(E_f)}{n_f}$$

5.4.2 Elements of a Self-Categorisation Theory-Compatible Utility Function

Consistent with standard assumptions in microeconomic theory, the utility function is taken to be strictly increasing, strictly quasi-concave, and continuously differentiable throughout its domain. In addition to these general assumptions, the principles and hypotheses outlined in Section 5.2 suggest the following additional design principles:

1. The utility function should represent social-identity salience dynamically, capturing variation in agents' identification along a continuous spectrum of group salience. At one extreme lies the *individual* pole, where self-categorisation operates solely at the personal

level: self-perception is based on similarities and differences with other in-group members, emphasising uniqueness. At the opposite extreme lies the *social* pole, where the self becomes fully depersonalised, and the agent perceives themself as an **interchangeable** member of the in-group. Most individuals occupy an intermediate position on this continuum; moving toward the social pole implies a gradual acceptance of shared values.

2. The utility function should embody the *functional antagonism* between levels of self-categorisation postulated by Turner [1987]. Excluding the superordinate level from consideration, this principle implies that greater salience of the personal level reduces the salience of the intermediate (social) level—and vice versa. In other words, individual and group frames of self-definition compete for cognitive dominance.

3. At the individual pole of self-categorisation, the utility function should reduce to the standard economic form.

4. At the social pole, by contrast, the utility function should depend on the characteristics of the group prototype—the shared mental image of the individual who best embodies the group's defining traits, the group's shared result, the group norms and the firm culture.

To operationalise these requirements within the model, the next step is to formalise how identity salience itself is endogenously generated—namely, through the meta-contrast ratio.

5.4.3 Formalising the Meta-Contrast Ratio

The meta-contrast ratio (MCR) provides a quantitative measure of how well a firm functions as a psychologically coherent group. The idea is simple: a collection of agents forms a salient category to the extent that its members are more similar to each other than they are to outsiders. Accordingly, the MCR compares average inter-group dissimilarity to average intra-group dissimilarity.

Let each agent i be characterised by a feature vector $\boldsymbol{\theta_i} = (\theta_i^1, \theta_i^2, \ldots, \theta_i^d)$, where θ_i^j represents agent i's position on a stereotypical dimension j. Consider a focal firm f (the in-group) and its complement f' (the out-group). The MCR quantifies the extent to which f is perceived as a distinct category relative to f'. Intra-group dissimilarities are computed over unique unordered pairs to avoid double-counting:

$$\mathrm{MCR}_f = \frac{\dfrac{1}{|f|\,|f'|}\displaystyle\sum_{a\in f}\sum_{b\in f'} d(\boldsymbol{\theta_a}, \boldsymbol{\theta_b})}{\dfrac{2}{|f|(|f|-1)}\displaystyle\sum_{\substack{a,c\in f\\ a<c}} d(\boldsymbol{\theta_a}, \boldsymbol{\theta_c})} \tag{5.1}$$

Here, $d(\boldsymbol{\theta_a}, \boldsymbol{\theta_b})$ measures the multidimensional dissimilarity between two agents, defined as a weighted Euclidean distance:

$$d(\boldsymbol{\theta_a}, \boldsymbol{\theta_b}) = \sqrt{\sum_{j=1}^{d} \omega^j \left|\theta_a^j - \theta_b^j\right|^2} \tag{5.2}$$

where ω^j denotes the feature weight of dimension j, and is subject to $\sum_k \omega^k = 1$. The weights allow dimensions of comparison to contribute unequally to perceived similarity depending on their *fit* to category specifications. A higher value of MCR_f indicates that members of f are, on average, more internally homogeneous and more distinct from the rest of the population. In SCT, such configurations make the in-group more likely to dominate the agent's self-categorisation process.

Given this firm-level measure of contextual contrast, identity salience can now be specified as a smooth function of MCR_f.

5.4.4 A Function for Social Identity Salience

Drawing on the principles of SCT, the salience of a social identity corresponds to the degree to which an agent cognitively shifts from an individual to a group-based frame of self-categorisation. To model the continuous nature of this transition, identity salience for agent i is specified as the logistic activation

$$\phi_i(\mathrm{MCR}_f) = \frac{1}{1 + \mathrm{e}^{-\kappa_i(\mathrm{MCR}_f - M_{0,i})}} \tag{5.3}$$

where $M_{0,i}$ denotes the contrast threshold at which in-/out-group differentiation becomes perceptually meaningful, and κ_i controls the steepness of the response around that threshold. Agents with larger κ_i react more sharply to changes in MCR_f, whereas higher values of $M_{0,i}$ imply that more contextual contrast is required to trigger the same salience level.

Throughout the simulations, the identity-salience parameters $(\kappa_i, M_{0,i})$ are calibrated jointly and—unless otherwise specified—are taken to be homogeneous across agents (*i.e.*, $\kappa_i \equiv \kappa$ and $M_{0,i} \equiv M_0$). This allows the analysis to isolate the endogenous impact of contextual contrast as

encoded by MCR_f, while systematically exploring global shifts in the shared contrast regime (M_0, κ).

Selection of the salience parameters

From the logistic form in Eq. (5.3), it follows directly that

$$\kappa = \ln\left(\frac{\phi(\mathrm{MCR}_f)}{1-\phi(\mathrm{MCR}_f)}\right)\frac{1}{\mathrm{MCR}_f - M_0} \tag{5.4}$$

In words: once the threshold M_0 and a target salience level ϕ at a reference contrast MCR_f are chosen, the steepness parameter κ is pinned down.

A natural benchmark is $\mathrm{MCR}_f = 1$, which corresponds to parity between inter- and intra-group differences. In such cases, the group does not perceptually stand out and salience should remain minimal ($\phi \approx 0$). Let $\varepsilon > 0$ denote this minimal salience such that $\phi(1) = \varepsilon$. Substituting into Eq. (5.4) yields

$$\kappa = \ln\left(\frac{\varepsilon}{1-\varepsilon}\right)\frac{1}{1-M_0} \tag{5.5}$$

This calibration ensures that salience remains latent in low-contrast environments while rising predictably as the firm becomes more distinct. Having specified how identity salience responds to social contrast, we can now incorporate this mechanism into a utility function that governs effort choice and mobility.

5.4.5 *A Utility Formulation for Firm Identity and Effort Contribution*

With identity salience now defined as a continuous function of contextual contrast, the next step is to embed this cognitive shift into agents' optimisation behaviour. To do so, the job-mobility and effort-contribution model introduced by Axtell et al. [1999] and later refined in Axtell [2016a] is extended, using it as a rigorous testbed for a salience-weighted, SCT-compatible utility. The goal is not simply to *add* prosociality, but to *leverage* identity salience as a contextual and continuous moderator of behaviour: when salience is low, choices collapse to standard income-leisure trade-offs; when salience is high, agents increasingly evaluate outcomes through the group prototype and norm. This raises a natural set of questions: how do such micro-level shifts propagate to **i)** intra-firm effort, **ii)** inter-firm mobility, and **iii)** macro-economic regularities? Addressing these questions clarifies when identity functions as "motivational capital" in decentralised labour markets while remaining consistent with bottom-up macro emergence.

The decision to extend Axtell's framework is deliberate and rests on three considerations:

1. **Continuity with earlier chapters.** Chapters 3-4 analysed teamwork as a *closed* unit in which agents choose how much to contribute given a fixed team. By contrast, Axtell's economy is an *open* system: agents may migrate across firms or found new ones through limited, networked search. Firms emerge, grow and disappear endogenously, and macro-structure arises from decentralised micro decisions. Embedding identity into this setting preserves the micro-to-macro narrative while adding the missing outer loop of labour mobility. This yields a unified story: when group boundaries are permeable, identity shapes both *where* agents work (mobility) and *how* they work (effort).

2. **A principled "stress test" for identity in a data-concordant ABM.** The baseline model in Axtell [2016b] reproduces several empirical regularities—Zipf/Pareto-like firm sizes, realistic job-to-job flows, entry/exit dynamics, wage and tenure distributions—*without* exogenous shocks: agents remain in disequilibrium while the aggregate approaches a steady state. These empirical moments provide a stringent benchmark. If introducing identity salience preserves these macro patterns, then identity can be integrated into microfoundations without breaking known macro facts; if it alters them, the model identifies boundary conditions under which identity reshapes market structure.

3. **Microfoundations aligned with earlier reward structures.** The original model [Axtell et al., 1999, Axtell, 2016a] features **i)** team production with increasing returns, **ii)** equal sharing of output, and **iii)** agents maximising a log-utility over income and leisure while considering a small set of outside options. These ingredients map cleanly onto the earlier chapters: Chapter 3 used CES-style aggregators and income-leisure trade-offs; Chapter 4 made effort sequential and information-local. Replacing the purely individualistic log-utility with a salience-weighted one that blends individual and group-prototypical evaluations yields a tractable formulation that **i)** reduces to the standard economic case when salience is low, **ii)** shifts weight toward group-norm-congruent evaluation as salience rises, and **iii)** allows identity to influence both effort and mobility. This provides a natural, endogenous bridge from individual to team reasoning—guided by the MCR and its salience transform—without imposing discrete, exogenous "identity types".

The key elements of a formal model integrating team production, effort choice, and endogenous labour mobility with identity-driven preferences are as follows.

Traits and social categorisation.

Each agent i is characterised by a set of *stereotypical dimensions* $\boldsymbol{\theta}_i \in \mathbb{R}^d$ that capture socially relevant attributes capable of inducing categorisation.[2]

As defined in Eq. (5.1), the Meta-Contrast Ratio (MCR) is computed with respect to a focal firm f (the agent's in-group) and its complement $f' = \mathcal{A} \setminus f$. Identity salience for agent i is then given by Eq. (5.3):

$$\phi_i(\mathrm{MCR}_f) = \frac{1}{1 + \mathrm{e}^{-\kappa(\mathrm{MCR}_f - M_0)}} \in (0, 1)$$

where (κ, M_0) are the universal salience parameters defined in Section 5.4.4.

Prototype, norms and values.

At the social pole of self-categorisation ($\phi \approx 1$), the agent's evaluation is framed entirely through the group prototype—the shared mental image of the individual who best embodies the defining traits of the firm. This prototype is an *idealised standard* derived from the distribution of attributes within the group rather than from any specific member. In the model, this social reference is operationalised through the firm's shared outcome and its behavioural norms. The firm's shared outcome is proxied by the median income $\widetilde{y}_f$, which provides a robust measure of the typical economic outcome of the group and is therefore appropriate for capturing *perceptions of common fate or shared destiny*. Conversely, the *contribution norm* perceived by agent i is represented by the mean effort of all *other members*, $\bar{e}_{-i}$, computed in a Leave-One-Out (LOO) fashion,[3] reflecting the fact that norms arise from observed behavioural regularities.

The firm's *normative preference* parameter $\widetilde{\alpha}_f$—already defined above as the within-firm median of the α_i—serves as a summary of the group's characteristic valuation of income relative to leisure. In this sense, $\widetilde{\alpha}$ encapsulates the firm's *values* in the broader, normative sense: the standards and orientations that guide appropriate effort provision. All these quantities are updated each timestep as firms evolve and memberships change, providing the dynamic inputs through which the social component of utility in Eq. (5.7) is continuously recalibrated.

[2] These dimensions may include **socio-demographic factors** (*e.g.*, gender, age, ethnicity), **organisational or professional attributes** (*e.g.*, hierarchical level, occupational specialisation, tenure), **value-based or attitudinal orientations** (*e.g.*, alignment with the firm's mission or cooperative norms), or **ideological and cultural dimensions** (*e.g.*, position along a political or moral continuum). Any such dimension can, under appropriate contextual conditions, trigger self-other comparisons and thus activate social identity processes within the firm.

[3] $\bar{e}_{-i} = \sum_{\substack{j \in f \\ j \neq i}} \frac{e_j}{|f|-1}$

Salience-weighted perceived utility.

In Axtell et al. [1999], agents derive individual utility from income y_i and leisure $1 - e_i$ according to the standard Cobb-Douglas form:

$$U_i^{\text{ind}}(y_i, e_i) = y_i^{\alpha_i}(1 - e_i)^{1-\alpha_i} \tag{5.6}$$

This formulation is retained as the *microeconomic core* of the model. The present framework does not replace it but *extends* it by embedding social-identity parameters derived from SCT into the same functional architecture. Specifically, the agent's perceived utility is modelled as a salience-weighted combination of two evaluative frames: the *individual frame*, identical to Axtell's specification, and the *social frame*, in which outcomes are assessed through the firm's prototypical values and norms. Substituting the Cobb-Douglas form into this salience-weighted structure yields:

$$U_i^{\text{SCT}} = \left[y_i^{\alpha_i}(1 - e_i)^{1-\alpha_i}\right]^{1-\phi_i} \left[\widetilde{y}_f^{\,\widetilde{\alpha}_f}\,(1 - \bar{e}_{-i})^{1-\widetilde{\alpha}_f}\right]^{\phi_i} \tag{5.7}$$

where $\bar{e}_{-i}$ is the Leave-One-Out (LOO) mean effort of the other members of firm f, $\widetilde{y}_f$ is the firm's median income (the proxy for common fate), and $\widetilde{\alpha}_f$ is the firm-level normative preference parameter. Maximising (5.7) with respect to e_i yields the agent's optimal contribution. As $\phi_i \to 0$, behaviour converges to the baseline income–leisure trade-off; as $\phi_i \to 1$, the decision is increasingly framed from the perspective of the firm prototype.

The careful reader may have noticed that, in this model, ϕ_i alters the *valuation frame* but does not *prescribe* behaviour. This can be seen by considering an agent with a highly salient social identity, such that only the social frame of the utility function is active:

$$\widetilde{y}_f^{\,\widetilde{\alpha}_f}\,(1 - \bar{e}_{-i})^{1-\widetilde{\alpha}_f}$$

When optimising effort e_i against this expression, the agent's choice enters the utility only through its effect on $\widetilde{y}_f$—and not through leisure. This is consistent with an extreme reading of SCT: when group identification is very strong, the individual no longer evaluates leisure in personal terms. Instead, the agent adopts a fully depersonalised evaluative frame in which only the *prototypical* leisure of the group, given by $(1 - \bar{e}_{-i})$, is taken into account. However, this does *not* force agent i to adhere to the group's contribution norm: there is no penalty for deviation. What this formulation captures is therefore a shift in the evaluative frame, but not behavioural homogenisation. Beyond its consistency with SCT, the use of the LOO statistic avoids circularity in the forthcoming comparative statics in Section 5.5.

Notice how the model presented above focuses exclusively on the cognitive shift in evaluative perspective, leaving the formalisation of prescriptive norm alignment to future work. This is a deliberate modelling choice. In its classical formulation [Turner, 1985, 1987, Turner and Haslam, 2001], Self-Categorisation Theory (SCT) specifies the *cognitive* mechanisms through which category salience reshapes perception and outcome evaluation, but it does not posit intrinsic prescriptive forces or behavioural obligations. Norm adherence and conformity—central in Social Identity Theory (SIT) [Tajfel and Turner, 2004] and in later work on group norms—arise only at subsequent stages of social influence.

Taking logs, Eq. (5.7) can be rewritten as:

$$\begin{aligned} V_i^{\mathrm{SCT}} = (1-\phi_i)\left[\alpha_i \ln y_i + (1-\alpha_i)\ln(1-e_i)\right] \\ + \phi_i\left[\widetilde{\alpha}_f \ln \widetilde{y}_f + (1-\widetilde{\alpha}_f)\ln(1-\bar{e}_{-i})\right] \end{aligned} \tag{5.8}$$

Expressing the SCT-compatible utility as the exponential of log-utilities implies that agents integrate their individual and social valuations *multiplicatively* rather than additively, that is, as a geometric mean. This formulation avoids additive inflation—where one evaluative frame could dominate merely due to scale—and ensures that both components contribute proportionally to perceived well-being.

Behaviourally, the logarithmic specification aligns with the postulates of SCT, according to which individuals integrate self- and group-based evaluations as functionally antagonist **cognitive frames** rather than as separable sources of payoff. In this sense, $U_i^{\mathrm{ind}}(y_i, e_i)$ and $U_i^{\mathrm{soc}}(\widetilde{y}_f, \bar{e}_{-i})$ represent **perceptual valuations of outcomes under distinct frames of self-categorisation**. Taking their weighted average in log space and then exponentiating corresponds to forming an *expected frame value*—the subjective evaluation an agent makes given the salience ϕ_i of the social identity.

This structural choice therefore embeds self-categorisation directly into the optimisation process, allowing the salience parameter ϕ_i—determined by the MCR—to modulate behaviour **endogenously** rather than being imposed exogenously.

Job mobility.

Agents periodically reassess their organisational membership. With an exogenous activation probability $\pi_m \in (0,1)$, agent i considers a finite set of candidate organisations:

- the current firm $\varphi(i)$;

- a local information set $\mathcal{N}_i \subseteq \mathcal{F}$ of observed firms (the agent's social/professional neighbourhood);
- the option of founding a single-person start-up.

This defines the candidate set

$$\mathcal{C}_i \stackrel{\text{def}}{=} \{\varphi(i)\} \cup \mathcal{N}_i \cup \{\text{start-up}\}$$

For each candidate $f' \in \mathcal{C}_i$, membership is evaluated through a one-step counterfactual. Let $f \stackrel{\text{def}}{=} \varphi(i)$ denote the current firm. The counterfactual environment in which agent i joins f' is constructed as follows:

1. **Firm size.** The headcount of f' is incremented by one if and only if $f' \neq f$:

$$n_{f'}^{+} = n_{f'} + \mathbf{1}[f' \neq f]$$

   while the sizes of all other firms remain unchanged.

2. **Identity frame.** When evaluating the current firm f, identity salience is given by the endogenous value $\phi_i(\text{MCR}_f)$ defined in Eq. (5.3). By contrast, prospective firms are assessed through the individual frame: for all $f' \neq f$ it is assumed that

$$\phi_i(\text{MCR}_{f'}) = 0$$

Again, this design choice is theoretically grounded in SCT: when evaluating prospective employers, agents are not yet members of those organisations, and therefore the necessary conditions for a psychological group are not met. Prospective workplaces are thus assessed through an individual frame of reference ($\phi_i = 0$), whereas identification and social framing emerge only after entry, once membership and interaction take place [Turner, 1987, Ashforth and Mael, 1989, Turner and Haslam, 2001].

Given the counterfactual parameters for each $f' \in \mathcal{C}_i$, the model computes the corresponding salience-weighted utility $U_i^{\text{SCT}}(f')$ and the implied best-response effort $e_i^\star(f')$. The agent then updates membership according to the myopic rule

$$\varphi(i) \leftarrow \arg\max_{f' \in \mathcal{C}_i} \mathbb{E}\left[U_i^{\text{SCT}}(f')\right] \tag{5.9}$$

where the expectation operator captures any stochastic elements in payoffs or information.

Before turning to the model's testable implications, Table 5.1 summarises how the key psychosocial constructs introduced above map onto their economic representations at each pole of the self-categorisation continuum. Figure 5.1 provides an overview of how identity salience, utility, effort, and mobility jointly shape the dynamic evolution of firms. The key contribution of the proposed framework lies in closing the feedback loop between organisational composition, social categorisation, and economic behaviour. Unlike existing identity-based economic models—where identity categories and their motivational effects are typically imposed exogenously—the present model endogenises identity salience through the MCR. Changes in firm composition alter the degree of inter- and intra-group contrast, which in turn reshapes agents' evaluative frames, effort provision, and mobility decisions. These behavioural adjustments then feed back into the composition of firms, generating a dynamic co-evolution between social identity and economic structure.

Table 5.1: **Elements of a Self-Categorisation Theory-compatible utility function.** Mapping between psychosocial constructs and their economic representations in the model at each pole of the self-categorisation continuum.

Level of self-categorisation	**Component**	**Formalisation**
Individual pole	Individual income	$y_i = \frac{O(E_f)}{n_f}$
	Individual leisure	$1 - e_i$
	Preference for income	α_i
Social pole	Shared fate	$\widetilde{y}_f = \frac{O(E_f)}{n_f}$
	Social effort norm	$\bar{e}_{-i} = \frac{1}{\lvert f \rvert - 1} \sum_{\substack{j \in f \\ j \neq i}} e_j$
	Firm values	$\widetilde{\alpha}_f = \text{median}\{\alpha_j : j \in f\}$

With the microeconomic structure in hand, let us now turn to the model's testable implications: a set of comparative statics and welfare results that pin down the direction and strength of the proposed mechanisms.

5.5 Comparative Statics

The analysis in this section maintains the equal-sharing rule introduced above: for any firm f with membership $\mathcal{A}_f$ and size $n_f = |\mathcal{A}_f|$, total output $O(E_f)$ is divided equally among its

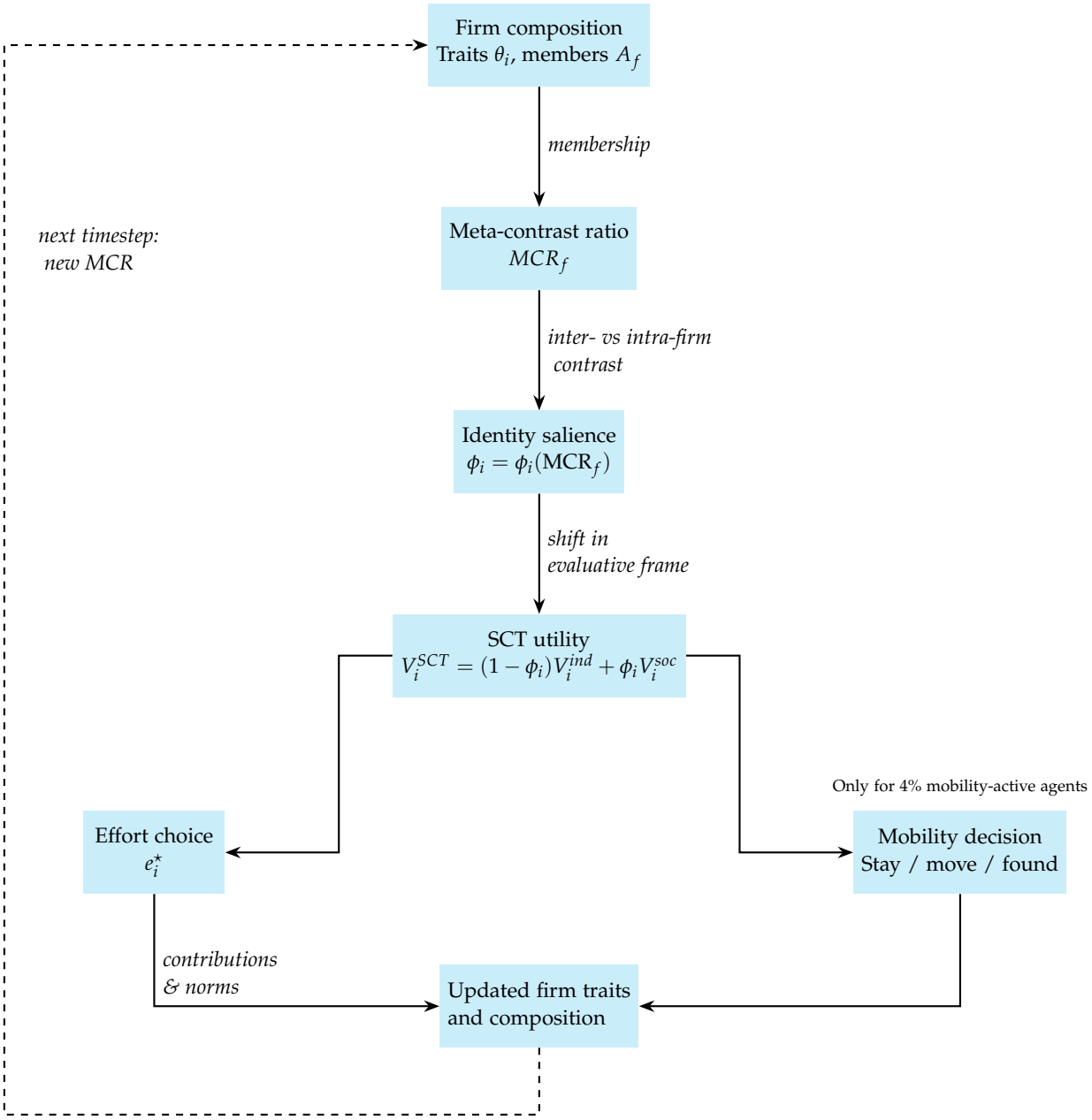


Figure 5.1: Computational pipeline of the proposed identity-enriched extension of Axtell's firm-formation model. Firm composition determines the meta-contrast ratio (MCR_f), which endogenously generates workers' identity salience ϕ_i. This salience modulates the SCT-compatible utility function, shaping both effort provision and mobility decisions. These individual choices subsequently alter firm composition and social contrast, closing the feedback loop between organisational structure, identity formation, and economic behaviour over time.

members, so that $y_i = O(E_f)/n_f$ for all $i \in \mathcal{A}_f$. Since within-firm incomes are identical under this assumption, the firm's median income satisfies $\widetilde{y}_f = O(E_f)/n_f$.

Because the comparative statics focus on a single firm, notation is simplified by dropping the firm subscript whenever no confusion arises. Accordingly,

$$E,\ n,\ y = \frac{O(E)}{n},\ \widetilde{y} = \frac{O(E)}{n},\ \text{and}\ \widetilde{\alpha}$$

is written instead of E_f, n_f, y_i, $\widetilde{y}_f$, and $\widetilde{\alpha}_f$. All results in this section are derived for an arbitrary firm f under this relabelling. Finally, all derivatives are computed given within-timestep quantities, as per the timing convention introduced in Section 5.4.1.

5.5.1 *Behavioural comparative statics I: Identity salience*

To connect the formal analysis with the motivating example introduced at the beginning of the chapter, consider Anna's situation again. Imagine that she is not the only one who feels detached from her company, but that this sentiment is shared by many senior employees who experienced the merger. In response, the firm decides to implement a series of initiatives aimed at rebuilding a shared sense of identity and cohesion among its staff.

The simplest version of Akerlof and Kranton [2000a]'s principal-agent model predicts that stronger identification with the firm should invariably translate into higher individual performance. The framework developed here partly challenges this assumption: it identifies parameter regimes under which a highly cohesive workforce may, *in fact*, perform worse.

Formally, recall the SCT-compatible log-utility in Eq. (5.8), which can be written as a convex combination of two evaluative frames,

$$V_i^{\mathrm{SCT}} = (1-\phi_i)\, V_i^{\mathrm{ind}}(y_i, e_i) + \phi_i\, V_i^{\mathrm{soc}}(\widetilde{y}, \bar{e}_{-i})$$

where V_i^{ind} and V_i^{soc} denote the individual and social log-utilities, respectively. Let

$$F(e_i, \phi_i) = \frac{\partial V_i^{\mathrm{SCT}}}{\partial e_i}$$

The optimal contribution $e_i^\star$ is characterised by the first-order condition $F(e_i^\star, \phi_i) = 0$. By the Implicit Function Theorem,[4] if $\frac{\partial F}{\partial e_i}(e_i^\star, \phi_i) \neq 0$, then $e_i^\star$ can be written locally as a differentiable

[4] *Implicit Function Theorem.* Let $F : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable in a neighbourhood of $(x^\star, t^\star)$ with $F(x^\star, t^\star) = 0$ and $\partial F/\partial x(x^\star, t^\star) \neq 0$. Then there exist open neighbourhoods and a continuously differentiable function $x(t)$ such that $F(x(t), t) = 0$ and

$$x'(t^\star) \;=\; -\left.\frac{\partial F/\partial t}{\partial F/\partial x}\right|_{(x^\star, t^\star)}$$

In the present context, taking $x = e_i$ and $t \in \{\phi_i,\, n\}$ yields the expressions in Eqs. (5.10) and (5.11).

function of identity salience, $e_i^\star(\phi_i)$, with

$$\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i} = -\left.\frac{\partial F/\partial \phi_i}{\partial F/\partial e_i}\right|_{(e_i^\star,\phi_i)} \tag{5.10}$$

The expression for $\frac{\partial F}{\partial e_i}$ follows directly from differentiating V_i^{SCT} with respect to e_i. Details of this derivation and its decomposition into production and leisure terms are presented in Appendix D.1. As shown there, the denominator of Eq. (5.10) is strictly negative for all $\beta \in (0, 4]$, which implies that the sign of $\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i}$ coincides with that of $\frac{\partial F}{\partial \phi_i}$.

Differentiating F with respect to ϕ_i yields:

$$\frac{\partial F}{\partial \phi_i} = \underbrace{(\widetilde{\alpha} - \alpha_i)\frac{O'(E)}{O(E)}}_{\text{(I) Marginal productivity / valuation shift}} + \underbrace{\frac{1-\alpha_i}{1-e_i}}_{\text{(II) Relief of the marginal cost of effort}}$$

Term (I) captures a *valuation shift* in the marginal benefit of effort. As shown in Appendix D.1, the marginal benefit component of $F(e_i, \phi_i)$ is:

$$\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right]\frac{O'(E)}{O(E)}$$

so increasing ϕ_i shifts the weight from the individual preference α_i towards the firm's normative parameter $\widetilde{\alpha}$. Since $\frac{O'(E)}{O(E)} > 0$, the sign of this shift is determined by $(\widetilde{\alpha} - \alpha_i)$: if $\widetilde{\alpha} > \alpha_i$, a rise in ϕ_i increases the marginal benefit of effort; if $\widetilde{\alpha} < \alpha_i$, it reduces it. In other words, in a group whose norm prescribes a stronger orientation towards income than the agent's own preference, moving along the identity spectrum towards the social pole raises the perceived payoff from exerting additional effort. The factor $\frac{O'(E)}{O(E)}$ scales this change by how responsive output, and hence per-capita income, is to total effort E.

Term (II) reflects the *relief of the marginal cost of effort*. In $F(e_i, \phi_i)$, the term

$$(1-\phi_i)\frac{1-\alpha_i}{1-e_i}$$

enters negatively, representing the individual cost of effort. Differentiating this negative contribution with respect to ϕ_i yields $\frac{1-\alpha_i}{1-e_i} > 0$. In other words, as ϕ_i increases—and the social frame gains weight—the cost term is scaled down: losing an extra unit of personal leisure counts less in the overall evaluation. Behaviourally, shifting towards the social pole partially alleviates the perceived marginal disutility of effort, because own effort is evaluated less through the individual income-leisure trade-off and more through the group frame.

For subsequent results, it is convenient to note that

$$\frac{O'(E)}{O(E)} = \frac{a + b\beta E^{\beta-1}}{aE + bE^{\beta}} \in \left[\tfrac{1}{E},\, \tfrac{\beta}{E}\right]$$

so the proportional responsiveness of output to total effort is always positive, and becomes smaller as total effort E grows (typically in larger firms). Appendix D.1 provides a decomposition of $\frac{O'(E)}{O(E)}$ in elasticity form and shows how these bounds arise from the technological parameters.

Proposition 5.1 (Identity salience and effort). *Assume $\beta \in (0, 4]$ and an interior solution $e_i^\star \in (0, 1)$. Then $\frac{\partial F}{\partial e_i} < 0$, and*

$$\operatorname{sign}\left(\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i}\right) = \operatorname{sign}\left(\frac{\partial F}{\partial \phi_i}\right) = \operatorname{sign}\left((\widetilde{\alpha} - \alpha_i)\,\frac{O'(E)}{O(E)} + \frac{1-\alpha_i}{1-e_i}\right)$$

In particular:

(i) If $\widetilde{\alpha} \geq \alpha_i$, then $\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i} > 0$: greater identity salience increases individual effort.

(ii) If $\widetilde{\alpha} < \alpha_i$, then $\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i} < 0$ iff

$$\boxed{(\alpha_i - \widetilde{\alpha})\,\frac{O'(E)}{O(E)} > \frac{1-\alpha_i}{1-e_i}}$$

Since $\frac{O'(E)}{O(E)} \in [1/E,\, \beta/E]$, the valuation channel weakens as total effort E increases. Assume $E = n \cdot \bar{e}$. In larger firms—where E is higher—increases in total effort translate into smaller marginal gains in output. Hence, an additional unit of individual effort generates a smaller marginal increase in per-capita income. In that case, the identity-backfire phenomenon of Case *(ii)* becomes less likely.

Corollary 5.2 (Backfire condition). *Under the same assumptions as Proposition 5.1, social identity **reduces** optimal effort whenever*

$$\boxed{\frac{\mathrm{d}e_i^\star}{\mathrm{d}\phi_i} < 0 \quad \Longleftrightarrow \quad (\alpha_i - \widetilde{\alpha})\,\frac{O'(E)}{O(E)} > \frac{1-\alpha_i}{1-e_i}}$$

*That is, **identity backfires** when the individual's excess valuation of income over the firm's social norm—scaled by the responsiveness of income to total effort, $\frac{O'(E)}{O(E)}$—exceeds the perceived relief in the marginal cost of effort, $\frac{1-\alpha_i}{1-e_i}$.*

Returning to Anna's case, this means that when her team is both small and lax, making firm identity more salient can actually drag her down: the more she sees herself through the eyes of a complacent group, the less worthwhile extra effort looks, and this loss in motivation dominates the relief in how costly effort feels, so her optimal effort falls.

5.5.2 *Behavioural comparative statics II: Firm size*

A second comparative-static exercise examines how changes in firm size affect a worker's optimal effort, *ceteris paribus*. In practical terms, can the company's expansion affect Anna's performance? And if it can, how? This section unravels three *incentive channels* through which a change in firm size affects Anna's individual predisposition to work: a production, an identitarian and a leisure one; and discusses how they interact to inhibit or encourage her engagement.

Note from Eq. (5.1) that identity salience for each employee may depend on firm size through the MCR. The analysis below therefore considers both the direct technological effect of size and the indirect effect operating through salience. Throughout, the standard decomposition whereby **total effort equals the product of mean effort and headcount**, $E = \bar{e}\, n$, is adopted.

Let $F(e_i, n) = \frac{\partial V_i^{\text{SCT}}}{\partial e_i}$. The optimal contribution $e_i^\star$ satisfies the first-order condition $F(e_i^\star, n) = 0$. By the Implicit Function Theorem, if $\frac{\partial F}{\partial e_i}(e_i^\star, n) \neq 0$, then $e_i^\star$ can locally be expressed as a differentiable function of firm size, $e_i^\star(n)$, and

$$\frac{\mathrm{d}e_i^\star}{\mathrm{d}n} = -\left.\frac{\partial F/\partial n}{\partial F/\partial e_i}\right|_{(e_i^\star, n)} \tag{5.11}$$

From Section 5.5.1, under the technology and preference assumptions used here the denominator $\frac{\partial F}{\partial e_i}(e_i^\star, n)$ is strictly negative for $\beta \in (0, 4]$.[5] Hence,

$$\operatorname{sign}\left(\frac{\mathrm{d}e_i^\star}{\mathrm{d}n}\right) \;=\; \operatorname{sign}\left(\frac{\partial F}{\partial n}\right)$$

To characterise $\frac{\partial F}{\partial n}$, define

$$R(n) \overset{\text{def}}{=} \frac{O'(E)}{O(E)} \qquad E = \bar{e}\, n$$

which measures how responsive output, and thus per-capita income, is to changes in total effort when firm size is n. For the production technology $O(E) = aE + bE^\beta$, Appendix D.2 shows that $R(n) > 0$ and $R'(n) < 0$ for all n and all $\beta \in (0, 4]$.

Allowing identity salience ϕ_i to depend on n through the MCR, differentiating F with respect to n yields

$$\frac{\partial F}{\partial n} = \underbrace{R'(n)\Big[(1-\phi_i)\,\alpha_i + \phi_i \widetilde{\alpha}\Big]}_{\text{(I) Marginal productivity shift}} + \underbrace{\frac{\partial \phi_i}{\partial n}\, R(n)\,(\widetilde{\alpha} - \alpha_i)}_{\text{(II) Salience-benefit channel}} + \underbrace{\frac{\partial \phi_i}{\partial n}\,\frac{1-\alpha_i}{1-e_i}}_{\text{(III) Salience-cost relief}} \tag{5.12}$$

[5] See Appendix D.1 for further details.

Term (I) captures the purely technological effect of size: as n grows (holding average effort fixed), $R(n) = O'(E)/O(E)$ falls. Since the preference weight $[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}]$ is positive, this implies that increasing firm size reduces the marginal benefit of effort: an additional unit of individual effort translates into a smaller proportional increase in per-capita income in larger firms. Terms (II) and (III) represent the additional impact of firm size through identity salience: changes in $\phi_i(n)$ alter both the marginal benefit of effort (via the shift from α_i to $\widetilde{\alpha}$) and the marginal cost of effort (via the weight on lost leisure). The full derivation of Eq. (5.12), together with a detailed sign analysis of its components, is provided in Appendix D.2. Here, the main implications are summarised.

Proposition 5.3 (Firm size and effort). *Assume an interior solution $e_i^\star \in (0,1)$ and $\beta \in (0,4]$. Then $\frac{\partial F}{\partial e_i} < 0$, and*

$$\operatorname{sign}\left(\frac{\mathrm{d}e_i^\star}{\mathrm{d}n}\right) \;=\; \operatorname{sign}\left(\frac{\partial F}{\partial n}\right) \;=\; \operatorname{sign}\Big(R'(n)\big[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\big] + \frac{\partial \phi_i}{\partial n}\Big[R(n)(\widetilde{\alpha}-\alpha_i) + \tfrac{1-\alpha_i}{1-e_i}\Big]\Big)$$

In particular:

(i) If identity salience is locally size-invariant $\big(\frac{\partial \phi_i}{\partial n} = 0\big)$, then $\frac{\mathrm{d}e_i^\star}{\mathrm{d}n} < 0$ for all feasible n: larger firms reduce optimal effort.

(ii) If $\frac{\partial \phi_i}{\partial n} \neq 0$, a necessary and sufficient condition for $\frac{\mathrm{d}e_i^\star}{\mathrm{d}n} > 0$ is

$$\boxed{\frac{\partial \phi_i}{\partial n}\Big[R(n)(\widetilde{\alpha}-\alpha_i) + \tfrac{1-\alpha_i}{1-e_i}\Big] \;>\; -R'(n)\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right]}$$

Case *(ii)* shows that firm growth may induce higher optimal levels through two antagonist channels:

Corollary 5.4 (Salience-mediated reversals). *Under the assumptions of Proposition 5.3, firm growth may* increase *optimal individual effort through two distinct channels.*

(i) ***Positive salience channel.*** *If $\frac{\partial \phi_i}{\partial n} > 0$ and $\widetilde{\alpha} \geq \alpha_i$, then a sufficient condition for $\frac{\mathrm{d}e_i^\star}{\mathrm{d}n} > 0$ is*

$$\frac{\partial \phi_i}{\partial n} \;>\; \frac{-R'(n)\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right]}{R(n)\,(\widetilde{\alpha}-\alpha_i) + \frac{1-\alpha_i}{1-e_i}}.$$

(ii) ***Norm disengagement channel.*** *If $\frac{\partial \phi_i}{\partial n} < 0$ and $\widetilde{\alpha} < \alpha_i$, firm growth may also increase optimal effort provided that*

$$-\,\tfrac{\partial \phi_i}{\partial n}\,R(n)\,(\alpha_i - \widetilde{\alpha}) \text{ is sufficiently large relative to } -\,R'(n)\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right].$$

In this case, expanding firm size weakens identification with a low-effort social norm, thereby restoring the individual's stronger valuation of income and raising effort incentives.

The three channels outlined—productivity, cost relief, and identity—will reappear systematically in the next welfare analysis, now determining not effort incentives but perceived well-being.

5.5.3 *Welfare and identity salience*

As detailed in Section 5.4, the model at hand grants two distinct levels of agency or decision-making: at the *closed-system* level, agents decide how much effort they are willing to put in teamwork; at the *open-system* level, they decide where they want to work. In Anna's case, this means choosing both how much to contribute to her current firm and whether to stay or move.

Understanding how her micro-based utility is affected by identity salience and firm size is thus key to grasping the emergent dynamics at the meso- and macro-levels. The focus of this section is on how changes in identity salience ϕ_i permeate into the utility function: will Anna's welfare improve or deteriorate as she identifies more strongly with the group? The forthcoming results show that the answer is relative to her individual status: *it depends on whether she stands to gain or lose from joining the group*. Firm size is held constant throughout.

By the Envelope Theorem,[6]

$$\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}\phi_i} \;=\; U_i^{\mathrm{SCT}\star}\,\frac{\partial V_i^{\mathrm{SCT}}}{\partial \phi_i}\bigg|_{e_i=e_i^\star} \tag{5.13}$$

From Eq. (5.8), the partial derivative of the log-utility with respect to salience is

$$\frac{\partial V_i^{\mathrm{SCT}}}{\partial \phi_i} \;=\; \big[\widetilde{\alpha}\ln\widetilde{y} + (1-\widetilde{\alpha})\ln(1-\bar{e}_{-i})\big] \;-\; \big[\alpha_i \ln y_i + (1-\alpha_i)\ln(1-e_i^\star)\big] \tag{5.14}$$

where $\widetilde{y}$ is the firm's median income, $\bar{e}_{-i}$ the leave-one-out mean effort of co-workers, and y_i the individual's income.

Since $U_i^{\mathrm{SCT}\star} > 0$, the sign of $\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}\phi_i}$ coincides with the sign of $\frac{\partial V_i^{\mathrm{SCT}}}{\partial \phi_i}$. This expression shows that the change in welfare an employee experiences *while working at her optimal effort* varies with

[6] *Envelope Theorem.* Let $f(x,t)$ be continuously differentiable and let $x^\star(t)$ solve $\max_x f(x,t)$ with an interior optimum. Define the associated value function $g(t) = f(x^\star(t),t)$. Then, under regularity conditions,

$$\frac{\mathrm{d}g}{\mathrm{d}t} \;=\; \frac{\partial f}{\partial t}\Big|_{x=x^\star(t)}$$

In the present model, f corresponds to the log-utility V_i^{SCT}, while $g(t) = V_i^{\mathrm{SCT}}(e_i^\star(t),t)$ and $U_i^{\mathrm{SCT}} = \exp(V_i^{\mathrm{SCT}})$. Applying the chain rule yields

$$\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}t} = U_i^{\mathrm{SCT}\star}\,\frac{\partial V_i^{\mathrm{SCT}}}{\partial t}\Big|_{x=x^\star(t)}$$

In the present context, taking $x = e_i$ and $t \in \{\phi_i, n\}$ yields the expressions in Eqs. (5.13) and (5.15).

the salience of her social identity and is driven by the difference between two evaluations: **i)** the log-utility of the *social* frame (a fully depersonalised evaluation through the group prototype) and **ii)** the log-utility of the *individual* frame.

If the social frame assigns a higher value than the individual frame, then $\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}\phi_i} > 0$: stronger identity increases perceived welfare. The welfare effect of identity is therefore exactly the gap between social and individual log-utilities.

Proposition 5.5 (Identity salience and welfare). *Fix firm size and the leave-one-out mean effort, and assume an interior optimum $e_i^\star \in (0,1)$. Then*

$$\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}\phi_i} \;=\; U_i^{SCT\star}\Big(\underbrace{\big[\widetilde{\alpha}\ln\widetilde{y}+(1-\widetilde{\alpha})\ln(1-\bar{e}_{-i})\big]}_{\textit{social frame}}-\underbrace{\big[\alpha_i\ln y_i+(1-\alpha_i)\ln(1-e_i^\star)\big]}_{\textit{individual frame}}\Big)$$

In particular:

(i) $\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}\phi_i} > 0$ *if and only if the social frame yields a higher log-utility than the individual frame at* $(e_i^\star, \bar{e}_{-i})$.

(ii) $\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}\phi_i} = 0$ *if and only if both frames evaluate the current situation equally.*

(iii) $\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}\phi_i} < 0$ *if and only if the individual frame dominates the social one at* $(e_i^\star, \bar{e}_{-i})$.

Corollary 5.6 (Equal income-sharing). *Under equal sharing, individual and median incomes coincide, $y_i \equiv \widetilde{y} \equiv \bar{y}$, and Eq. (5.14) becomes*

$$\frac{\partial V_i^{SCT}}{\partial \phi_i} = (\widetilde{\alpha}-\alpha_i)\,\ln\bar{y} + (1-\widetilde{\alpha})\ln(1-\bar{e}_{-i}) - (1-\alpha_i)\ln(1-e_i^\star)$$

Thus the welfare effect of higher salience decomposes into an income component, proportional to $(\widetilde{\alpha}-\alpha_i)\ln\bar{y}$, and a leisure component, driven by the comparison between group norm $\bar{e}_{-i}$ and individual effort $e_i^\star$. In the special case where the agent's preference for income is aligned with the group prototype ($\alpha_i = \widetilde{\alpha}$), the income term disappears and variation in welfare arises solely through the leisure channel.

In the context of the simulated economy, these welfare effects translate into a distinct asymmetry at the model's meso-scale. Consistent with the mobility specification in Section 5.6.3, social identity is inactive when agents evaluate potential alternative firms: outside options are assessed purely through the individual frame. Identity-based welfare gains, therefore, act solely as a *retention mechanism*. Stronger identification raises the perceived utility of staying within the current firm but does not increase the attractiveness of others, manifesting itself as endogenous talent retention rather than as recruitment.

5.5.4 *Welfare and firm size*

To conclude this section, it is useful to examine how variations in firm size affect workers' welfare. As in the previous subsection, this exercise helps to decouple workers' willingness to work from their well-being as firm size increases. For Anna, this amounts to asking whether a larger employer necessarily makes her better off. The careful reader will suspect by now that there is no universal answer and that, again, size, motives, and identity intertwine and affect incentives through different channels. Nonetheless, the forthcoming proposition outlines some conditions under which the identitarian channel dominates.

This analysis completes the macroscopic intuition of the model and anticipates the simulation results that follow by providing an answer to the question: *under what conditions do large firms attract more agents, and when does individual welfare decline with firm growth?*

As before, the Envelope Theorem is applied to obtain the derivative of welfare *at the worker's optimal effort* $e_i^\star$. Firm output is given by $O(E) = aE + bE^\beta$ as detailed at the beginning of this section, with total effort $E = \bar{e}\, n$. Thus, per-capita income under equal sharing is

$$\bar{y} = \frac{O(E)}{n} = a\bar{e} + b\,\bar{e}^\beta\, n^{\beta-1}$$

Applying the Envelope Theorem yields

$$\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}n} = U_i^{\mathrm{SCT}\star}\,\frac{\partial V_i^{\mathrm{SCT}}}{\partial n}\bigg|_{e_i=e_i^\star} \tag{5.15}$$

Substituting the log-utility formulation in Eq. (5.8) gives

$$\begin{aligned}\frac{\partial V_i^{\mathrm{SCT}}}{\partial n} = &\underbrace{\Big[(1-\phi_i)\,\alpha_i + \phi_i\,\widetilde{\alpha}\Big]\,\frac{(\beta-1)\,b\,\bar{e}^\beta\, n^{\beta-2}}{a\bar{e} + b\,\bar{e}^\beta\, n^{\beta-1}}}_{\text{(I) Technological / individual frame}} \\ &- \underbrace{\phi_i(1-\widetilde{\alpha})\,\frac{e_i^\star - \bar{e}}{(1-\bar{e}_{-i})(n-1)^2}}_{\text{(II) Leisure / social frame}} \\ &+ \underbrace{(V_i^{\mathrm{soc}} - V_i^{\mathrm{ind}})\,\frac{\partial \phi_i}{\partial n}}_{\text{(III) Identity-salience channel}}\end{aligned} \tag{5.16}$$

Here $V_i^{\mathrm{ind}} = \alpha_i \ln y_i + (1-\alpha_i)\ln(1-e_i^\star)$ and $V_i^{\mathrm{soc}} = \widetilde{\alpha}\ln\bar{y} + (1-\widetilde{\alpha})\ln(1-\bar{e}_{-i})$ denote the individual and social log-utilities, respectively, and $\bar{e}_{-i}$ is the leave-one-out mean effort of the worker's colleagues. The detailed derivation of the components in Eq. (5.16) is provided in Appendix D.3

These terms admit a natural interpretation. Term (I) is the *technological channel*: changes in firm size modify per-capita income through the production function, and this change is weighted by the salience-weighted valuation of income $\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right]$. Term (II) is the *leisure channel*: as n grows, the leave-one-out mean effort $\bar{e}_{-i}$ shifts relative to $e_i^\star$, altering the perceived value of foregone leisure under the social frame. Term (III) is the *identity-salience channel*: changes in n affect the salience $\phi_i(n)$, thereby re-weighting the individual and social frames according to the difference $V_i^{\text{soc}} - V_i^{\text{ind}}$

Under the parameter regime considered here, terms (I) and (II) decay in magnitude with firm size. **For sufficiently large** n the following result yields:

Proposition 5.7 (Firm size and welfare). *Fix $\bar{e}$ and $e_i^\star$, and let ϕ_i depend on firm size n. Then welfare at the individual optimum varies with firm size as in Eq. (5.16). In particular, because $U_i^{SCT\star} > 0$, the sign of $\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}n}$ coincides with the sign of $\frac{\partial V_i^{SCT}}{\partial n}$, and:*

(i) If $V_i^{soc} > V_i^{ind}$ and $\frac{\partial \phi_i}{\partial n} < 0$, then $\frac{\mathrm{d}U_i^{SCT\star}}{\mathrm{d}n} < 0$: welfare declines with firm growth as identity salience fades

(ii) If $V_i^{soc} < V_i^{ind}$, the sign reverses: lower salience can mitigate welfare losses in oversized firms

Comparative statics: main insights.

- Changes in the feeling of belonging—*i.e.* in the salience of social identity—and in headcount affect workers' incentives to *perform* at their current firm and to *jobhunt* through three distinct *incentive channels* shaping their **perceived income benefits**, **social identity** and **perceived foregone leisure**.
- An increased sense of belonging to a firm—more formally, an increased salience of firm identity—does **not** always lead to higher performance. Specifically, picture a highly work-oriented worker. If she joins a *small* firm with a *lax* work preference—relative to hers—, making firm identity more salient can actually drag her down: the more she sees herself through the eyes of a complacent group, the less worthwhile extra effort looks, and this loss in motivation dominates the relief in how costly effort feels, so her optimal effort falls.
- Belonging to larger firms typically **inhibits** individual contributions, unless firm growth either strengthens identification with a high-effort organisational identity or weakens attachment to a low-effort social normparticularly in highly responsive production environments.
- An increased salience of firm identity does **not** necessarily lead to higher well-being; it depends on whether the worker stands to gain or lose well-being from identifying with the group.
- Changes in perceived well-being as headcount rises are dominated by the identitarian channel when firms are sufficiently large. If the worker *gains* utility from identifying with the firm, and social identity *dilutes* as the company grows, then firm growth will hurt her well-being. And vice versa: if the worker *loses* utility from identifying with the group, a weaker sense of shared identity in larger firms mitigates the welfare loss and makes expansion comparatively easier to bear.
- Overall, the effects of identity on incentives and well-being are *intrinsically asymmetric*: they reinforce advantages for aligned workers and aggravate disadvantages for misaligned ones.

With these theoretical results in place, we now turn to the computational model that tests, extends, and reveals the dynamic consequences of these comparative statics.

5.6 Simulation Design and Implementation

This section specifies the simulation setup: state variables, information structure, mobility routine, and model parameterisation.[7]

5.6.1 One-Dimensional Traits

For simulation purposes, one-dimensional traits ($d = 1$) are adopted with $\omega^1 = 1$ in Eq. (5.2): each agent is characterised by a single scalar trait $\theta_i \in [0, 1]$, representing an abstract latent dimension of social comparison capable of inducing categorisation (*e.g.*, ideological, cultural, or professional position). Distances reduce to $d(\theta_a, \theta_b) = |\theta_a - \theta_b|$.

5.6.2 Information, Network and Candidate Sets

Agents are embedded in a static social/professional random graph $\mathcal{G}$, where the connectivity parameter ν is sampled from a uniform distribution, $\nu_i \sim \mathcal{U}(2, 6)$. In each mobility assessment, agent i forms a candidate set consisting of their current firm, the firms observed through neighbours in $\mathcal{G}$, and the start-up option. Firms outside this neighbourhood are unobserved, preserving the limited-information assumption.

5.6.3 Local Evaluation Heuristic

Operationally, the mobility rule in Eq. (5.9) is implemented via a local evaluation heuristic. For each candidate firm $f' \in \mathcal{C}_i$, the simulator constructs the counterfactual environment described in Section 5.4.5, computes the associated salience-weighted utility $U_i^{\mathrm{SCT}}(f')$ and the corresponding best-response effort $e_i^\star(f')$, and selects the maximising option.

5.6.4 Calibration and Parameters

Identity salience is governed by the logistic specification introduced in Section 5.4.4, where ε denotes the minimal salience at the neutrality benchmark $\mathrm{MCR}_f = 1$, M_0 is the contrast threshold at which salience reaches its inflexion point, and κ controls the steepness of the transition. Once (ε, M_0) are chosen, κ follows from the logistic mapping.

[7] The corresponding multi-agent framework is publicly available and can be consulted at the project's GitHub repository: https://github.com/aloaberasturi/IdentityFields.

In the simulations, the minimal salience is fixed at $\varepsilon = 0.1$, ensuring that firms with negligible contrast remain effectively non-salient.[8] The contrast threshold is varied over the range $M_0 \in [1.3, 1.6]$, so that $(M_0 - 1)$ directly measures the additional contrast required for group identity to become perceptually meaningful.

Inter-group distance via sampling

The inter-group term in the numerator of Eq. (5.1) is computed by sampling a shared subset of out-group agents for each firm. Let f denote the focal firm (size $|f|$) and $f' = \mathcal{A} \setminus f$ its complement (size $|f'|$). For a fixed sampling budget $s = \min\{k_{\text{out}}, |f'|\}$, draw a subset $S \subseteq f'$ of size s uniformly at random (with replacement) and define

$$\widehat{D}_{\text{inter}}(f, f') \;=\; \frac{1}{|f|\, s} \sum_{a \in f} \sum_{b \in S} d(\theta_a, \theta_b) \tag{5.17}$$

This yields an unbiased estimator of the mean inter-firm distance with computational cost $\mathcal{O}(|f|\, s)$ per firm. Using a shared out-group sample S for all $a \in f$ keeps variance low for moderate sampling budgets.[9]

Intra-group distance

The intra-group term is the mean pairwise distance among members of the firm. For one-dimensional traits, the exact $\mathcal{O}(|f| \log |f|)$ formula is used:

$$\overline{D}_{\text{intra}}(f) \;=\; \frac{2}{|f|(|f|-1)} \sum_{k=1}^{|f|} \big(2k - |f| - 1\big)\, \theta_{(k)}$$

where $\theta_{(k)}$ is the k-th sorted trait value within the firm. This identity matches exactly the average of all absolute pairwise differences without constructing an $|f| \times |f|$ matrix.

Sampled MCR

Combining the two components gives the estimated meta-contrast ratio for firm f:

$$\widehat{\text{MCR}}_f \;=\; \frac{\widehat{D}_{\text{inter}}(f, f')}{\overline{D}_{\text{intra}}(f)} \tag{5.18}$$

When $|f| < 2$ or when f coincides with the entire population (so $|f'| = 0$), $\widehat{\text{MCR}}(f)$ is undefined, as either the within-firm or the between-firm distance cannot be computed. In these cases, the MCR update is skipped.

[8] Values in the range $\varepsilon \in [0.01, 0.1]$ were tested and produced qualitatively identical results. The choice $\varepsilon = 0.1$ improves readability in the visualisations.

[9] $k_{\text{out}} = 200$ by default, clipped to $|f'|$ when smaller.

Base Parameters

The remaining parameters follow the configuration reported in Axtell [2016b], summarised in Table 5.2. The number of agents is set to $|\mathcal{A}| = 1 \times 10^5$, which preserves the macro-level regularities of the Axtell economy—such as Zipf's law for firm sizes—while keeping run-time manageable.

Table 5.2: **Simulation parameters, distributions, and initialisation.**

Symbol	Meaning	Calibration / Distribution
(a, b, β)	Micro-production parameters	$(\mathcal{U}(0,1), \mathcal{U}(0,\frac{1}{2}), \mathcal{U}(\frac{3}{4}, \frac{5}{4}))$
Δ	Time/energy endowment	1
α_i	Income–leisure preference	$\mathcal{U}(0,1)$
ν_i	Node connectivity (degree)	$\mathcal{U}(2,6)$
π_m	Mobility activation rate	0.04 (4% of agents per timestep)
T	Timesteps simulated	1000

In each timestep, a random 4% of agents are activated for potential mobility. All activated agents simultaneously evaluate alternative firms according to the mobility rule in Eq. (5.9). The full timestep-by-timestep dynamics are made explicit in the pseudocode featured in Appendix D.5. The interested reader may refer there for details on the activation step, the evaluation of alternative membership options, the best-response computation, and the subsequent update of firm states.

5.7 Experiments

Applying the parameter specification from Section 5.6.4, this section now presents the main results. Recall that the feature space is one-dimensional and that the salience parameters $[M_0, \kappa]$ are homogeneous across agents and calibrated through their functional relation in Eq. (5.4). The economy runs with $|\mathcal{A}| = 10^5$ agents over $T = 1000$ timesteps, with a mobility activation rate $\pi_m = 0.04$ per timestep.

The simulations address the overarching question posed in this chapter: *What micro- and macro-patterns emerge in an economy of job mobility and effort provision à la* Axtell [2016a] *once social categorisation is embedded in the utility function?* The core objective is to assess whether identity salience is *compatible* with the economic backbone of the baseline model (*e.g.*, Zipf/Pareto-like

firm size distributions, realistic mobility and tenure) or whether it *reconfigures* market structure. Two complementary scenarios that differ in the source of social identity are considered:

Scenario A (endogenous identity from traits). Agents are endowed with latent traits $\theta_i \in [0,1]$ (Section 5.6.1). Firm-level MCR then *emerges* from the within- and between-firm distributions of these traits via Eq. (5.1), which in turn maps to identity salience through Eq. (5.3). Behaviour (effort, mobility) responds to this endogenous salience. To probe how social proximity structures categorisation, two trait topologies are analysed:

1. **i.i.d. traits:** $\theta_i \sim \mathcal{N}(\mu, \sigma^2)$, truncated to $[0,1]$. Results for this i.i.d. baseline are reported in Appendix D.7 to avoid clutter. The main text focuses on the network-homophily case below.

2. **Network-homophily traits:** starting from $\theta_i \sim \mathcal{N}(\mu, \sigma^2)$, a local averaging operator is applied over the social/professional graph $\mathcal{G}$ (Section 5.6.2) to induce similarity by homophily. Concretely, with an initial vector $\boldsymbol{\theta}^{(0)}$, traits evolve for $t = 1, \ldots, 10$ iterations as

$$\boldsymbol{\theta}^{(t)} \;=\; 0.95\,\boldsymbol{\theta}^{(t-1)} \;+\; 0.05\,W\,\boldsymbol{\theta}^{(t-1)}$$

   where W is an adjacency matrix normalised by rows (each row sums to one), with a self-loop added so that each node retains part of its own trait during propagation. The operator preserves isolated nodes and converges to a locally smoothed field that is more similar along edges of $\mathcal{G}$.

Scenario B (exogenous identity via firm-level MCR). To isolate the effect of *contrast* itself—independently of the stochastic structure of trait distributions—each firm is instead assigned an *exogenous* MCR drawn from a log-normal distribution. These sampled MCR values are mapped to identity salience through Eq. (5.3). Traits play no role in this scenario; identity operates as an exogenous firm-level contextual variable. Results are reported in Appendix D.8.

The reason to choose such log-normal specification is that it produces a positively skewed distribution of contrast intensities: most firms exhibit weak or moderate identity salience, while a minority display very high contrast and thus strong corporate identity. This functional form is used purely as a modelling device: in the absence of empirical data on the distribution of identity salience across firms, the log-normal provides a plausible solution to generate heterogeneous and strictly positive contrasts.

5.7.1 *Experiment: Endogenous Identity with Network-Homophily Traits*

This section reports the simulation results for Scenario A with network-homophily traits. The remaining configurations are deferred to Appendix D.7 (Scenario A with i.i.d. traits) and Appendix D.8 (Scenario B with exogenous identity via sampled MCR) for clarity of exposition. The homophily case is chosen to be in the main text because it **i)** most closely approximates realistic professional networks, and **ii)** activates all feedback loops in the model—endogenous identity, evolving firm norms and prototypes, and networked mobility.

Taken together, the appendices provide a parsimonious identification of mechanisms: Appendix A removes homophily to isolate the "pure" endogenous-MCR channel, while Appendix B fixes salience exogenously to identify identity's causal impact cleanly. Let's now turn to the main scenario.

Hypotheses and Links to Theory

The comparative statics and welfare analyses from Section 5.5 suggest three falsifiable patterns:

(H1) **Baseline and the social limit.** The model admits two analytically tractable boundary regimes:

- When $\phi \approx 0$, the baseline model of Axtell [2016b] is recovered. Agents behave as standard utility-maximisers, optimising a conventional leisure-income trade-off.
- When $\phi \approx 1$, agents are fully aligned with their firm and evaluate costs and welfare in terms of the firm's prototype. In this degenerate social case, assuming the firm's production technology follows $O(E) = aE + bE^{\beta}$, the derivative of social utility with respect to firm size is:

$$\frac{\partial V_i^{\text{soc}}}{\partial n} = \widetilde{\alpha}\,\frac{(\beta-1)\,b\,\bar{e}^{\beta}\,N^{\beta-2}}{a\,\bar{e}+b\,\bar{e}^{\beta}\,N^{\beta-1}} - (1-\widetilde{\alpha})\,\frac{e_i^{\star}-\bar{e}}{(1-\bar{e}_{-i})(N-1)^2}$$

  Further clarification on this derivation is provided in Appendix D.3. As discussed in Appendix D.6, for $\beta < 1$ social utility decreases with firm size, making large firms unattractive; whereas for $\beta > 1$, the marginal contribution to social welfare remains positive as $n \to \infty$. In such cases, after sufficient simulation time, agents are expected to cluster in firms with the highest β values.

In summary, under a fully individualistic regime, the model collapses to the baseline. Conversely, under a fully social regime driven by increasing returns to scale, the economy should converge towards a monopolistic configuration.

(H2) **Effort-salience slope.** Proposition 5.1 predicts $\frac{de_i^\star}{d\phi_i} > 0$ when $\widetilde{\alpha} \geq \alpha_i$, and the backfire condition (Corollary 5.2) when $\widetilde{\alpha} < \alpha_i$ and the valuation channel dominates. Empirically, as ϕ_i increases, the distribution of individual effort should traverse distinct regimes. Approaching the backfire region, contributions progressively decline and mass accumulates at low effort. This is expected in lax, moderately sized firms where E is still low enough that $\frac{O'(E)}{O(E)}$ remains high and the marginal-productivity channel is not yet attenuated. Beyond that regime, higher ϕ_i raises effort for agents with $\widetilde{\alpha} \geq \alpha_i$. Consequently, a bimodal distribution should emerge at the population level, provided there is sufficient heterogeneity in α_i relative to $\widetilde{\alpha}$.

(H3) **Effort-size slope.** Proposition 5.3 implies $\frac{de_i^\star}{dn} < 0$ holding ϕ_i fixed, but a possible reversal if firm growth either strengthens identification with a high-effort organisational identity or weakens attachment to a low-effort social normparticularly in highly responsive production environments (Corollary 5.4). Hence, a weakly negative effort-size gradient is expected.

In the following sections, the methods and simulations used to test these hypotheses are described, along with the results obtained.

Methods

Having laid out the calibration procedure in previous sections (Section 5.6.4) and the homophily-based trait sampling process, we now detail the experimental parameter grid employed and the outcomes measured.

Clustering Thresholds. In Scenario A, the clustering threshold is swept:

$$M_0 \in \{1.3, 1.4, 1.5, 1.6\}$$

For each M_0, the steepness κ is pinned down by Eq. (5.5) with $\varepsilon = 0.1$ at $MCR = 1$.

In general, a clustering threshold

$$M_0 = 1 + c$$

implies that agents require a $c\%$ difference between in-group and out-group members for their social identity to become perceptually salient—that is, for $\phi_i = 0.5$.

Outcome Measures. For each configuration, micro-, meso- and macro-level statistics are recorded:

- **Micro (agent-level):** effort e_i, social identity salience ϕ_i and income y_i, distributional summaries (means, quantiles) and conditional plots $\mathbb{E}[e_i \mid \phi_i]$.
- **Meso (firm-level):** firm size n_f, firm tenure, firm total output, within-firm free-riding tendencies, median $\widetilde{\alpha}$, estimated $\widehat{\text{MCR}_f}$, within-firm salience $\bar{\phi}$, and conditional plots $\mathbb{E}[e_i \mid n_f]$
- **Macro (economy-level):** firm size and firm tenure distribution (Zipf/Pareto fit) and mobility flows.

For a given firm f with production parameters (a, b, β), firm output follows $O(E) = aE + bE^{\beta}$ and per-capita income **under equal sharing** is $y_i = O(E)/n_f$ (Section 5.4). When applicable, empirical power-law fits using the methods in Clauset et al. [2009] are reported.

Results and Discussion

This section presents the key simulation results for Scenario A, characterised by network-homophily traits. It begins with the baseline and degenerate cases, which illustrate the model's boundaries and expected dynamics under limiting conditions. Next, it examines how the salience parameters influence the emergence of social identity, shifting agents' evaluative frames from the individual to the social pole. The remainder of the section tests hypotheses H2-H3 and reports additional findings on the model's macro-level structure. For each experimental condition, 10 independent simulations (1,000 timesteps each) were ran with different random seeds. Metrics were exported to TensorBoard and aggregated across runs by timestep.

Baseline and degenerate cases (H1). To study the limiting configurations, a fixed, arbitrary threshold $M_0 = 1.5$ was chosen and Eq. (5.5) was used to calibrate the salience parameter κ as follows:

- In the baseline case ($\phi_i(\text{MCR}_f) = 0$), social identity is never triggered. In practice, ϕ is set to $\phi(100) = \varepsilon$.

- In the degenerate social case ($\phi_i(\text{MCR}_f) = 1$), social identity is triggered to its maximum. In practice, ϕ is set to $\phi(1) = 1 - \varepsilon$ so it takes minimal contrast between the in- and out-group for the agents to switch into the fully social frame.

With $\varepsilon = 0.01$.[10] Once the pair $[\kappa, M_0]$ was calibrated for each configuration, the resulting dynamics of firm formation, growth and dissolution were examined. Figure 5.2 illustrates the baseline economy. Firm sizes increase during the early timesteps as random matching and market learning take place, but they soon stabilise around an average of fourteen agents per firm.

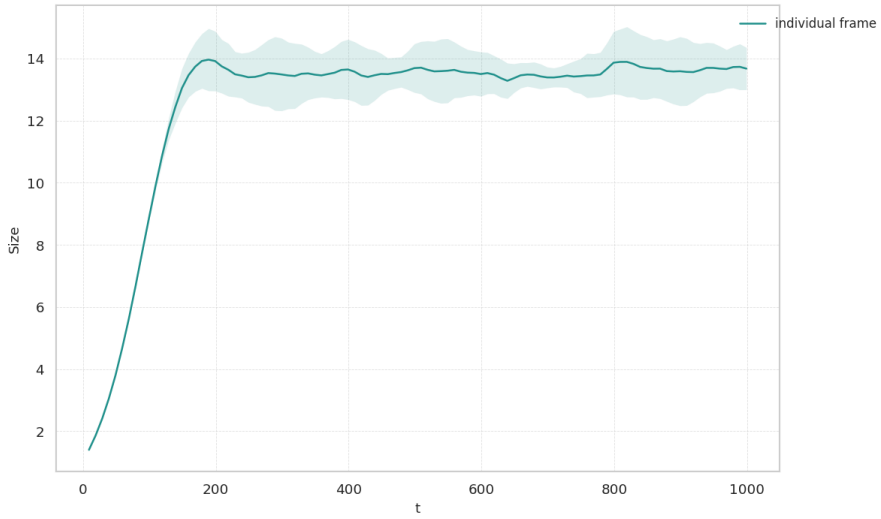


Figure 5.2: **Average firm size over time in the baseline economy** ($\phi \approx 0$). The shaded area represents one standard deviation across simulation runs.

When social salience increases to its upper bound, the system exhibits a qualitatively different trajectory. Firm sizes expand rapidly and eventually converge towards a few large, stable organisations, as can be inferred from the average firm size reported in Figure 5.3. This collapse results from agents' evaluative frames shifting from the individual to the social pole, reinforcing coordination and reducing mobility once they *fall* into a big company with high increasing returns to effort (*i.e.*, with high β), consistent with Appendix D.6.

Together, these contrasting patterns validate Hypothesis 1: while the baseline economy remains decentralised and dynamically balanced, *maximal salience induces a collapse of the economy into a few dominant firms*.

[10] Other values $\varepsilon \in [0.01, 0.1)$ were also tested to confirm robustness. We ultimately chose 0.01 for clearer plots.

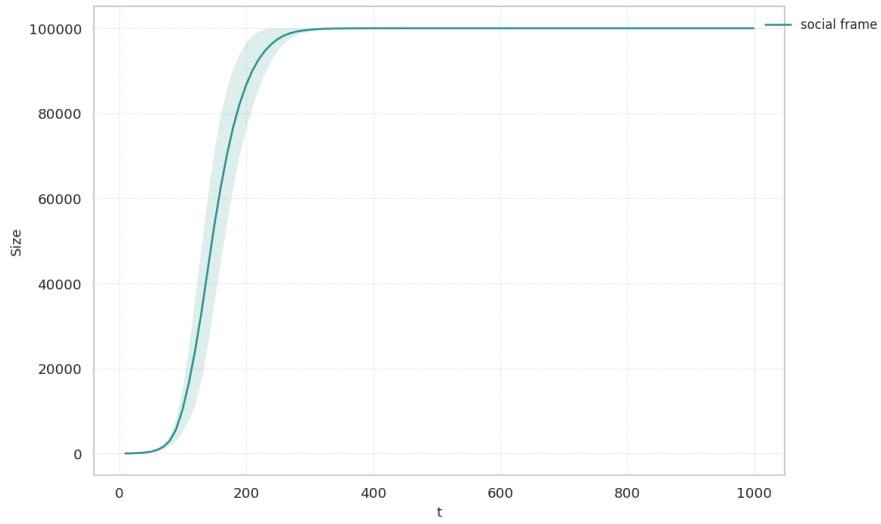


Figure 5.3: **Average firm size over time in the degenerate social economy** ($\phi \approx 1$).

Effects of M_0 on salience. Next, the model's sensitivity to variations in the clustering threshold M_0 was tested in the main scenario. As expected, higher values of M_0 attenuate the activation of social identity. Figure 5.4 displays this inverse relationship between M_0 and average social-identity salience.

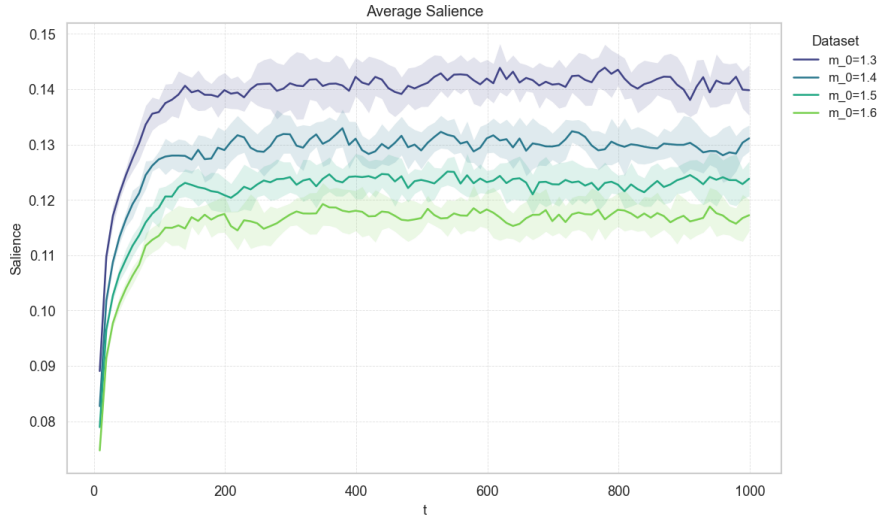


Figure 5.4: **Average social-identity salience for different clustering thresholds** M_0.

In words, *the higher the clustering threshold, the stronger the contrasts between in- and out-groups must be to elicit an equivalent salience response.*

Effect of social-identity salience on individual behaviour (H2). Proposition 5.1 implies heterogeneous effort responses to salience: $de_i^\star/d\phi > 0$ when $\widetilde{\alpha} \geq \alpha_i$, and potential declines when $\widetilde{\alpha} < \alpha_i$ and the valuation channel dominates (Corollary 5.2). Empirically, Fig. 5.5 shows that the typical effort distribution evolves through distinct regimes as ϕ rises: in the baseline ($\phi \approx 0$) the distribution is already bimodal (free-riding coexisting with a high-effort mode); at low-to-intermediate ϕ mass shifts towards low effort; and at higher ϕ the high-effort mode re-emerges.

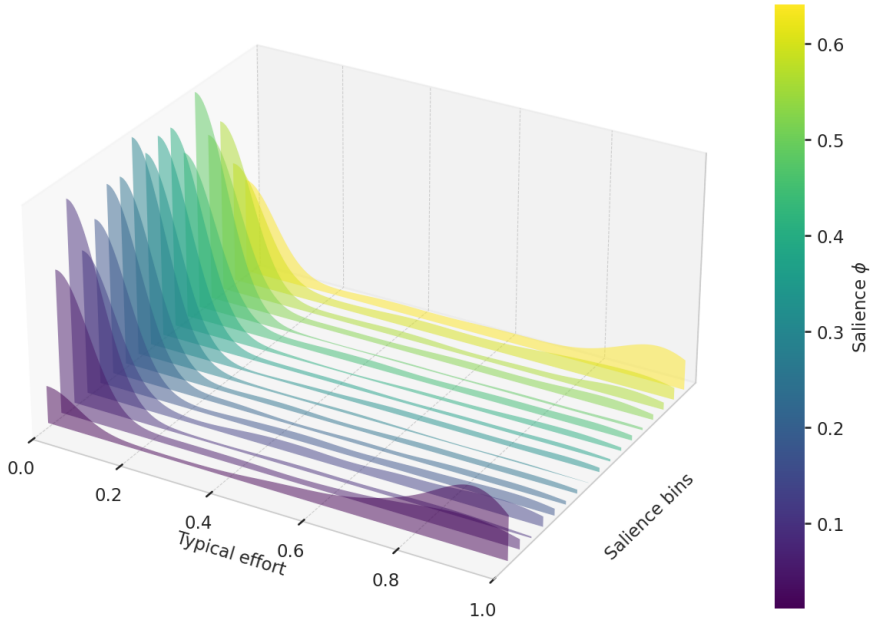
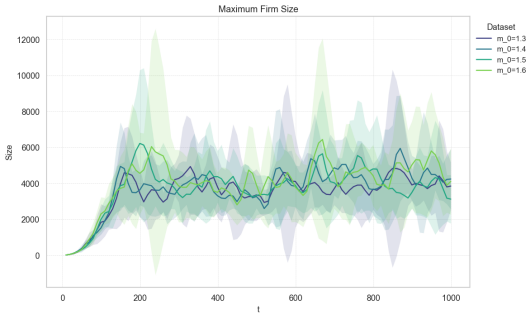


Figure 5.5: **Cross-sectional effort distributions by salience**. The colourmap represents the centres of equal-width bins.

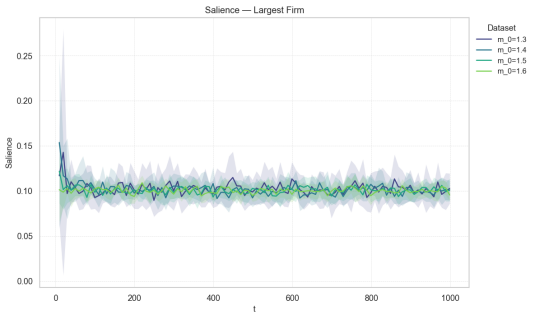
As ϕ increases, agents' decisions are re-weighted towards the firm frame. As explained in Proposition 5.1, in lax, moderately sized firms, the factor $\frac{O'(E)}{O(E)}$ can not yet attenuate the excess valuation of income for agents fulfilling $\alpha_i > \widetilde{\alpha}$, which depresses effort. For the case $\alpha_i < \widetilde{\alpha}$, an increase in social identity encourages effort. Because α_i is heterogeneous and firms differ in size, these behavioural regimes coexist. In practice, this produces an early *lazy* phase for low—but non-zero—values of ϕ, during which mass moves towards free-riding strategies, followed by a progressive shift of probability towards higher typical effort as social identity

becomes increasingly salient.

Effect of firm size on individual behaviour (H3). Figure 5.6b shows that, once the transient phase is over, the salience of social identity in the largest firm stabilises at a low level (around 0.10), with no upward drift. This pattern is robust across M_0 values. Figure 5.6a complements this observation: although the size of the largest firm fluctuates markedly over time—sometimes increasing by almost two orders of magnitude—the corresponding salience remains essentially flat. In other words, in this model, the MCR mechanism does *not* induce a positive relationship between firm size and ϕ.



(a) **Maximum firm size over time.** The size of the largest firm fluctuates substantially across timesteps, with large swings but no systematic upward or downward trend across M_0 settings.



(b) **Identity salience in the largest firm.** Salience rapidly converges to a stable level (around 0.10) and remains largely insensitive to firm size, even when the largest firm grows dramatically.

Figure 5.6: **Maximum firm size and identity salience over time.** The left panel shows the evolution of the largest firm's size; the right panel shows the corresponding identity salience. Despite large fluctuations in firm size, salience remains approximately constant.

Under these conditions, Proposition 5.3 implies that the technological channel dominates: the marginal return to individual effort falls with firm size. Figure 5.7 confirms this prediction empirically: median effort declines monotonically as firms grow.

Macro structure. Noting the probability of realising a firm of size n in the steady state as $\mathbb{P}(n)$, let us use the same parameterisation described in Axtell et al. [1999] of the discrete power-law distribution:

$$\mathbb{P}(n) = \left(\frac{n}{n_0} \right)^{-(1+\mu)}$$

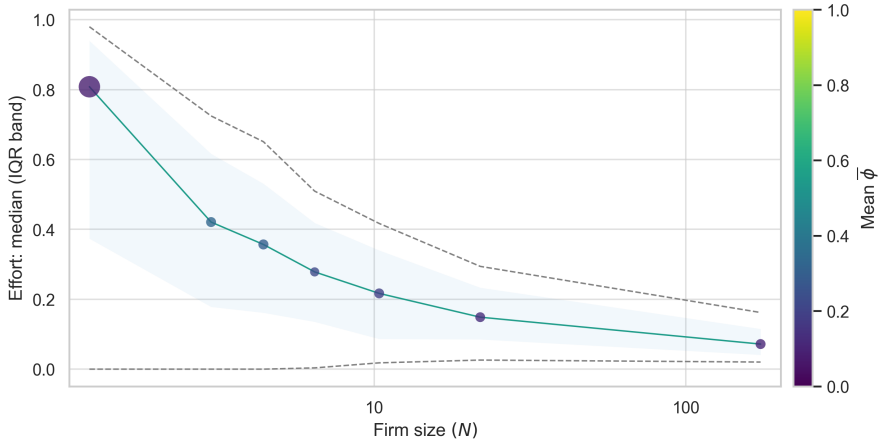


Figure 5.7: **Individual effort (median) versus firm size (n, logarithmic scale).** The solid line shows the median effort within each firm-size bin. The shaded blue band represents the interquartile range (IQR, $25-75\%$). The colour map encodes the mean social-identity salience for each bin, and dot size is proportional to the number of observations in each bin.

With n_0 being a minimal firm size. Data from the simulations are shown in doubly logarithmic (log-log) coordinates in Fig. 5.8. Across all M_0 values, the distribution of firm sizes exhibits a heavy-tailed pattern consistent with scaling behaviour. Fitting a discrete power-law model with $n_0 = 20$ yields the tail exponents μ and goodness-of-fit statistics reported in Table 5.3.

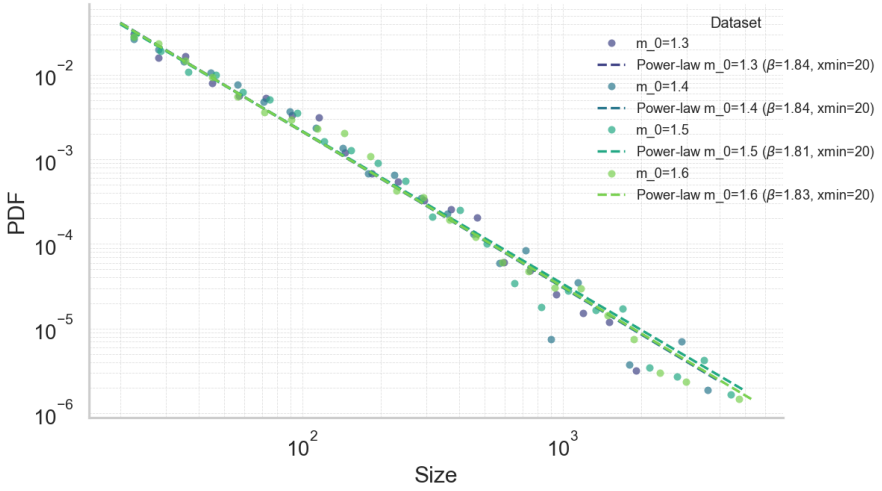


Figure 5.8: **Probability Mass Function (PMF) of firm size** by number of employees across different values of the clustering threshold, M_0. The distribution exhibits a heavy-tailed pattern consistent with scaling behaviour.

The Zipf law, which is the empirical macroeconomic regularity consistently observed in

real data and reproduced in Axtell [2016b], implies that the exponent μ should approach 1 in the asymptotic limit. In the present model, however, the estimated exponent remains remarkably stable around $\mu \simeq 0.88$ (corresponding to a slope of $1 + \mu \approx 1.88$). Kolmogorov-Smirnov distances range between 0.05 and 0.08 (see Table 5.2 for further detail), indicating that the power-law model provides an acceptable description of the upper tail. Likelihood-ratio tests show that the exponential alternative is decisively rejected ($R > 4$, $p < 0.001$). In contrast, the lognormal offers a marginally better fit ($R < 0$, $|R| \approx 3 - 4$), consistent with the empirical ambiguity typically observed between lognormal and power-law scaling in firm-size data (cf. [Clauset et al., 2009]).

Importantly, this deviation from the canonical Zipf exponent is not specific to the introduction of social-identity mechanisms: *the same value emerges in the baseline replication of Axtell's model when the simulated economy is limited to* 10^5 *agents*. The most parsimonious explanation is therefore a finite-size effect arising from the restricted support of the empirical distribution. With fewer agents, the upper tail is necessarily truncated, reducing the frequency of extreme realisations and slightly flattening the estimated slope. Such downward bias in power-law fits under limited sample size is a well-documented artefact in the statistical literature [Newman, 2005] and does not contradict the qualitative consistency of the results with Zipf's law. Taken together, these results confirm that the simulated economies reproduce a realistic heavy-tailed size distribution, with an exponent close to the empirical Zipf benchmark.

Table 5.3: **Pareto tail estimates for firm size across M_0.** The Kolmogorov-Smirnov statistic between the empirical and fitted power-law distributions is also reported.

M_0	Tail exponent	KS-distance
1.3	1.838	0.071
1.4	1.843	0.074
1.5	1.806	0.085
1.6	1.835	0.054

Overall, these results show that introducing self-categorisation and identity salience into a decentralised labour economy reproduces the empirical regularities of firm size while revealing new, interpretable regularities at the micro scale.

Key Takeaways

In summary, the simulations in Scenario A (endogenous identity with network-homophily traits) yield the following empirical patterns:

- **Identity salience drives a non-monotonic effort response.** As ϕ increases, the effort distribution first concentrates on low effort and later reactivates the high-effort mode. Agents in the $\widetilde{\alpha} \geq \alpha_i$ region increase their contributions at high ϕ, whereas agents with $\widetilde{\alpha} < \alpha_i$ reduce effort at low-to-intermediate ϕ, reproducing the predicted backfire regime.

- **Firm size does not increase identity salience.** The largest firms stabilise at low salience levels ($\phi \approx 0.1$), even when their size fluctuates by orders of magnitude. The MCR mechanism therefore does **not** generate a positive size-salience gradient.

- **Effort decreases as firms grow.** Because salience remains approximately flat with respect to firm size, the technological dilution channel dominates. Median effort declines monotonically with n, consistent with the theoretical prediction $\mathrm{d}e_i^\star/\mathrm{d}n < 0$ when $\partial\phi/\partial n \approx 0$.

- **The boundary regimes replicate the analytical limits.** When $\phi \approx 0$, the economy reproduces Axtell's baseline dynamics. When $\phi \approx 1$, workers cluster into a small number of large firms with high increasing returns (β), as predicted by the fully social limit of the model.

- **The macrostructure remains heavy-tailed.** Across all M_0, the firm-size distribution follows a stable heavy tail well-approximated by a discrete power law with exponent $\mu \simeq 0.88$. These results show that the introduction of identity salience preserves the macroeconomic backbone of the baseline model.



6 *Conclusion*

> Learning to craft rules that attract and encourage individuals who share norms of reciprocity and trustworthiness, or who learn them over time, is a fundamental skill needed in all democratic societies.
>
> Elinor Ostrom

6.1 Synthesis of Contributions *200*
6.2 Integrative Insights *204*
6.3 Limitations and Future Work *207*
6.4 Applied Directions and Transfer Opportunities *208*
6.5 Outstanding Questions and Concluding Remarks *210*

This thesis has proposed an approach towards cooperation by delineating key mechanisms—across different levels of analysis—that explain when cooperation should be expected (and when it should not) in settings where it is not compulsory.

Although computational at its core, the framework is grounded in several disciplines. To formalise the class of situations under study, it has drawn on game theory. This structural scaffold has made it possible to formally integrate "soft" constructs from work and social psychology into the body of strategic and economic models.

The presentation of the micro-theories at each chapter has followed an incremental progression: from a static conceptualisation of teamwork to dynamic models of open teams; from tangible and static alignment mechanisms, through intangible and temporary ones, to abstract mechanisms concerning the very perception of identity among team members.

Across these increasingly abstract models, each has, in its own minimalist iteration, circled back to the same essential question: *why do we cooperate?*

Alongside each theoretical model, a companion simulation environment has been built. These environments serve as mirrors: they test whether the theory holds when instantiated in interacting agents, and they reveal aggregate, evolving behaviours that no equation alone can predict. In the tradition of agent-based modelling and complex systems physics, the dissertation thus pairs every microscopic assumption with a computational world designed to show how simple rules can scale into the collective patterns we recognise as cooperation.

Together, these models outline a multi-level computational theory of cooperation grounded simultaneously in incentives, learning dynamics, and social identity. The remainder of this chapter distils the insights gained across these levels, highlights their unifying principles, examines the limitations of the adopted approaches, and reflects on the implications for the study of cooperative behaviour in both artificial and human systems.

6.1 Synthesis of Contributions

The three core chapters develop these ideas through distinct formal frameworks, each illuminating a different facet of cooperative behaviour. This section draws them together and highlights the specific contributions that each chapter makes to the broader theory developed

across the thesis. What follows is a concise synthesis of these contributions, organised around the mechanisms examined in Chapters 3, 4, and 5.

6.1.1 Chapter 3: Structural Alignment

Chapter 3 opens the analysis of alignment mechanisms by examining the most directly designable factors that foster cooperation in teamwork settings. Among the three theories developed in this dissertation, it is the one grounded in the most tangible variables—and it provides the structural foundation upon which the subsequent chapters build.

Conceptually, the chapter examines teamwork at the intersection of game theory and multi-agent learning. Whereas classical game-theoretic accounts model a team as a coalition with a shared utility and study cooperation through cooperative game theory, this chapter introduces a novel class of mixed-motive games—*Teamwork Games*—that capture settings in which **agents decide their voluntary contributions to a shared task**. Building on well-established theories of group performance from social psychology, the model incorporates variables often absent from game-theoretic treatments of teamwork, such as team roles, task type, expertise, interdependence, and evaluation functions.

Formally, the **Nash equilibria** of Teamwork Games are analytically characterised, showing both how they extend the equilibria of general public-good games and how the model's structural components systematically affect rational behaviour. Computationally, the chapter proposes a **Multi-Agent Multi-Armed Bandit (MA-MAB)** framework in which agents learn contribution strategies that approximate these equilibria, *CollabQuest*. Contributions are conceptualised as bandit arms, and agents adapt their behaviour using Boltzmann exploration. Even with simple independent learners—without opponent modelling—MA-MAB reliably converges to approximations of the analytically derived equilibria.

Qualitatively, the chapter conducts simulation analyses to examine how task type, evaluation functions, and agent expertise shape group performance. The findings align with established theories in social psychology, offering computational support for Steiner's hypotheses concerning *task interdependence*, *evaluative thresholds*, and *team composition*. The analysis is further extended to tasks with discontinuous evaluation functions, showing that the framework generalises to settings lacking closed-form solutions.

Taken together, these contributions establish Teamwork Games as a tractable structural model for analysing—and designing—voluntary teamwork.

6.1.2 *Chapter 4: Temporal Alignment*

Chapter 4 extends the analysis of alignment mechanisms by introducing temporality as a central dimension of teamwork. Whereas Chapter 3 examined cooperation in structurally defined and essentially static settings, this chapter argues that time fundamentally reshapes how incentives evolve, how agents coordinate, and how team processes emerge. When interactions unfold sequentially, agents influence not only one another's immediate payoffs but also the future configuration of the task, the team, and their own opportunity structure. This shift places the study of teamwork within a richer dynamic framework aligned with both multi-agent learning and theories of group development.

Conceptually, the chapter integrates Teamwork Games with Ilgen et al.'s Input-Mediator-Output-Input (IMOI) model of dynamic teamwork, offering a formalisation of cooperative tasks under **temporal pressure**, heterogeneous roles, and partial observability. The result is a Markov Game formulation—*Sequential Teamwork Games*—that preserves the soft constructs introduced earlier—roles, interdependence, evaluation functions—while embedding them in a sequential decision-making environment. This formulation captures how team processes can arise endogenously, even in the absence of explicit communication or centralised coordination.

Computationally, the chapter introduces *TEamwork under Multi-agent Policy Optimisation (TEMPO)*, a multi-agent simulation environment that operationalises Sequential Teamwork Games. Mirroring the role of *CollabQuest* in Chapter 3, *TEMPO* provides a concrete instantiation of the theoretical model, enabling systematic exploration of cooperative behaviour under controlled structural and temporal conditions. Using empirical game-theoretic analysis over realised agent policies, the chapter examines when sequential teamwork satisfies the criteria for a Sequential Social Dilemma (SSD), thereby extending SSD analysis beyond classical repeated-matrix settings and connecting to work such as Perolat et al. [2017]. In doing so, it demonstrates that sequential interdependence can generate pro-social dynamics in environments where static incentives alone would not predict them.

Finally, *TEMPO* is designed as an extensible platform, supporting further operationalisations of group behaviour, decision-making, and coordination mechanisms. It thus provides both a testbed for empirical investigation and a bridge between computational modelling and psychological theories of team dynamics.

Taken together, these contributions position Chapter 4 as a temporal and behavioural deepening of the theory introduced in Chapter 3. They underscore that cooperation in teams is not

a fixed disposition but an emergent, context-sensitive dynamic shaped by interdependence, evaluation, heterogeneity, and the unfolding structure of interaction. By situating teamwork within the methodological toolkit of multi-agent reinforcement learning, the chapter lays the pathway for future research at the intersection of psychology, economics, and artificial intelligence.

6.1.3 Chapter 5: Social Alignment

Chapter 5 adds a final layer of abstraction to the study of cooperation by examining how workers' sense of belonging to an in-group shapes their behaviour. Here, cooperation no longer concerns performance within a team whose composition is fixed; instead, the group—more precisely, the firm—is treated as an *open system* that workers may join or leave **at will**. Firms thus form, grow, and collapse endogenously, and meso- and macro-level economic patterns emerge from the interaction of micro-level motivations.

Conceptually, the chapter integrates grounded psycho-social theories into the decision-making processes of utility-maximising agents. In particular, it offers a formal, micro-founded representation of how **social identity processes** influence economic behaviour, drawing on Self-Categorisation Theory (SCT). The central modelling insight is that social identity can be represented as a scalar field in the physical sense: a value assigned to each worker, generated endogenously through the dynamic contrast between the focal firm and the set of alternative firms visible to the worker.

This formalisation advances the field of Identity Economics by providing a modelling approach more firmly grounded in SCT, where social norms, self-categorisation, and identity salience emerge endogenously and shape the **evaluative frame** through which benefits and costs are assessed. Through comparative statics, the chapter characterises the joint effects of identity salience and firm size on workers' optimal effort and welfare. The results clarify when marginal increases in belongingness or headcount raise—or lower—a worker's performance and well-being, refining the seminal insights of Akerlof and Kranton [2000b] and specifying the conditions under which non-pecuniary motivations act as substitutes for wages.

Computationally, these cognitive processes are embedded into agents' optimisation rules within Axtell et al. [1999]'s labour-mobility agent-based model. Extending Axtell's framework provides a principled "stress test" for identity in a data-concordant agent model. Both analytically and through simulations, the chapter shows that activating firm identity systematically

shapes cooperation patterns, welfare outcomes, and the emergent structure of the economy, all while preserving the economic backbone of the *homo economicus* baseline.

The SCT-compatible formulations introduced in Chapter 5 are, in principle, applicable to any utility function, opening a pathway toward a more comprehensive identity-based economics grounded in social psychology.

In doing so, this chapter completes the dissertation's progression—from structural to learned to social moderators of cooperation—bridging micro-level cognition and macroeconomic coordination within a unified computational framework.

The next section distils the insights that emerge when these intertwined levels are considered as a unified whole. Beyond their individual theoretical contributions, the three models jointly suggest a broader perspective on cooperation as a multi-level phenomenon shaped simultaneously by incentives, temporal dynamics, learning, and social identity. This integrative view not only helps explain cooperative behaviour across human and artificial systems, but also points toward practical applications in organisational design, multi-agent artificial intelligence, and hybrid human–AI environments. The next sections develop these broader implications, discuss the limitations of the present framework, and outline potential avenues for transfer and future research.

6.2 Integrative Insights

Descriptive research opens the door to environmental engineering. The overarching idea is that we can only expect unenforced cooperation when it lies within an agent's self-interest. Understanding why and when cooperation arises enables us to design environments and conditions that encourage cooperation and coordination—an approach that is ultimately more sustainable and stable than relying on punitive systems for defection or incentive schemes for cooperation. A system becomes more aligned when conditions for cooperation are ripe—not when enforcement is more intense. Viewing behaviour as revealed utility maximisation invites a reinterpretation of what other disciplines have classed as irrational, emotional, or psychosocial behaviours—*e.g.*, *altruism*, *loss of motivation*, *social categorisation*—as conduct that maximises a utility function that is not strictly pecuniary, drawing on the tools of game theory and be-

havioural economics. This perspective opens the door to modelling reward-maximising agents who are boundedly rational *and* compatible with "soft" behaviour.

Cooperation as a multi-level phenomenon. Cooperation is a faceted, multi-level phenomenon that requires interdisciplinary study in order to be understood. The very description of an environment at different levels of analysis may yield completely different predictions about the behaviour of the agents involved. For example, the time agents spend succeeding or failing while learning as a team—*i.e.*, the fraction of time they spend in a *good* or *bad* state—affects, and can significantly alter, their incentives to cooperate relative to the one-shot version of the same problem. Moreover, analysing the group as a collection of interdependent agents (Chapters 3 and 4) will lead to conclusions about cooperation that differ from those obtained when the group is studied as a psychological entity that induces shifts in its members' evaluative frames (Chapter 5). Realistic prediction, therefore, requires an integral and nuanced perspective that incorporates all alignment mechanisms in their proper measure.

Observability as a transversal variable. When speaking of interdependence, we refer not only to material interdependence but also to informational interdependence: what is observed, what is attributed, what is valued, and what is learned. More concretely, at every level, what matters is as much what is seen as what is believed—group progress, contribution signals, evaluation, social network structure, intergroup comparison, and so forth.

Three sources of alignment. When we cooperate with others, we are immersed in a network of interdependence that shapes our motivations and the impact of our actions—and those of others—on our own and the group's outcomes. This interdependence has different roots. This thesis has identified and modelled three types: structural, temporal or dynamic, and socio-identity-based. When considered together, a single principle underlies them all: to cooperate is to achieve alignment between group incentives and the perceived selfish incentive (evaluative frame), stabilised by experience acquired through learning.

The system is self-reinforcing. The temporality and openness of the system in the models of Chapters 4 and 5 make clear that temporal structure brings feedback loops and endogeneity: insofar as cooperative policies are adopted within a system, they affect the state of the environment and the very suitability of continuing to adopt cooperative strategies or to defect. This ongoing rewriting of the system is too complex to be studied analytically, making simulation in the presence of temporality indispensable for a deep understanding of the micro-, meso-, and macro-scale.

Multi-agent simulation as a laboratory. Multi-agent simulation techniques can be used as *in vitro* laboratories in which to test the macroscopic dynamics that emerge from the specification of microscopic motivation and behaviour. These simulations serve a bidirectional role: they corroborate the plausibility of theoretical models by revealing their macroscopic and dynamic consequences, and they facilitate the parsimonious study of complex and formally intractable systems. The simulation environments at each level illuminate the levers that an environment designer can manipulate to incentivise cooperation in an organic, environmental manner. For example, structural-level simulations—*i.e.*, *CollabQuest*—inform us of the *type of task in which a team is most likely to thrive*, and can therefore be used as prescriptions when the aim is to maximise performance. Alternatively, suppose that the aim is to achieve an equitable distribution of effort. In that case, even when the underlying task technology cannot be changed, the incentive system can be manipulated to modify its practical typology. For instance, a task may be disjunctive by nature—solving a mathematical problem on a whiteboard—but it can be framed as conjunctive to force a bottleneck coordination that discourages free-riding behaviour: all members of the group must have contributed to some part of the exercise for the group to receive a positive evaluation. Preliminary steps toward extending this simulation-based logic to human teams are illustrated by *The Island* (Appendix C), a pilot platform designed to reproduce the same structural conditions as in *CollabQuest* and *TEMPO* empirically.

Together, these insights point toward a broader claim: **the emergence of cooperation is best understood not through equilibrium analysis alone, nor through learning dynamics alone, but through the integrated study of alignment mechanisms that operate simultaneously at structural, temporal, and social levels**. While this perspective outlines a coherent account of cooperation across levels, it also makes explicit the boundaries of what this dissertation can claim. The theoretical scope of the models, the simplifications they adopt, and the scale at which they operate inevitably leave questions open. The following section discusses these limitations and the avenues they naturally suggest for future research.

6.3 *Limitations and Future Work*

This section outlines the main limitations of the present work. It does so by focusing on the two levels that have remained in continuous dialogue throughout the dissertation—*theory* and *simulation*. Within this scope, the modelling strategy has been to isolate a small set of mechanisms with minimal confounds, rather than to maximise ecological realism.

Conceptually, the models in Chapters 3 and 4 restrict attention to **unitary tasks**, where team members differ in their level of expertise along a single continuous dimension, but not in the *type* of contribution they can make. In other words, all agents act on the same production function using varying amounts of a single, shared skill. Ongoing work will extend this framework to **divisible tasks**, following established distinctions from social psychology. This generalisation will allow agents to specialise across qualitatively different subcomponents of a task, providing a more faithful representation of variety as a form of team heterogeneity.

Both Teamwork Games and Sequential Teamwork Games also focus on **positive outcome interdependence**: evaluation functions that reward or penalise all agents based on the collective outcome. A natural extension is to incorporate **competitive interdependence within the team**—for example, by making evaluation contingent on relative rather than collective performance. Tournament-style mechanisms would offer a first level of structural analysis for this alternative organisational design.

In practice, the simulation environments developed in this dissertation prioritise **parsimony over realism**. For instance, the theoretical formulation of Teamwork Games allows for heterogeneity in preferences for leisure time—a plausible and important real-world variable—yet *CollabQuest* was instantiated with **a common preference parameter for all agents**. This choice enabled a clean assessment of the structural determinants of cooperation but necessarily limited the generality of the resulting conclusions. A natural next step is therefore to allow the leisure–effort trade-off parameter α to vary across agents.

A similar simplification appears in the modelling of **agent expertise**. Across Chapters 3 and 4, expertise levels are fixed: they do not evolve through exposure to others, nor through *human-like learning* (as opposed to bandit-learning or policy optimisation). Likewise, agents do not model one another; they treat the behaviour of their teammates as part of the environment's stochasticity. These are principled simplifications, not oversights: they make it possible to isolate the effects of task interdependence, partial observability, and temporally delayed evaluation without conflating them with skill acquisition or social cognition. By holding p_i

constant, the models provide controlled testbeds to examine how cooperation and free-riding emerge from incentives and team structure alone.

Future work will relax these assumptions by incorporating **evolving abilities**, **social learning mechanisms**, and **explicit modelling of other agents**. Similarly, communication was intentionally excluded to avoid confounds arising from signalling or protocol design. Once minimal coordination principles are properly understood, examining how communication channels alter role differentiation, policy convergence, and emergent organisation becomes both feasible and conceptually informative.

With respect to the empirical component of the thesis, while Appendix C introduces *The Island* as a preliminary platform to operationalise the structural conditions studied in *CollabQuest* and *TEMPO* with human participants, it should be emphasised that the current implementation is a **pilot** and **does not constitute empirical validation** of the models. The platform primarily supports **experimental control** (internal validity) rather than establishing **generalisable behavioural claims** (external validity). Accordingly, any behavioural regularities observed in small-scale deployments should be interpreted as **illustrative** rather than confirmatory. Establishing external validity will require **larger and more heterogeneous samples** and systematic replication across task types and evaluation regimes.

Finally, this dissertation has only **partially integrated** the models introduced across Chapters 3, 4, and 5. Sequential Teamwork Games extend the production geometry of Teamwork Games into a temporal domain, while the model in Chapter 5 shares the assumption of equal sharing of output and utility maximisation over income and leisure. However, because the identity-based model is embedded in an existing labour-market environment, it does not incorporate multi-agent learning or non-linear production functions. A compelling direction for future research is therefore to embed **social identity–compatible reward functions** into learning-based multi-agent environments—*e.g.*, *TEMPO*—to study how identity-sensitive agents learn to cooperate within dynamic teams.

6.4 *Applied Directions and Transfer Opportunities*

Beyond the potential extensions and integrations discussed earlier, this dissertation suggests avenues for application and transfer that, while outside the immediate scope of the present

work, arise naturally from its theoretical and computational contributions. In what follows, some concrete opportunities for deployment in real-world domains are presented, illustrating how the theoretical insights of this research may yield tangible impact.

Organisational and Team Design. The structural results of Chapter 3 make it possible to use game-theoretic models such as *CollabQuest* as tools for team planning and organisational design. In particular, they may inform practice along at least three avenues:

1. task design and decomposition strategies that minimise free-riding and maximise joint performance,
2. incentive schemes that align individual effort with collective outcomes,
3. team composition policies based on complementarities in expertise and interdependence requirements.

Real-world scenarios such as collaborative research, software development, and distributed work environments would potentially benefit from the application of these approaches.

Identity-driven Cooperation in Artificial Intelligence Systems. The social categorisation mechanism developed in Chapter 5 suggests new avenues for building socially aligned artificial systems. By endowing agents with a notion of group membership that modulates their willingness to comply with shared norms, it becomes possible to model—and potentially engineer—cooperative behaviour emerging from identity-driven incentives rather than from purely pecuniary or reward-maximising considerations.

Such identity-aware agents could be used to investigate how socially grounded norms may be leveraged to steer behaviour in ways consistent with designer goals and system-level welfare, opening a path towards artificial societies whose collective behaviour is guided by norm adherence rather than external enforcement.

This form of alignment may be particularly relevant for applications involving large populations of autonomous agents operating in shared environments—for example, fleets of autonomous vehicles, distributed robotic systems, or multi-agent service platforms—where cooperative adherence to social conventions (*e.g.*, sharing resources, avoiding congestion) is essential for safety, efficiency, and public trust.

Human Experiments and Hybrid Systems. The pilot platform *The Island* provides a foundation for empirical validation and for experiments involving hybrid human-AI interaction. In addition to the experimental agenda outlined in Section 6.3, future iterations could support

interfaces in which humans and learning agents collaborate or compete, enabling controlled studies of how the introduction of one or more artificial agents—cooperative or defecting—shapes the evolution of teamwork and affects the stability of cooperation.

Such experiments may help identify disruptive factors that destabilise cooperative groups and inform the design of mechanisms to mitigate them. Beyond organisational settings, this line of work also suggests applications in areas such as online moderation, where the deployment of generative or LLM-based agents acting as cooperative participants could foster prosocial interaction and attenuate hostile dynamics within social media environments.

6.5 Outstanding Questions and Concluding Remarks

Against this backdrop, several outstanding questions arise that extend the theoretical and computational framework developed in this dissertation. These questions are posed to point out the natural next steps once the core structural, temporal, and social mechanisms have been isolated:

- **Competitive interdependence within teams.** What happens when interdependence among team members is competitive rather than positive? What are the equilibrium solutions of static games with relative or tournament-style evaluation, and to what extent do they complement the results obtained for Teamwork Games or align with established theories of intra-group competition and group productivity?
- **Relationship between structural equilibria and learned policies.** What is the connection between the Nash Equilibria characterised in Teamwork Games and the policies learned by agents in Sequential Teamwork Games? Do the structural equilibria act as attractors for the policies that emerge through learning in the sequential setting?
- **Human behaviour in analogous settings.** In situations analogous to those modelled in Teamwork Games and Sequential Teamwork Games, how do people actually behave? What behavioural traces can be collected from human participants in controlled environments, and how might such data be used to refine the structural, cognitive, and motivational assumptions of the models?

- **Extending self-categorisation models to full social identity dynamics.** Given an endogenous model of self-categorisation, can the framework be extended to multi-agent systems with learning so as to capture the broader identity-driven processes typically associated with social identity theory? More specifically, how might phenomena such as prejudice, norm adherence, or intergroup aggression be represented in environments populated by self-categorising, learning agents?

These questions indicate merely a fraction of the work that remains to be done.

Ultimately, the models developed in this dissertation open concrete pathways for future work. Chapter 3 provides a principled basis for the design of environments in which cooperation becomes self-sustaining through the structure of the task and its evaluation. Chapter 4 shows that as agents learn, they do not merely respond to incentives: they change them. By altering which states are visited and how often, learning dynamics modulate the cooperative possibilities identified in Chapter 3. This points to the need for models that treat incentives and learning as mutually shaping processes rather than as independent lenses. Chapter 5 demonstrates that social identity can be expressed in computational terms without sacrificing analytical tractability, suggesting a route toward agents whose motivations extend beyond pecuniary payoff yet remain formally grounded. Together, these contributions outline a research programme for studying cooperation across multiple levels of organisation within a single modelling architecture.

In tackling cooperation from an interdisciplinary standpoint, the work highlights the need for a genuine dialogue between strategic, psychosocial, and economic formalisations in order to approach a question that touches on our very nature as a species and arguably distinguishes us from others: *why do we cooperate?*

The value of the approach lies not in providing a definitive answer, but in clarifying how such an answer might be pursued. By rendering the mechanisms of cooperation explicit, formal, and testable, the dissertation opens a parsimonious and transparent pathway toward the investigation of why we cooperate—one that is open to verification, replication, and falsification through the systematic use of computational simulation.

Acronyms

ABM Agent Based Modelling. 150, 157, 166

AI Artificial Intelligence. 5, 6, 45, 46

CES Constant Elasticity of Substitution. 23, 40, 49, 51, 57, 68–72, 102, 104, 105, 108, 110, 125, 127, 166, 220, 221, 227, 229

CPR Common-Pool Resource. 101

DRL Deep Reinforcement Learning. 27, 28

GAE Generalised Advantage Estimation. i, 30, 31, 111, 113, 238

GT Game Theory. 24, 44

I-P-O Input-Process-Output. 98

IMOI Input-Mediator-Output-Input. ii, 94, 97–99, 138, 202

iPPO independent Proximal Policy Optimization. 14, 15, 110, 111, 243, 250

LOO Leave-One-Out. 167, 168

MA-MAB Multi-Agent Multi-Armed Bandit. ii, 12, 14, 15, 24, 35–37, 67, 69–71, 75, 77, 78, 81, 87, 201

MARL Multi-Agent Reinforcement Learning. i, 9, 10, 12, 15, 45, 46, 68, 95, 101, 103, 106, 113, 127

MAS Multi-Agent Systems. 7, 43, 45, 68, 95, 100, 103, 151, 157, 158

MCR Meta-Contrast Ratio. iii, v, ix, 14, 151, 155, 160, 163, 166, 167, 169, 171, 176, 184, 186, 187, 196, 265, 277–279

NE Nash Equilibrium. ii, vii, 20, 37, 53, 55–57, 59, 75–78, 210, 229, 230

NMI Normalised Mutual Information. 142, 143, 145, 146

PMF Probability Mass Function. viii, 194

PPO Proximal Policy Optimization. i, 28, 30–32, 111, 113, 150

QRE Quantal Response Equilibrium. 26

RL Reinforcement Learning. 9, 45

SCT Self-Categorisation Theory. iii, viii, xii, 10, 12, 14, 149–152, 154, 156–160, 162, 164, 165, 168–173, 203, 204

SSD Sequential Social Dilemma. iv, vii, viii, 53, 94, 100, 101, 112, 115–117, 119, 120, 122–125, 127, 202, 237, 240, 245

SVO Social Value Orientation. 9, 102, 104

TEMPO TEamwork under Multi-agent Policy Optimisation. 12, 103, 107, 112, 113, 115, 117, 127, 139, 140, 150, 202, 206, 208

TMS Transactive Memory System. 139, 140, 142, 145, 146



A Teamwork Games

A.1 Proof of Lemma 3.9 220
A.2 Proof of Lemma 3.10 and Proposition 3.12 220
A.3 Analysis of a player's replacement function in teamwork games with conjunctive tasks 222
A.4 Proof of Lemma 3.11 224
A.5 Teamwork Games as Social Dilemmas 227
A.6 Proof of Proposition 3.14 231
A.7 Individual efforts in disjunctive and conjunctive tasks: tables 232

A.1 Proof of Lemma 3.9

Let $\widetilde{u}_i(x_i, \sigma(G)) = x_i^\alpha \, \sigma(G)$ with $\alpha > 0$ and σ an evaluation function. For convenience, we analyse $\widetilde{u}_i$ as a function of (x_i, G) rather than $(x_i, \sigma(G))$. Since $\sigma(\cdot)$ is strictly increasing, both approaches are equivalent for quasi-concavity and monotonicity.

Differentiability and monotonicity. Both x_i^α and $\sigma(G)$ are differentiable functions whose derivatives are continuous. Therefore their product $\widetilde{u}_i(x_i, \sigma(G))$ is also continuously differentiable in (x_i, G).

Moreover,

$$\frac{\partial \widetilde{u}_i}{\partial x_i} = \alpha x_i^{\alpha-1} \sigma(G) > 0, \qquad \frac{\partial \widetilde{u}_i}{\partial G} = x_i^\alpha \sigma'(G) > 0$$

because $\alpha > 0$, $x_i > 0$, and $\sigma'(G) > 0$ by Definition 3.6. Hence $\widetilde{u}_i$ is strictly increasing in both arguments.

Strict quasi-concavity. Consider the logarithm of $\widetilde{u}_i$:

$$\log \widetilde{u}_i(x_i, \sigma(G)) = \alpha \log x_i + \log \sigma(G)$$

The function $\log x_i$ is strictly concave on $(0, \infty)$. Condition 3 of Definition 3.6 is equivalent to

$$\frac{d^2}{dG^2}\big(\log \sigma(G)\big) < 0$$

so $\log \sigma(G)$ is strictly concave in G.

Since a positive linear combination of strictly concave functions is strictly concave, the function $\log \widetilde{u}_i$ is strictly concave in (x_i, G) on the domain where $\widetilde{u}_i > 0$. A positive function whose logarithm is strictly concave is strictly quasi-concave, so $\widetilde{u}_i$ is strictly quasi-concave.

Conclusion. The utility $\widetilde{u}_i(x_i, \sigma(G)) = x_i^\alpha \sigma(G)$ is continuously differentiable, strictly increasing in both arguments, and strictly quasi-concave. Therefore, it satisfies Assumption 3.8. □

A.2 Proof of Lemma 3.10 and Proposition 3.12

In a central lemma in [Cornes and Hartley, 2007b], the authors go on to show the circumstances under which given a general public good game with a concave CES aggregator, a well-defined

replacement function exists:

Lemma A.1 (Cornes & Hartley). *Let $\Gamma = (\mathcal{A}_i, \widetilde{u}_i)_{i\in\mathcal{I}}$ be a generalised aggregative game featuring a concave CES social composition function. Under increasing, normal preferences, there is a unique $g_i \in (0, w_i]$ for each player satisfying the first-order condition:*

$$\left.\frac{\partial \widetilde{u}_i}{\partial g_i}\right|_{g_{-i}} \geq 0$$

with equality if $g_i < w_i$. Furthermore, the feasibility condition $G^\rho \geq \beta_i g_i^\rho$ is satisfied for $G \geq \overline{G}_i$ if $0 < \rho < 1$ and for $G \leq \overline{G}_i$ if $\rho < 0$.

This lemma implies that, in general public-good games with a concave CES aggregator, the replacement correspondence $r_i(G)$ is a well-defined function satisfying the first-order condition and with domain $[\overline{G}_i, \infty)$ if $0 < \rho < 1$ and $[0, \overline{G}_i]$ if $\rho < 0$.

A later proposition in the same paper establishes existence and uniqueness of equilibrium:

Proposition A.2 (Cornes & Hartley). *Under the conditions of Lemma A.1, the game has a unique equilibrium whenever the utility function $\widetilde{u}_i(x_i, G)$ is such that the corresponding indifference map is asymptotic to the axes.*

We now adapt these results to Teamwork Games. In a Teamwork Game, player i's reduced utility depends on the private good x_i and on the evaluated public good $\sigma(G)$:

$$\widetilde{u}_i(x_i, \sigma(G))$$

By Assumption 3.8, the preferences represented by $\widetilde{u}_i(x_i, \sigma(G))$ are continuously differentiable, strictly increasing and strictly quasi-concave in $(x_i, \sigma(G))$, and both x_i and $\sigma(G)$ are normal goods. Because $\sigma(G)$ is strictly increasing, $\widetilde{u}_i(x_i, \sigma(G))$ induces the same preference ordering as a utility function defined over (x_i, G), preserving normality, monotonicity and strict quasi-concavity. Hence, we may define

$$u_i(x_i, G) \equiv \widetilde{u}_i(x_i, \sigma(G))$$

as a utility function defined on (x_i, G), and the hypotheses of Lemma A.1 are satisfied whenever the CES aggregator $G(\boldsymbol{g})$ is concave in g_i.

This is precisely the case in Teamwork Games involving a conjunctive task ($\rho < 1$, $\rho \neq 0$), where G is concave in each g_i. In our setting, the budget constraint is

$$\Delta = x_i + \frac{g_i}{p_i}$$

so the maximal feasible gift is $g_i = p_i\Delta$. Thus the condition $g_i < w_i$ in Lemma A.1 is replaced by $g_i < p_i\Delta$.

It follows that, in a Teamwork Game $(\Gamma, \mathcal{T}, \sigma)$ with a conjunctive task, the replacement correspondence $r_i(G)$ is in fact a function satisfying the first-order condition

$$\left.\frac{\partial \widetilde{u}_i}{\partial g_i}\right|_{g_{-i}} \geq 0$$

with equality whenever $g_i < p_i\Delta$, and with domain $[\overline{G}_i, \infty)$ if $0 < \rho < 1$ and $[0, \overline{G}_i]$ if $\rho < 0$. This proves Lemma 3.10.

Finally, adapting Proposition A.2 to the utility $\widetilde{u}_i(x_i, \sigma(G))$ yields Proposition 3.12: under the conditions of Lemma 3.10, the Teamwork Game has a unique equilibrium whenever the indifference curves of $\widetilde{u}_i(x_i, \sigma(G))$ are asymptotic to the axes. □

A.3 *Analysis of a player's replacement function in teamwork games with conjunctive tasks*

Let us analyse how the replacement function in a conjunctive task ($\rho < 1$) is impacted by changes in each of the elements of a teamwork game:

$$r_i(G)^{\rho-1}\left(\Delta - \frac{r_i(G)}{p_i}\right) = \frac{\alpha}{\beta_i p_i} G^{\rho-1} \frac{\sigma(G)}{\sigma'(G)}$$

- Deriving both sides of the equation *w.r.t.* G shows that the derivative of the left-hand side must be negative:

$$\frac{dr_i}{dG} r_i^{\rho-2}\left[\rho\left(\Delta - \frac{r_i}{p_i}\right) - \Delta\right] = \frac{\alpha}{\beta_i p_i} G^{\rho-2}\left[(\rho-1)\frac{\sigma(G)}{\sigma'(G)} + G\,\frac{\sigma'(G)^2 - \sigma''(G)\sigma(G)}{\sigma'(G)^2}\right]$$

  Indeed, using that $\frac{r_i(G)}{p_i} \in [0, \Delta]$ and $(\rho - 1) < 0$, we see that the term in brackets in the left-hand side of the equation above must be negative. Thus, there are two behavioural regimes for $r_i(G)$ as G increases

  - If the term in brackets of the right-hand side of the equation is positive, then $\frac{dr_i(G)}{dG} < 0$. This is true whenever $\frac{|\rho-1|}{G} < \frac{\sigma'(G)^2 - \sigma''(G)\sigma(G)}{\sigma'(G)\sigma(G)}$.

– If the term in brackets of the right-hand side of the equation is negative, $\frac{dr_i(G)}{dG} > 0$, which happens whenever $\frac{|\rho-1|}{G} > \frac{\sigma'(G)^2-\sigma''(G)\sigma(G)}{\sigma'(G)\sigma(G)}$.

This resonates with the qualitative regimes that Cornes and Hartley [2007b] distinguished (see section 3.4.2): if ρ is sufficiently negative as in the second case, a substitution effect occurs and encourages player i to contribute more since a given increase in G generated by their contribution now has a lower opportunity cost in terms of private consumption. Otherwise, an income effect dominates and the preferred contributions by i will decrease as G increases.

- When deriving *w.r.t* α, we obtain the equation:

$$\frac{dr_i}{d\alpha} r_i^{\rho-2} \left((\rho-1)\Delta - \frac{\rho}{p_i} r_i \right) = \frac{G^{\rho-1}}{\beta_i p_i} \frac{\sigma(G)}{\sigma'(G)}$$

The right-hand side of this equation is positive because $\sigma'(G) > 0$, and the term in parentheses on the left is negative by the same argument as above. Hence, $\frac{dr_i(G)}{d\alpha} < 0$. This result is in line with what was observed in section 3.4.2: as the importance that a player gives to their leisure time increases, their optimal contributions decrease.

- Deriving *w.r.t.* β:

$$\frac{dr_i}{d\beta_i} r_i^{\rho-2} \left((\rho-1)\Delta - \frac{\rho}{p_i} r_i \right) = -\frac{G^{\rho-1}\alpha}{\beta_i^2 p_i} \frac{\sigma(G)}{\sigma'(G)}$$

one sees that the right-hand term is negative. Again, the factor in parentheses on the left is negative. Hence, the derivative is positive to maintain the sign: $\frac{dr_i(G)}{d\beta_i} > 0$. The higher the weight or importance that a player's contribution has, the higher the optimal contribution will be.

- Regarding the term $\frac{\sigma(G)}{\sigma'(G)}$, it had been already established that this coefficient represents the ratio between the incentive $\sigma(G)$ and the slope of the tangent curve to the incentive $\sigma(G)$, and thus is a measure of *opportunity cost* in terms of leisure. Deriving with respect to it yields:

$$\frac{dr_i}{d(\sigma/\sigma')} r_i^{\rho-2} \left((\rho-1)\Delta - \frac{\rho}{p_i} r_i \right) = \frac{\alpha}{\beta_i p_i} G^{\rho-1}$$

Using similar arguments to the above, it is straightforward to see that $\frac{dr_i(G)}{d(\sigma/\sigma')} < 0$. Just as in Section 3.4.2, higher values of this ratio (*i.e.* higher opportunity cost in terms of enjoyed leisure) induce lower contributions.

- When deriving with respect to the turn duration, Δ, one obtains:

$$\frac{dr_i}{d\Delta} = \frac{p_i\, r_i}{\rho\, r_i + p_i(1-\rho)\Delta}$$

Using $\rho - 1 < 0$, we see that the fraction's numerator and denominator are positive. Thus, $\frac{dr_i(G)}{d\Delta} > 0$, meaning that the more time a player has to perform a team task, the more they will contribute.

- Finally, the derivative *w.r.t.* the expertise level p_i yields, after some re-arrangement of the terms:

$$\frac{dr_i}{dp_i}\left[r_i^{\rho-2}(\rho-1)\left(\Delta - \frac{r_i}{p_i}\right) - \frac{r_i^{\rho-1}}{p_i}\right] = -\frac{\alpha}{\beta_i\,(p_i)^2} G^{\rho-1}\frac{\sigma(G)}{\sigma'(G)} - \frac{r_i^{\rho}}{(p_i)^2}$$

Using that $\rho - 1 < 0$ and $r_i(G) \in [0, p_i\Delta]$, we can see that $\frac{dr_i(G)}{dp_i} > 0$. Again, just as in the canonical case, higher expertise levels induce higher contributions to player i.

A.4 Proof of Lemma 3.11

Given a **general public good game** with Cobb-Douglas preferences and $\rho > 1$, Cornes and Hartley [2007b] demonstrate in a lemma the existence of three potential behavioural regimes within the set of best responses:

Lemma A.3. *There exists a threshold value $G^*_{-i} > 0$ and a positive real-valued function $\widetilde{b}_i$ on $\left[0, G^*_{-i}\right]$ such that:*

$$\mathcal{B}_i(G_{-i}) = \begin{cases} \{0\} & \text{if } G_{-i} > G^*_{-i} \\ \left\{0, \widetilde{b}_i(G_{-i})\right\} & \text{if } G_{-i} = G^*_{-i} \\ \left\{\widetilde{b}_i(G_{-i})\right\} & \text{if } G_{-i} < G^*_{-i} \end{cases} \tag{A.1}$$

Their proof revolves around the observation that the stationary points of the Cobb-Douglas utility function satisfy an expression that can be decomposed into a term dependent on g_i, denoted $\phi(g_i)$, and another dependent on G_{-i}:

$$\frac{d\widetilde{u}_i}{dg_i} = \phi(g_i) - \alpha\, G_{-i}^{\rho} = 0 \tag{A.2}$$

The authors establish that $\phi(g_i)$ attains a maximum $\phi_{\max}$, implying that Equation (A.2) holds as long as $\alpha\, G_{-i}^{\rho}$ is less than or equal to $\phi_{\max}$. Therefore, $\phi_{\max} \geq \alpha\, G_{-i}^{\rho}$ serves as a necessary and sufficient condition for the stationary points of the utility function. Completing the proof involves: 1) demonstrating the well-defined nature of G_{-i}^{*}, 2) assigning the curve in the upper panel of Figure 3.2 to the case $G_{-i} > G_{-i}^{*}$, 3) associating the middle panel with $G_{-i} = G_{-i}^{*}$, and 4) assigning the lower panel to the case $G_{-i} < G_{-i}^{*}$.

The approach that we will follow regarding teamwork games mirrors that of Cornes and Hartley: we will take the utility function to be

$$\widetilde{u}_i = x_i^{\alpha}\, \sigma(G)$$

and we will break down $\frac{\partial \widetilde{u}_i}{\partial g_i}$ into two components: one dependent on $(g_i, \sigma(G))$, and another related to G_{-i}; and we will prove that $\phi(g_i, \sigma(G))$ attains a maximum. After that, we will show that G_{-i}^{*} is well-defined. Then, we will reason how each regime of G_{-i} relates to the panels in Figure 3.2.

The stationary points of utility function satisfy:

$$\frac{\partial \widetilde{u}_i}{\partial g_i} = \phi(g_i, \sigma(G)) - f(G_{-i})$$

By definition, the threshold points (g_i^{*}, G_i^{*}) would satisfy $\frac{\partial \widetilde{u}_i}{\partial g_i} = 0$. If one demonstrates that $\phi(g_i, \sigma(G))$ has a maximum $\phi_{\max}\, \forall G$, then there will exist critical points (g_i^{*}, G_i^{*}) as long as $f(G_{-i}) \leq \phi_{\max}$.

We have established that the first-order condition $\frac{\partial \widetilde{u}_i}{\partial g_i} = 0$ is equivalent to

$$\frac{\sigma(G)}{\sigma'(G)} G^{\rho-1} = \left(\Delta - \frac{g_i}{p_i}\right) \frac{\beta_i p_i}{\alpha} g_i^{\rho-1}.$$

Rearranging,

$$\alpha^{\frac{\rho}{\rho-1}} \left(\beta_i g_i^{\rho} + G_{-i}^{\rho}\right) = \left[\left(\Delta - \frac{g_i}{p_i}\right) \left(\frac{\sigma'(G)}{\sigma(G)}\right) \beta_i p_i\right]^{\frac{\rho}{\rho-1}} g_i^{\rho}.$$

Define

$$\phi(g_i, \sigma(G)) = \left[\left(\Delta - \frac{g_i}{p_i}\right) \left(\frac{\sigma'(G)}{\sigma(G)}\right) \beta_i p_i\right]^{\frac{\rho}{\rho-1}} g_i^{\rho} - \alpha^{\frac{\rho}{\rho-1}} \beta_i g_i^{\rho}.$$

Clearly,

$$\phi(0) = 0, \qquad \phi(p_i \Delta) = -\alpha^{\frac{\rho}{\rho-1}} \beta_i (p_i \Delta)^{\rho} < 0,$$

and fixing G the equation $\phi(g_i) = 0$ has a second root at

$$c = p_i\Delta - \frac{\alpha}{\beta_i^{1/\rho}} \frac{\sigma(G)}{\sigma'(G)}, \qquad 0 < c < p_i\Delta.$$

To determine whether the interior extremum of ϕ is positive, differentiate ϕ. Imposing $\frac{\partial \phi}{\partial g_i} = 0$ yields

$$\left(\Delta - \frac{g_i}{p_i}\right)^k \left(\frac{\sigma'}{\sigma}\right)^k (\beta_i p_i)^k g_i^\rho \left(\frac{-k/p_i}{\Delta - \frac{g_i}{p_i}} + k\frac{\sigma\sigma'' - (\sigma')^2}{\sigma\sigma'} \beta_i \left(\frac{g_i}{G}\right)^{\rho-1} + \frac{\rho}{g_i} \right) - \alpha^k \beta_i \rho g_i^{\rho-1} = 0, \qquad k = \tfrac{\rho}{\rho-1}.$$

Extracting the common factor

$$\left(\Delta - \frac{g_i}{p_i}\right)^k \left(\frac{\sigma'}{\sigma}\right)^k (\beta_i p_i)^k g_i^\rho$$

and substituting into ϕ gives that the interior critical point satisfies $\phi(g_i^*) > 0$ if and only if

$$\frac{-k/p_i}{\Delta - \frac{g_i}{p_i}} + k\frac{\sigma\sigma'' - (\sigma')^2}{\sigma\sigma'} \beta_i \left(\frac{g_i}{G}\right)^{\rho-1} < 0.$$

Condition **3)** in Definition 3.6 ensures that this inequality holds for every G. Hence, any interior maximiser g_i^* satisfies $\phi(g_i^*) > 0$.

Since ϕ is continuous on the compact interval $[0, p_i\Delta]$, by the Weierstrass Theorem it attains a global maximum and a global minimum on this interval. Because $\phi(0) = 0$, $\phi(p_i\Delta) < 0$ and the global maximiser yields $\phi(g_i^*) > 0$, the maximum must be attained at an interior point and satisfy

$$\phi_{\max} > 0 \qquad \forall G.$$

Therefore, ϕ achieves a strictly positive maximum for every G, which establishes the existence of the threshold G_{-i}^*, following the structure of Cornes and Hartley [2007b].

The correspondence of each panel in Fig. (3.2) to the different regimes of G_i follows the same reasoning as in [Cornes and Hartley, 2007b]. Due to space constraints, we do not repeat that reasoning here.

We can solve for G_i^ρ in the global maximum condition (Eq. (3.18)) and substitute in the local maximum condition (3.17) to check if G_{-i}^* is well-defined:

$$\left(\Delta - \frac{g_i^*}{p_i}\right)^\alpha \sigma\left(\left(\frac{\sigma'(G_i^*)}{\sigma(G_i^*)} \frac{\beta_i}{\alpha} p_i\right)^{\frac{1}{\rho-1}} \left(\Delta - \frac{g_i^*}{p_i}\right)^{\frac{1}{\rho-1}} g_i^* \right) = \Delta^\alpha \sigma\left(\left[\left(\frac{\sigma'(G_i^*)}{\sigma(G_i^*)} \frac{\beta_i}{\alpha} p_i\right)^{\frac{\rho}{\rho-1}} \left(\Delta - \frac{g_i^*}{p_i}\right)^{\frac{\rho}{\rho-1}} - \beta_i \right]^{\frac{1}{\rho}} g_i^* \right)$$

For $\beta_i > 0$, the argument of σ in this equation's right-hand side (RHS) is smaller than that of the left-hand side (LHS). Given that σ is monotonically increasing, we have that the quotient $\frac{\sigma_{\text{RHS}}}{\sigma_{\text{LHS}}} < 1$. Thus, we can write

$$g_i^* = p_i \Delta \left(1 - \left(\frac{\sigma_{\text{RHS}}}{\sigma_{\text{LHS}}}\right)^{1/\alpha}\right) > 0 \tag{A.3}$$

By differentiating both sides of the global maximum condition (3.18) and substituting in the local maximum equation (3.17), we derive an expression of G_{-i}^* in terms of g_i^*:

$$\frac{\sigma(G_{-i}^*)}{\sigma'(G_{-i}^*)} G_{-i}^{*(\rho-1)} = \left(\Delta - \frac{g_i^*}{p_i}\right) \frac{\beta_i p_i}{\alpha} g_i^{*(\rho-1)}$$

which proves that $G_{-i}^* > 0$, confirming that G_{-i}^* is well-defined. □

A.5 Teamwork Games as Social Dilemmas

This appendix provides an intuitive static characterisation of *Teamwork Games* as instances of classical social dilemmas. For simplicity, we consider a two-player version ($i \in \{1,2\}$) in a *binary pass–fail* setting with the three canonical task types. Each player has two actions: *cooperate* (C), meaning to exert a prescribed cooperative effort (defined below), and *defect* (D), meaning to exert no effort. As usual, player i exerts some effort a_i, which is converted into a *gift* g_i:

$$g_i = a_i p_i \Delta$$

where p_i is expertise and Δ the duration of the decision period. With role weights $\beta_i = 1$, the CES production function becomes

$$G(g_1, g_2) = \begin{cases} g_1 + g_2, & \rho = 1 \quad \text{(additive)}, \\ \min\{g_1, g_2\}, & \rho \to -\infty \quad \text{(conjunctive)}, \\ \max\{g_1, g_2\}, & \rho \to +\infty \quad \text{(disjunctive)}. \end{cases}$$

The task has a performance requirement $G^{\mathcal{T}}$. Under a hard-threshold approximation, evaluation is binary:

$$\mathbb{H}(G - G^{\mathcal{T}}) = \begin{cases} 1, & \text{if } G \geq G^{\mathcal{T}}, \\ 0, & \text{otherwise.} \end{cases}$$

Utility is given by:

$$\widetilde{u}_i(a_i, G) = x_i^{\alpha_i}\, \mathbb{H}(G - G^{\mathcal{T}}), \qquad x_i = \Delta\,(1 - a_i)$$

so that:

- If the task fails ($\mathbb{H} = 0$), then $\widetilde{u}_i = 0$ regardless of effort.
- If the task succeeds ($\mathbb{H} = 1$), then $\widetilde{u}_i = x_i^{\alpha_i}$: success is valuable, but effort reduces leisure.

In what follows, we assume symmetric expertise and preferences, $p_1 = p_2 = p$ and $\alpha_1 = \alpha_2 = \alpha$.

Cooperative effort and induced payoffs

Define the *prescribed cooperative contribution* as the minimal symmetric gift level $g_C^\star$ such that, if both players choose it, the task meets the threshold:

$$G(g_C^\star, g_C^\star) = G^{\mathcal{T}}. \tag{A.4}$$

The corresponding effort is the cooperate action C, $a_C^\star = g_C^\star/(p\Delta)$, and the defect action D is $a_D = 0$, $g_D = 0$. For each task type:

- **Additive** ($\rho = 1$): $G = g_1 + g_2$, so

  $$g_C^\star = \frac{G^{\mathcal{T}}}{2}, \qquad a_C^\star = \frac{G^{\mathcal{T}}}{2p\Delta}$$

  Both players must cooperate to reach the threshold.

- **Conjunctive** ($\rho \to -\infty$): $G = \min\{g_1, g_2\}$, so

  $$g_C^\star = G^{\mathcal{T}}, \qquad a_C^\star = \frac{G^{\mathcal{T}}}{p\Delta}$$

  Any deviation to D drops the minimum to zero and guarantees failure.

- **Disjunctive** ($\rho \to +\infty$): $G = \max\{g_1, g_2\}$, so

  $$g_C^\star = G^{\mathcal{T}}, \qquad a_C^\star = \frac{G^{\mathcal{T}}}{p\Delta}$$

  A single cooperator suffices for success and only (D, D) fails.

When the task succeeds, Player i's payoff is

$$R := x_i^\alpha > 0$$

if she cooperates, and

$$T := \Delta^\alpha > R$$

if she defects. When the task fails, the payoff is 0 regardless of effort.

Additive tasks: a degenerate Assurance Game

Given $g_C^\star = G^\mathcal{T}/2$, the additive production function satisfies:

$$G(C,C) = g_C^\star + g_C^\star = G^\mathcal{T} \quad \Rightarrow \quad \text{SUCCESS}$$

and

$$G(C,D) = G(D,C) = g_C^\star < G^\mathcal{T}, \qquad G(D,D) = 0 < G^\mathcal{T} \quad \Rightarrow \quad \text{FAILURE}$$

Success thus occurs only if both players cooperate. The payoff matrix for Player 1 is:

	C	D
C	R	0
D	0	0

This is a weak coordination game: (C,C) and (D,D) are both Nash Equilibria, and (C,C) strictly Pareto-dominates (D,D). Notice that additivity would normally allow for partial compensation across members, but the discontinuity at $G^\mathcal{T}$ suppresses this substitutability by awarding zero evaluation to any sub-threshold outcome. The resulting game is therefore a *degenerate Assurance Game*: the fear-based incentive is induced by the evaluation rule rather than by intrinsic bottlenecks in production.

Conjunctive tasks: an Assurance Game (Stag Hunt)

For conjunctive tasks, $G = \min\{g_1, g_2\}$ and success requires both gifts to exceed $G^\mathcal{T}$. With $g_C^\star = G^\mathcal{T}$ we have:

$$G(C,C) = \min\{g_C^\star, g_C^\star\} = G^\mathcal{T} \quad \Rightarrow \quad \text{SUCCESS}$$

and any deviation to D yields

$$G(C,D) = G(D,C) = G(D,D) = 0 < G^\mathcal{T} \quad \Rightarrow \quad \text{FAILURE}$$

The payoff matrix is identical to the additive case:

	C	D
C	R	0
D	0	0

Again, (C, C) and (D, D) are both Nash Equilibria, with (C, C) the Pareto-optimal outcome. Here, however, the Assurance (Stag-Hunt) structure is technological rather than evaluative: group performance is determined by the weakest contributor, so reaching the evaluation threshold requires *each* player to meet the prescribed effort. Any unilateral deviation collapses the minimum and guarantees failure. The fear-driven incentive to coordinate is thus structurally built into conjunctive production, yielding a *pure Stag Hunt*.

Disjunctive tasks: a Volunteer's Dilemma

For disjunctive tasks, $G = \max\{g_1, g_2\}$ and success requires that at least one player meets the threshold. With $g_C^\star = G^\mathcal{T}$:

$$G(C,C) = \max\{g_C^\star, g_C^\star\} = G^\mathcal{T}, \quad G(C,D) = G(D,C) = \max\{g_C^\star, 0\} = G^\mathcal{T}$$

and

$$G(D,D) = 0 < G^\mathcal{T}$$

Any profile with at least one cooperator succeeds; only (D, D) fails. Player 1's payoff matrix is:

	C	D
C	R	R
D	T	0

with $T > R > 0$. This is the *Volunteer's Dilemma*: if the other defects, cooperating is optimal to avoid joint failure; if the other cooperates, defecting is optimal to free-ride. Asymmetric profiles (C, D) and (D, C) are strict Nash equilibria, while (D, D) is inefficient and unstable. The tension is greed-driven: each player prefers that someone else bear the cost of making the task succeed.

Soft thresholds

The binary hard-threshold analysis above assumes that success requires hitting $G^\mathcal{T}$ exactly. In the main text, this is replaced by a smooth logistic evaluation,

$$\sigma(G) = \frac{d}{1 + \mathrm{e}^{-\kappa(G - G^\mathcal{T})}},$$

which awards partial credit below the nominal threshold. For finite steepness κ, three qualitative effects arise:

1. The effective cooperative requirement is reduced: it is no longer necessary to hit $G^{\mathcal{T}}$ exactly to obtain non-zero evaluation.

2. Deviations from the prescribed effort can still be beneficial if they raise G.

3. The production technology (additive, conjunctive, disjunctive) is unchanged, and so is the *type* of social tension (greed vs. fear), even if its intensity is modulated.

Task-specific consequences follow. In **additive tasks**, partial credit allows agents to benefit from a partner's high contribution while exerting strictly less than the prescribed effort, thereby *introducing a greed motive* absent under the hard threshold: the incentive structure shifts from fear-dominated coordination to a Prisoner's-Dilemma-like temptation to defect. In **conjunctive tasks**, the minimum operator continues to bind, so the penalty for unilateral under-contribution remains large: fear persists as the dominant incentive, albeit in attenuated form. In **disjunctive tasks**, where one strong contribution suffices, smoothing the threshold *amplifies* the temptation to rely on a "volunteer", as non-contributing agents can still obtain a high evaluation provided that someone else contributes enough. In summary, softening evaluation does not change which social dilemma each task type instantiates, but it systematically modulates incentive strength: it introduces greed in additive tasks, preserves fear in conjunctive tasks, and reinforces greed in disjunctive tasks.

A.6 Proof of Proposition 3.14

By Def. 2.6, the share correspondence is a function whenever the replacement correspondence is a function. In particular, given a teamwork game involving a disjunctive task and under the conditions of the Proposition, we know by Corollary 3.13 that both the null component and the positive component of the replacement correspondence are single-valued, and thus $s_i(G)$ can also be decoupled in two components: a null one and a positive one defined over the same domains as $r_i(G)$. By definition of the share function, of G_i^* and of $\overline{G}_i$, we obtain the conditions that $s_i(G)$ must satisfy in the proposition.

To see that the positive component of s_i is an increasing function, we use Eq. (3.17) and write it in terms of the share function (Definition 2.6):

$$G = \Delta \frac{p_i \beta_i^{1/\rho}}{s_i(G)^{1/\rho}} - \frac{\sigma(G)}{\sigma'(G)} \frac{\alpha}{s_i(G)} \tag{A.5}$$

Which yields:

$$\frac{dG}{ds_i(G)} = \left(1 + \frac{\sigma'(G)^2 - \sigma(G)\,\sigma''(G)}{\sigma'(G)^2} \frac{\alpha}{s_i(G)}\right)^{-1} \frac{1}{s_i(G)} \left(\frac{\sigma(G)}{\sigma'(G)} \frac{\alpha}{s_i(G)} - \frac{p_i}{\rho} \frac{\beta_i^{1/\rho}\Delta}{s_i(G)^{1/\rho}}\right)$$

Given the properties of the evaluation function, it is straightforward to see that the first factor in this equation is positive for all G. Using the condition $\frac{\rho}{\Delta p_i} \geq \frac{\sigma'(\overline{G}_i)}{\sigma(\overline{G}_i)} \frac{\beta_i^{1/\rho}}{\alpha}$, and the fact that for disjunctive tasks $s_i(G)^{-1/\rho} < s_i(G)^{-1}$, we can check that the second factor is positive too. Hence, $s_i(G)$ is a strictly increasing function of G in the domain $\left[G_i^*, \overline{G}_i\right]$. Furthermore, $s_i(G)$ is the inverse of a continuous function (Eq. (A.5)) on a compact domain and therefore itself continuous. These observations allow us to adapt Proposition 6.1 in [Cornes and Hartley, 2007b] into Proposition 3.14.

□

A.7 Individual efforts in disjunctive and conjunctive tasks: tables

Table A.1: **Individual effort across different teams for a conjunctive task ($\rho = -10$) under three different assessment treatments**. For simplicity, we fill only the upper diagonal of the matrices, which are symmetric.

(a) Soft Assessment Treatment ($G^{\mathcal{T}} = 3$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	(28%, 28%)	(46%, 30%)	(53%, 26%)	(56%, 22%)
0.50		(48%, 48%)	(53%, 40%)	(55%, 34%)
0.70			(44%, 44%)	(46%, 37%)
0.90				(38%, 38%)

(b) Medium Assessment Treatment ($G^{\mathcal{T}} = 5$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	(29%, 29%)	(49%, 31%)	(56%, 27%)	(59%, 23%)
0.50		(57%, 57%)	(68%, 52%)	(72%, 45%)
0.70			(64%, 64%)	(67%, 55%)
0.90				(58%, 58%)

(c) Hard Assessment Treatment ($G^{\mathcal{T}} = 7$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	(29%, 29%)	(49%, 31%)	(56%, 27%)	(59%, 23%)
0.50		(57%, 57%)	(69%, 52%)	(73%, 46%)
0.70			(69%, 69%)	(77%, 64%)
0.90				(72%, 72%)

Table A.2: **Individual effort across different teams for a disjunctive task ($\rho = 10$) under three different assessment treatments**. For simplicity, we fill only the upper diagonal of the matrices, which are symmetric.

(a) Soft Assessment Treatment ($G^{\mathcal{T}} = 3$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	[(63%, 0%), (0%, 63%)]	[(63%, 0%), (0%, 60%)]	(0%, 50%)	(0%, 41%)
0.50		[(60%, 0%), (0%, 60%)]	[(60%, 0%), (0%, 50%)]	[(60%, 0%), (0%, 41%)]
0.70			[(50%, 0%), (0%, 50%)]	[(50%, 0%), (0%, 41%)]
0.90				[(41%, 0%), (0%, 41%)]

(b) Medium Assessment Treatment ($G^{\mathcal{T}} = 5$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	[(67%, 0%), (0%, 67%)]	[(67%, 0%), (0%, 78%)]	(0%, 71%)	(0%, 61%)
0.50		[(78%, 0%), (0%, 78%)]	(0%, 71%)	(0%, 61%)
0.70			[(71%, 0%), (0%, 71%)]	[(71%, 0%), (0%, 61%)]
0.90				[(61%, 0%), (0%, 61%)]

(c) Hard Assessment Treatment ($G^{\mathcal{T}} = 7$)

p_2 \ p_1	0.30	0.50	0.70	0.90
0.30	[(67%, 0%), (0%, 67%)]	[(67%, 0%), (0%, 80%)]	(0%, 84%)	(0%, 78%)
0.50		[(80%, 0%), (0%, 80%)]	(0%, 84%)	(0%, 78%)
0.70			[(84%, 0%), (0%, 84%)]	[(84%, 0%), (0%, 78%)]
0.90				[(78%, 0%), (0%, 78%)]

B Multi-Agent Sequential Teamwork

B.1 Implementation Details for Sequential Teamwork Games 237
B.2 Convergence Criterion for Empirical Payoff Estimation 239
B.3 Mapping of Sequential Social Dilemma Presence Across Task Types 240
B.4 Team Composition 240
B.5 Cooperativeness in Heterogeneous Teams 241
B.6 Moderators of Cooperativeness in Heterogeneous Teams 242
B.7 Proxies of Implicit Coordination: Tables 244

B.1 Implementation Details for Sequential Teamwork Games

This section documents the implementation choices, wrappers, and training utilities used in the experiments of Chapter 4. The full codebase is available at: https://github.com/aloaberasturi/TEMPO.

B.1.1 Environment Construction

The environment is implemented as a custom `PettingZoo` `ParallelEnv`. To enable efficient experience collection, the base environment is vectorised using `SuperSuit`:

- `pettingzoo_env_to_vec_env_v1` wraps the `ParallelEnv` into a vectorised interface.
- `concat_vec_envs_v1` concatenates multiple copies to allow simultaneous rollouts.

Online observation normalisation is implemented through a custom `NormalizeObservation` wrapper that maintains running means and variances. Rewards are standardised using an exponential moving average implemented in a `NormalizeReward` wrapper.

B.1.2 Training Loop

Training follows the standard PPO alternation between data collection and policy updates:

1. Collect batches of transitions (observations, actions, log-probabilities, rewards, and value estimates) across all parallel environments.
2. Compute advantages using GAE.
3. Update actor and critic networks via PPO's clipped surrogate objective.

Learning-rate annealing and KL-based early stopping are supported.

B.1.3 Attribute Shuffling Procedure

During training, the environment applies an optional *shuffle* operation when reset, randomly permuting agents' intrinsic attributes (p, α, β) at the start of every episode. These attributes are never observed directly—agents only receive task-level information and per-timestep contributions—so shuffling does not increase observability. Instead, it breaks any stable correlation between an agent's identity and its latent ability.

This mechanism serves two purposes. First, it prevents agents from *overfitting to structural roles*: without shuffling, an identity consistently associated with high expertise might learn to "lead", while a low-expertise identity might converge to persistent freeriding. Second, it ensures that learned policies remain *exchangeable*. The empirical game-theoretic analysis used later in Section 4.5.1 relies on sampling policies into arbitrary team compositions; shuffling exposes each agent to the full distribution of attribute assignments during learning, promoting robustness to heterogeneous team contexts.

In summary, attribute shuffling mitigates spurious specialisation, promotes generalisation, and aligns the training procedure with the methodological assumptions underlying the analyses presented in this chapter.

B.1.4 Neural Architectures

Each agent maintains an actor network (categorical policy), and a critic network (state-value function). Both networks share the same architecture:

- two hidden layers of 128 units,
- Leaky ReLU activations,
- Xavier-uniform weight initialisation,
- zero-initialised biases.

They only differ in the last layer: the actor's final layer produces unnormalised logits corresponding to the action space. During training, actions are sampled from this distribution; during evaluation, actions are selected greedily by taking the mode. The critic outputs a scalar estimate of the state value via a linear projection.

B.1.5 Logging and Diagnostics

Training statistics—including losses, entropy, KL divergence, and rollout throughput—are logged to `TensorBoard` and optionally integrates with `Weights & Biases`.

B.2 Convergence Criterion for Empirical Payoff Estimation

For each payoff value used in the Schelling diagram analysis, episodes are simulated sequentially until the estimate of the focal agent's return becomes sufficiently precise.

After each episode, let $R_1, R_2, \ldots, R_m$ denote the returns observed so far. The sample mean $\bar{R}_m$ and standard error SE_m are computed. A two-sided 95% confidence interval for the mean is constructed:

$$\bar{R}_m \pm t_{m-1,\,0.975}\,\mathrm{SE}_m$$

where $t_{m-1,\,0.975}$ is the standard t-value for a 95% interval with $m-1$ degrees of freedom.

Convergence is declared when the *half-width* of this interval,

$$\text{half-width}_n = t_{n-1,\,0.975}\,\mathrm{SE}_n$$

falls below the tolerance $\delta = 0.05$. This guarantees that the estimated mean return is accurate to within ± 0.05 with 95% confidence. If the threshold is not reached after 10 000 episodes, the procedure terminates and returns the current estimate.

B.3 Mapping of Sequential Social Dilemma Presence Across Task Types

B.4 Team Composition

Below are the expertise of the four agents in each team used throughout the simulations in Chapter 4. The top table contains information about **homogeneous** teams (low expertise dispersion, $\sigma(p) < 0.145$). The bottom table shows team composition in **heterogeneous** teams ($\sigma(p) > 0.145$). Expertise of agent i is denoted by p_i.

Table B.1: **Homogeneous teams** (4 agents per team).

ID	p_1	p_2	p_3	p_4	Mean ($\overline{p}$)
1	0.10	0.10	0.10	0.10	0.10
2	0.10	0.10	0.20	0.20	0.15
3	0.10	0.20	0.20	0.30	0.20
4	0.10	0.20	0.30	0.30	0.23
5	0.20	0.30	0.30	0.40	0.30
6	0.20	0.30	0.40	0.50	0.35
7	0.30	0.40	0.40	0.50	0.40
8	0.30	0.40	0.50	0.60	0.45
9	0.40	0.50	0.50	0.60	0.50
10	0.40	0.50	0.60	0.70	0.55
11	0.50	0.60	0.60	0.70	0.60
12	0.50	0.60	0.70	0.80	0.65
13	0.60	0.70	0.70	0.80	0.70
14	0.60	0.70	0.80	0.90	0.75
15	0.70	0.80	0.80	0.90	0.80
16	0.70	0.80	0.90	1.00	0.85
17	0.80	0.90	0.90	1.00	0.90
18	0.90	1.00	1.00	0.90	0.95
19	1.00	1.00	1.00	1.00	1.00

Table B.2: **Heterogeneous teams** (4 agents per team).

ID	p_1	p_2	p_3	p_4	Mean ($\overline{p}$)
1	0.10	0.10	0.20	0.70	0.28
2	0.10	0.20	0.20	0.60	0.28
3	0.10	0.10	0.40	0.50	0.28
4	0.10	0.10	0.10	0.80	0.28
5	0.10	0.10	0.10	0.90	0.30
6	0.10	0.10	0.10	1.00	0.33
7	0.10	0.10	0.90	0.90	0.50
8	0.10	0.50	0.50	0.90	0.50
9	0.10	0.20	0.80	0.90	0.50
10	0.20	0.50	0.50	0.80	0.50
11	0.10	0.40	0.60	0.90	0.50
12	0.20	0.20	0.90	0.90	0.55
13	0.20	0.70	0.80	0.90	0.65
14	0.30	0.70	0.80	1.00	0.70
15	0.10	0.90	0.90	0.90	0.70
16	0.20	0.80	0.80	1.00	0.70
17	0.30	0.90	0.90	0.90	0.75
18	0.10	1.00	1.00	1.00	0.78
19	0.20	1.00	1.00	1.00	0.80

B.5 Cooperativeness in Heterogeneous Teams

This appendix shows how heterogeneous teams in Sequential Teamwork Games converge to different cooperativeness patterns depending on their average expertise and task type. It supplements the analysis of Section 4.5.2 on homogeneous teams. Similar to that case, *cooperativeness* is measured here as the—per turn—average percentage of time that an agent spends working on the time. As such, it is a measure of effort—and not of productivity. This allows decoupling cooperative willingness from expertise. Figure B.2 shows how task interdependence (ρ) and the group's average expertise ($\overline{p}$) shape cooperative behaviour when teams are **heterogeneous** in ability.[1]

Conjunctive tasks. The presence of a weak link dominates. In the experimental grid, heterogeneous teams rarely (or never) activate cooperation nor reach the evaluation threshold, consistent with bottom-down coordination around the least capable member. This mirrors the

[1]Unless otherwise stated, heterogeneity refers to teams with expertise dispersion above the threshold used in the main text (*cf.* Sec. 4.5.2); difficulty is set to *medium*, leisure preferences are fixed at $\alpha_i = 1\ \forall i$, and $\gamma = 0.9$. Colours denote average team effort (time on task) aggregated across timesteps; dark blue indicates collective inaction, while green-yellow reflects increasing engagement.

homogeneous-case pattern but is amplified by dispersion in expertise.

Additive tasks. Heterogeneous teams activate at similar $\overline{p}$ as homogeneous teams. Conditional on being active, their *mean* effort is qualitatively akin to that of homogeneous teams.

Disjunctive tasks. In terms of activation, heterogeneous teams benefit most under disjunctive interdependence: the transition from collective inaction to active engagement occurs at lower $\overline{p}$ than in homogeneous teams, and—once active—effort remains stable. This aligns well with the general narrative throughout the chapter: high-skill outliers tend to shoulder the task while others conserve leisure.

Across task types, average contributions differ significantly ($p < 0.001$)[2]. In sum, ρ remains a structural driver of cooperation: disjunctive tasks elicit volunteer's dilemma dynamics (few carry many), additive tasks sustain intermediate, shared effort once activated, and conjunctive tasks suppress cooperation under ability dispersion due to the weakest-link constraint.

As per *intra-group effort dispersion*, the overall pattern resembles that observed in homogeneous teams, with the caveat that in conjunctive tasks there is no active regime—and thus there are no data to compute differences with the other conditions. On average, additive tasks exhibited greater variability than weakly conjunctive (180.8%, p-value < 0.001); and lower variability than weakly disjunctive (-16.8%, p-value < 0.001) and strongly disjunctive tasks (-10.2%, p-value < 0.001). Interestingly, dispersion in weakly disjunctive tasks was higher than in strongly disjunctive ones ($+7.9\%$, p-value < 0.05).

B.6 Moderators of Cooperativeness in Heterogeneous Teams

This appendix complements the analyses in Section 4.5.2, focusing on how the preference for leisure (α) and the task difficulty—operationalised through the passing threshold ($G^{\mathcal{T}}$)—moderate cooperative behaviour in *heterogeneous* teams.

[2]All contrasts reported in this appendix rely on Welch's t-test ($p < 0.05$) applied to replicate runs per condition. Global summaries pool across active cells, and intra-team variability is assessed analogously using the standard deviation of contributions.

Preference for Leisure (α). Figure B.3 shows how α modulates cooperation as a function of expertise. Results correspond to medium-difficulty tasks, with heterogeneous teams, a fixed discount factor ($\gamma = 0.9$), and a uniform leisure preference across agents ($\alpha_i = \alpha \in \{0.5, 1.0, 1.5\}$). As in the main text, the visualisation is restricted to the three paradigmatic task structures: conjunctive, additive, and disjunctive.

Overall, the patterns mirror those reported for homogeneous teams (see Fig. 4.9), with some caveats:

1. **Conjunctive tasks.** Heterogeneous teams performing medium-difficulty conjunctive tasks *never* activate. The persistent weakest link renders the passing threshold unattainable, causing agents to learn that effort is futile. This is consistent with the collapse-of-cooperation phenomenon discussed in the main text.

2. **Additive tasks.** As in the homogeneous case, higher preference for leisure leads to *less* and *later* cooperation, and effort decays as teams become more skilled. The resulting curves retain the characteristic monotonic decrease in effort with increasing $\overline{p}$.

3. **Disjunctive tasks.** The delay-and-inhibit pattern is less pronounced. Heterogeneous teams often maintain nearly constant—and occasionally slightly increasing—effort as expertise rises. This aligns with the intuition from the heterogeneous heatmaps (Fig. B.2), where skilled volunteers can stabilise team performance.

Passing Threshold ($G^{\mathcal{T}}$). The moderating effect of evaluation hardness is more nuanced in heterogeneous teams than in homogeneous ones:

1. **Conjunctive tasks.** When conjunctive tasks are *easy*, only sufficiently skilful teams activate, for the coordination reasons discussed earlier. Even in these cases, effort patterns are noticeably less stable than in homogeneous teams: some configurations consistently fail despite being technically capable of passing. This suggests that iPPO agents face greater learning challenges in conjunctive settings, likely due to sparse and bottleneck-driven rewards.

2. **Additive tasks.** Increasing difficulty reliably delays activation and, once active, induces higher effort levels. In this sense, α and $G^{\mathcal{T}}$ produce analogous activation dynamics but push behaviour in opposite directions: higher α discourages cooperation, whereas higher thresholds demand it.

3. **Disjunctive tasks.** The qualitative patterns resemble those observed for additive tasks, though the shifts are less crisp. Once a competent agent can ensure success, the sensitivity to $G^{\mathcal{T}}$ diminishes.

Across all task types, once heterogeneous teams enter their active regime, they tend to sustain *stable effort levels* for a given passing threshold—that is, average effort does not clearly decline as $\overline{p}$ increases. The corresponding panels are in Fig B.4

B.7 Proxies of Implicit Coordination: Tables

Table B.3: **Anticipation proxy β_1 by task type**. Values are mean $\pm$ s.d. across seeds.

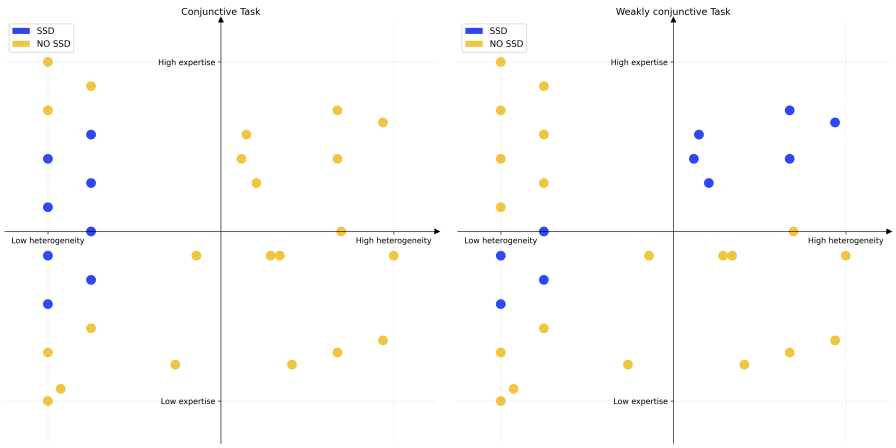
Task type	β_1 **(mean $\pm$ s.d.)**
Conjunctive	0.277 ± 0.033
Additive	0.209 ± 0.056
Disjunctive	0.031 ± 0.019

Table B.4: **Dynamic adjustment proxy $\overline{D}$ by task type**. Values are mean $\pm$ s.d. across seeds.

Task type	$\overline{D}$ **(mean $\pm$ s.d.)**
Conjunctive	0.894 ± 0.016
Additive	0.441 ± 0.035
Disjunctive	0.343 ± 0.017

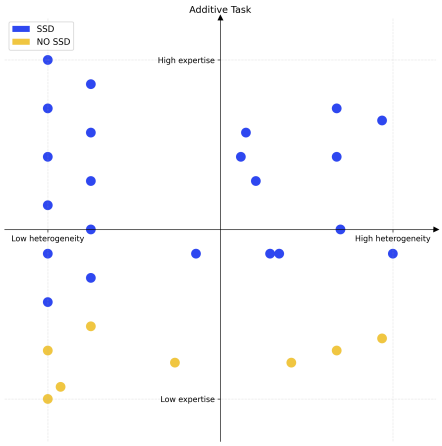
Table B.5: **Role specialisation proxy** (NMI) **by task type**. Values are mean $\pm$ s.d. across seeds.

Task type	**NMI (mean $\pm$ s.d.)**
Conjunctive	$0.00241 \pm 9\exp -5$
Additive	$0.0978 \pm 9\exp -4$
Disjunctive	0.53 ± 0.02

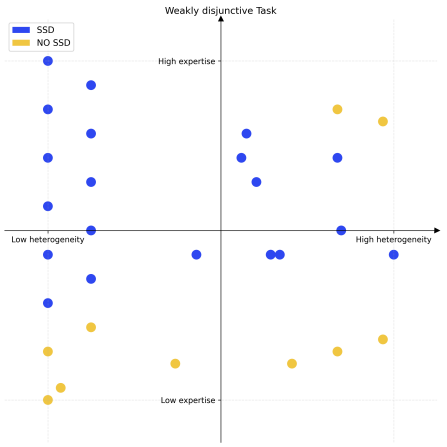


(a) Conjunctive

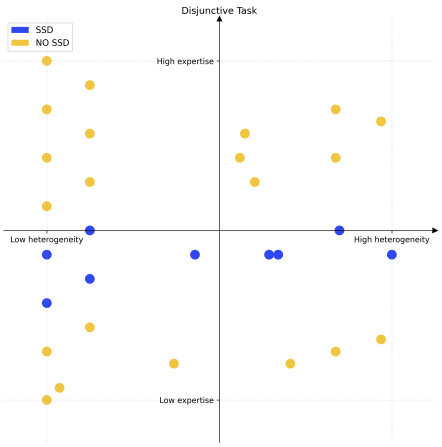
(b) Weakly Conjunctive



(c) Additive



(d) Weakly Disjunctive



(e) Disjunctive

Figure B.1: Mapping of Sequential Social Dilemma Presence Across Task Types

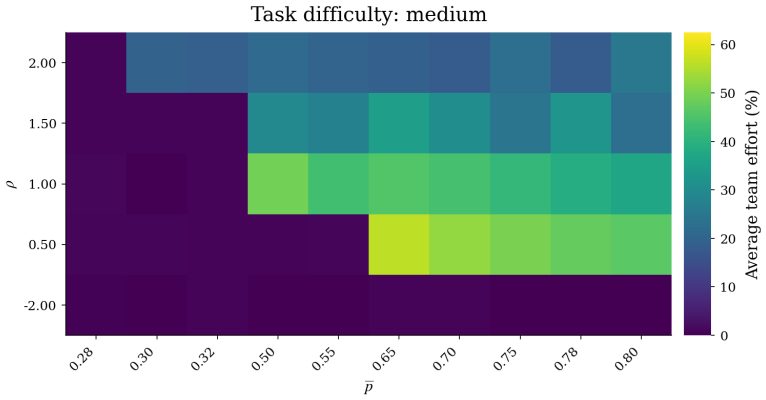


Figure B.2: **Average team effort (percentage of time) in heterogeneous teams** across task interdependence (ρ) and average expertise $\overline{p}$ (difficulty =medium, $\gamma = 0.9$, $\alpha_i = 1\ \forall i$). Dark regions denote collective inaction; green-yellow indicates increasing engagement.

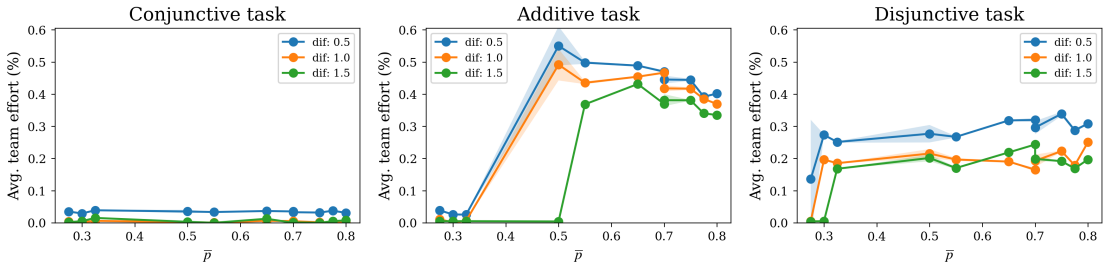


Figure B.3: **Effect of the preference for leisure (α) on effort–expertise curves** across task types (heterogeneous teams, medium difficulty, uniform α across agents).

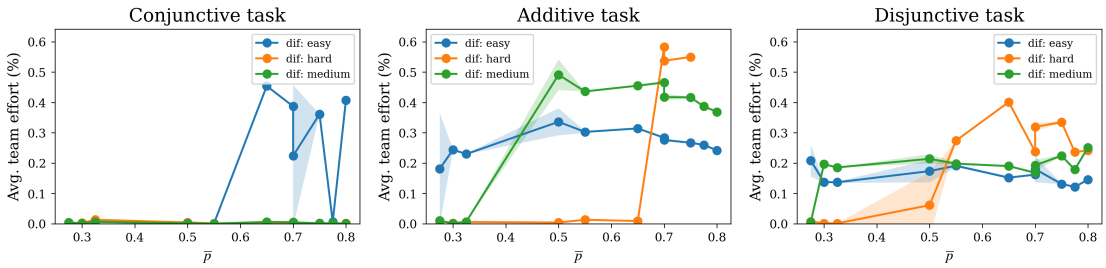


Figure B.4: **Effect of the passing threshold ($G^{\mathcal{T}}$) on effort-expertise curves** across task types (heterogeneous teams, uniform $\alpha = 1$ across agents)

C The Island

C.1 Introduction 250
C.2 Game Design 251
C.3 Conclusions 262

C.1 Introduction

This appendix presents an interactive platform developed to generate and record human behavioural data in tasks structurally analogous to those performed by multi-armed bandits in *CollabQuest* (Chapter 3) and by independent Proximal Policy Optimization (iPPO) agents in *TEMPO* (Chapter 4). The platform—henceforth *The Island*—comprises three interactive mini-games.

When logging in, each player on *The Island* is assigned a fixed avatar with immutable characteristics—specifically, a predetermined level of ability (or expertise) in each mini-game. All mini-games share the same duration, which can be configured by the experimenter. From the player's perspective, the interaction remains constant across games: at each turn, she chooses—via a clickable bar—how many resources (time) to keep for herself and how many to contribute to the team. The details of this decision process and its consequences are outlined in the sections that follow.

The Island was conceived as an empirical bridge between the theoretical and simulation-based frameworks introduced in Chapters 3 and 4. Although still a pilot, the platform has the potential to help close the loop between formal models and observed human behaviour as it is refined and empirically validated in future iterations. In particular, it enables systematic comparisons between the learned policies of artificial agents and the decisions of human participants within a controlled, easily customisable environment.

These comparisons, in turn, can inform improvements to the models and simulations underlying *CollabQuest* and *TEMPO*, advancing our understanding of unenforced cooperation and the mechanisms that align individual and collective goals in teamwork settings. In this way, *The Island* embodies the continuous feedback loop between theory, simulation, and empirical observation that underpins the methodological approach of this dissertation.

C.2 Game Design

C.2.1 Platform Architecture

The Island runs as a web-based client-server application built mainly in Python and JavaScript. The server, written in Python, handles all the internal logic of the game: it manages user logins, forms teams, updates scores, and stores every action in a local SQLite database. The visual interface that players interact with is written in JavaScript and runs directly in their web browser, communicating with the server in real time through WebSocket connections. To make the platform easy to install and run on any computer, both the server and the client are packaged inside a Docker container, which includes all the necessary software dependencies. During development, changes to the code or database can be made without rebuilding the container. In production, the platform runs on the IIIA's[1] Kubernetes server, which automatically launches the Docker images and keeps the database stored safely between sessions.

C.2.2 User Interaction and Interface

Players gain access through a pre-assigned username and password. Upon logging in, they can select their preferred language from English, Spanish, or Catalan (Fig. C.1).

After logging in, players are presented with the general instructions for the game (Fig. C.2). These remain on screen until the user presses the CONTINUE button.

Free players—*i.e.*, those who are online but not currently playing because they have either just logged in or have recently finished a mini-game—are added to a pool and then assigned to random teams of four via a queue system, detailed in the following section.

While a player is waiting for their team to be formed on the fly, a `FORMING TEAM` screen is displayed (see Fig. C.3), which reminds them of the following:

- The number of games they have already played and those remaining.
- Their accumulated score.
- The immutable abilities—or *level*, in the game's terminology—of their avatar in each of the three mini-games.

[1] IIIA stands for Artificial Research Institute—*Instituto de Investigación en Inteligencia Artificial* in Spanish.

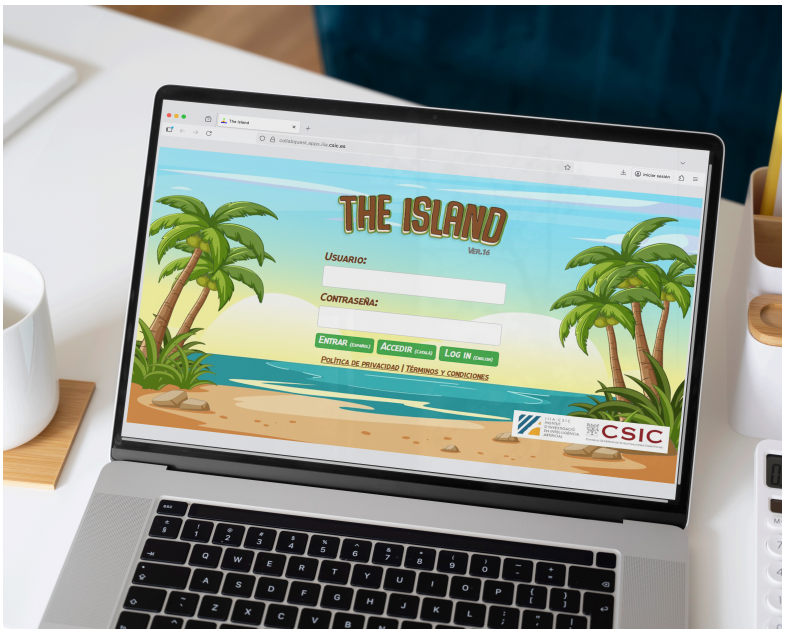


Figure C.1: Login screen of *The Island*. Players access the platform using a pre-assigned username and password, and can select their preferred language (English, Spanish, or Catalan) before proceeding.



Figure C.2: Instruction screen presented to players after logging in. The general rules and objectives of the game are displayed until the user confirms understanding by pressing the `CONTINUE` button.

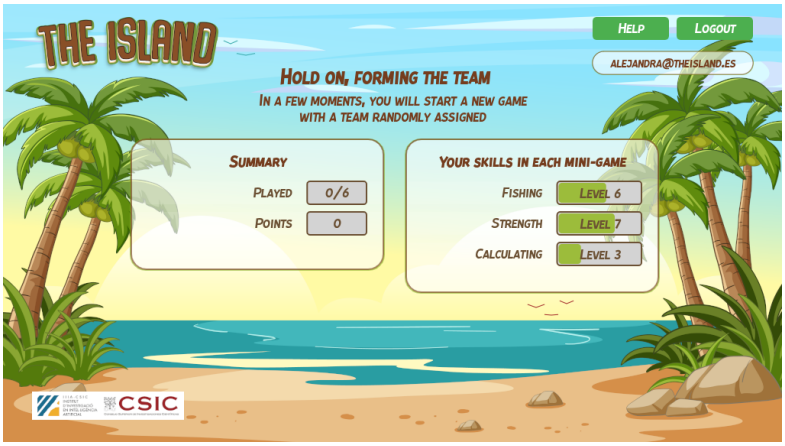


Figure C.3: Forming team screen. While waiting for a team to be formed on the fly, players are reminded of their progress so far and their avatar's levels in each mini-game.

Once all four players have acknowledged the instructions by pressing CONTINUE, a new mini-game begins. Players are anonymous at all times and do not know the identities of their teammates. To mitigate the risk of retaliation, team composition is re-randomised between mini-games. Participants are notified of this beforehand.

At any point, all players have access to two additional buttons:

- The HELP button opens a direct chat to send messages to the platform administrator.
- The LOGOUT button logs the player out—even in the middle of a game.

C.2.3 *Player Assignment and Team Formation*

Team formation in *The Island* is handled automatically by the server through an asynchronous game loop. At regular intervals, the server checks which players are connected and waiting to play, and groups them into teams of four whenever enough participants are available for the same mini-game. The matching algorithm ensures that all team members have the same upcoming task in their queue. To avoid learning biases due to task ordering, the algorithm is designed to prioritise a random sequence in which the three mini-games are presented to each participant. Once a new team is created, a corresponding entry is added to the database, each player's pending task list is updated, and all team members are simultaneously moved to the `FORMING TEAM` state. When everyone is ready, the server transitions the group to the active game phase and creates the first turn. Throughout the session, the same loop also monitors

disconnections, finalises completed tasks, and reassigns players whose teammates have left, ensuring that the experiment proceeds smoothly without manual supervision.

C.2.4 Administrator Dashboard

In addition to the `player` class, the game includes an `admin` class. Administrators have access to a dedicated dashboard from which they can monitor key aspects of ongoing games and, if necessary, intervene to control their flow. The dashboard is organised into six tabs, arranged from left to right as follows:

- The SUMMARY tab provides a summary of current activity on the platform, showing the total number of active players and ongoing games.
- The PLAYERS tab (Fig. C.4) lists all registered players, along with key information such as their accumulated score, connection status (`connected` / `disconnected`), the state of their most recent—or current—mini-game, the number of mini-games completed, and the queue of remaining mini-games before reaching the final screen. Administrators can revoke a player's access by changing their access state from `granted` to `ungranted`.
- The TASKS tab displays all ongoing and completed mini-games, indicating the duration of each game and the players involved.
- The CHAT tab manages communication between players and the administrator. Just as players can contact the administrator at any time, the admin can reply directly or broadcast messages to all players simultaneously.
- The TERMINAL tab provides live access to detailed logs of all ongoing mini-games, effectively functioning as a debugger console.
- Finally, the ACTIONS tab manages the database in which all player actions—and their corresponding timestamps—are stored. From this tab, administrators can create backup copies, download data, or delete existing records.

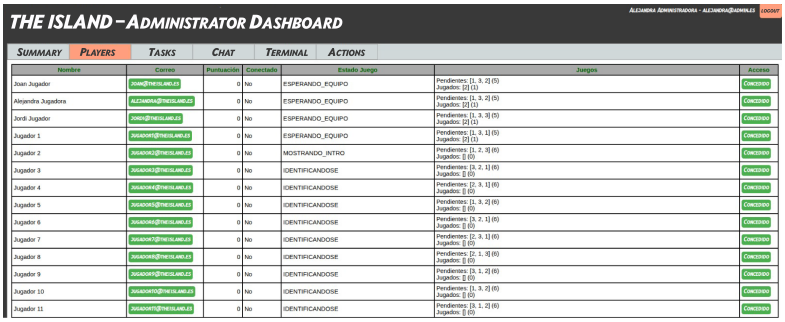


Figure C.4: Administrator Dashboard. The administrator class has access to detailed game and player data, including current activity on the platform, currently connected players, and pending games. It is also possible to create backup copies, download or delete the database and broadcast messages to all the players at once.

C.2.5 Game Logic

Across the three mini-games, the user experience is intentionally consistent: each involves a decision about how many resources—time—the player wishes to allocate to the team and how many to keep for herself. What changes between games is the way in which the resources contributed by the team are combined, and therefore—as explained in Subsection C.2.6—the optimal strategy for each player.

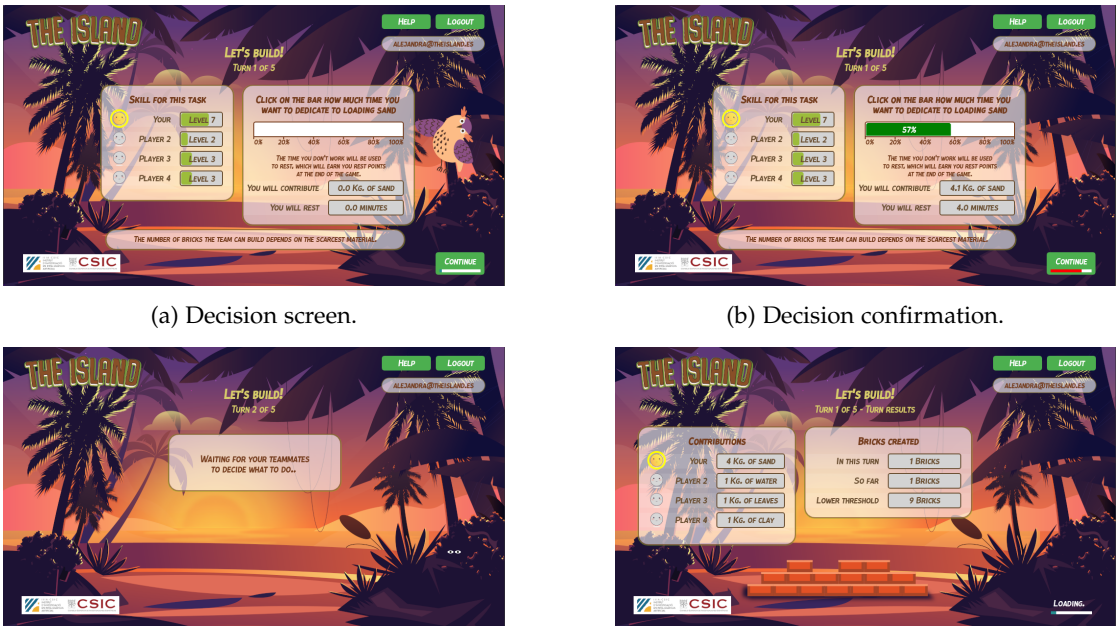
Figure C.5 illustrates the sequence of screens that a player encounters during a mini-game. Each mini-game has a fixed duration—five turns in this example. At each turn, the player clicks on a white selection bar (Fig. C.5a) to indicate how much time she wishes to dedicate to the task. Depending on the position of the pointer on the bar, a specific contribution is displayed below it (Fig. C.5b). Once satisfied with her choice, the player presses the CONTINUE button—unless the timer runs out, in which case zero resources are assigned to the group and the game proceeds to the next turn.

While waiting for the other team members to make their choices, the player sees a waiting screen (Fig. C.5c). Once all players have decided, a results screen appears (Fig. C.5d), showing:

- Each player's contribution for that turn (left panel).
- How the collective resources are combined to produce a joint outcome (*Bricks created in this turn*).
- The team's cumulative progress so far (*Bricks created so far*).

- The minimum collective requirement to complete the mini-game successfully (*Lower threshold*).

The details of how these contributions are aggregated, how the lower threshold is computed, and the consequences of meeting or failing to meet it are explained in the next subsection. Optionally, the information about other players' individual contributions can be hidden, so that only the ego-player's own action and the collective outcome are displayed. This setting allows the experimenter to manipulate the level of observability and examine its effect on cooperative behaviour.



(a) Decision screen.

(b) Decision confirmation.

(c) Waiting screen.

(d) Turn results screen.

Figure C.5: Sequence of screens presented to a player during a mini-game in *The Island*, illustrating the decision, waiting, and feedback stages. **(a): Decision screen.** The player selects how much time to dedicate to the task by clicking on the white bar. Her skill level and those of the other players are displayed on the left. **(b): Decision confirmation.** Once the bar has been set, the corresponding contribution and rest time are shown below. **(c): Waiting screen.** After confirming her decision, the player waits while her teammates make their own selections. **(d): Turn results screen.** Once all decisions are made, the team's individual and collective outcomes for the turn are displayed, together with cumulative progress and the lower threshold.

Optionally, information about other players' individual contributions can be hidden, so that only the ego-player's own action together with the collective outcome is displayed. This allows the experimenter to manipulate within-team transparency.

C.2.6 *Mini-game Structure and Scoring*

Each of the three mini-games on *The Island* corresponds to one of Steiner's canonical task types [Steiner, 1972]: additive, conjunctive, and disjunctive. Although their visual interface and player interaction are identical, the way in which individual contributions are aggregated at the team level varies according to the task type.

Avatar skills. Every player controls an avatar with fixed, immutable skill levels for each mini-game. These levels are displayed to the player at all times (see Fig. C.3) and act as multiplicative factors determining the efficiency of their actions in each task. On every turn t, the player allocates a fraction $a_i^t \in [0,1]$ of her available time to the team task and the rest to rest. The individual contribution g_i^t is computed as:

$$g_i^t \;=\; p_i^{\mathcal{T}} \times a_i^t \times 10$$

where $p_i^{\mathcal{T}} \in [0,1]$ denotes the avatar's skill in mini-game $\mathcal{T}$ and the constant 10 represents the turn duration (chosen for display convenience). Thus, contributions scale linearly with both skill and invested time.

Aggregation rules. At the end of each turn, the four players' contributions $\{g_i^t\}_{i=1}^4$ are aggregated through an aggregator function $\mathrm{Agg}(\cdot)$ into a team output:

$$\mathrm{Agg}(\boldsymbol{g^t}) = \begin{cases} \sum_i g_i^t & \text{additive: } \textit{Fishing}; \\ \min_i g_i^t & \text{conjunctive: } \textit{Building the Shelter}; \\ \max_i g_i^t & \text{disjunctive: } \textit{Navigation}. \end{cases}$$

Team outcome is initialised as zero, $G^0 := 0$, and evolves similarly to the task evolution described by Eq (4.2) in Chapter 4:

$$G^{t+1} = G^t \;+\; \mathrm{Agg}(\boldsymbol{g^t}), \qquad t = 0, \ldots, H-1 \tag{C.1}$$

Where H is the number of turns. Optionally, information about other players' individual contributions can be hidden, so that only the ego-player's own action and the collective outcome are displayed. This feature allows the experimenter to manipulate the level of observability of the game.

Threshold and collective success. Each mini-game defines a minimum collective performance threshold that determines group success. In parallel to the team's actual cumulative output G^t, the platform maintains an attainable bound M^t representing the maximum cumulative outcome achievable by a reference average-skill team. Both evolve over the same discrete time steps $t = 0, \ldots, H$, following a linear rule:

$$M^{t+1} = M^t + m, \quad \text{with} \quad M^0 = 0$$

The per-turn ceiling m is initialised as:

$$m = \begin{cases} 6 \times \texttt{NUM_TEAM_PLAYERS} & \text{additive (Fishing)}, \\ 6 \times 1 & \text{conjunctive (Building)}, \\ 6 \times 1 & \text{disjunctive (Navigation)}. \end{cases}$$

Here, the constant 6 represents the expected contribution of an avatar with average skill (*i.e.*, 0.6×10 for a single turn of duration 10). Thus, in the additive case (Fishing), the per-turn ceiling scales with team size, whereas in the conjunctive and disjunctive cases it depends only on a single representative player (the limiting or leading one, respectively).

At the end of the final turn H, the team's cumulative output G^H is compared with a fixed fraction of the attainable maximum M^H. The team passes if

$$G^H \;\geq\; \alpha\, M^H$$

where $\alpha \in (0, 1)$ is the assessment threshold. Otherwise it fails. This binary gating acts as a Heaviside step on the normalised team performance.

Individual scoring. Within each mini-game, players accumulate *rest points* proportional to the fraction of unused time $(1 - a_i^t)$ at each turn. These are summed to obtain $R_i = \sum_{t=1}^{H} 10 \times (1 - a_i^t)$ for the mini-game, where 10 is turn duration.

Let $\mathbb{H} : \mathbb{R} \to \{0, 1\}$ be the Heaviside step function,

$$\mathbb{H}(x) \;=\; \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

and define the team success indicator at the end of the game as

$$\mathbb{H}(G^H - \alpha M^H)$$

The player's final score P_i is then:

$$P_i = \mathbb{H}(G^H - \alpha M^H) \times R_i$$

Thus, if $\mathbb{H}(G^H - \alpha M^H) = 0$, all players lose the rest points accumulated during that mini-game.

C.2.7 *Mini-game Descriptions*

Each mini-game instantiates one of Steiner's canonical task types [Steiner, 1972], with identical player interaction but different aggregation rules at the team level.

Fishing (Figure C.6). Players decide what proportion of their turn to dedicate to fishing. Individual contributions are converted into "fish" in proportion to the player's task-specific skill. Team output for the turn is the *sum* of individual contributions—and thus the task is **additive**.



Figure C.6: Fishing mini-game (**additive task**). Each player decides what fraction of her time to devote to fishing. Individual catches depend on the avatar's fishing skill, and the team's total catch in each turn is the sum of all individual contributions.

Building the shelter (Figure C.7). Players allocate time to loading construction materials. Team output for the turn is the *minimum* of the four individual contributions, capturing the **conjunctive** nature of the task (the scarcest material limits production).

Navigation (Figure C.8). Players devote time to calculations for navigation. Team output for the turn is the *maximum* individual contribution, reflecting a **disjunctive** task in which the best individual performance determines the team outcome.

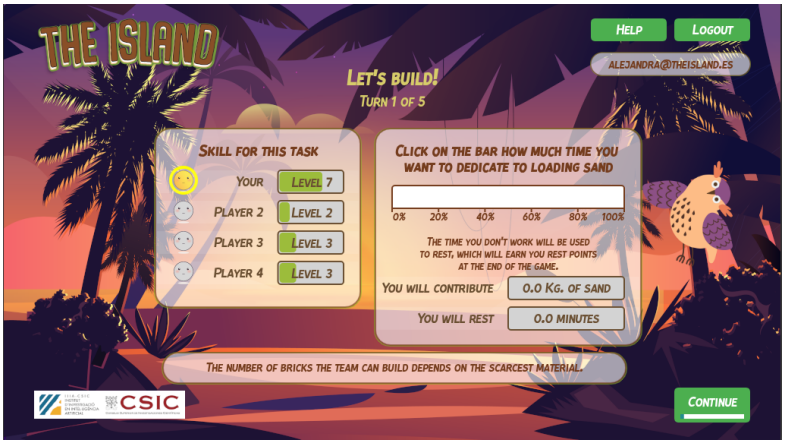


Figure C.7: Building the shelter mini-game (**conjunctive task**). Players allocate time to carrying building materials. The team's progress in each turn is limited by the smallest individual contribution, illustrating the conjunctive rule where collective output depends on the least-performing member.

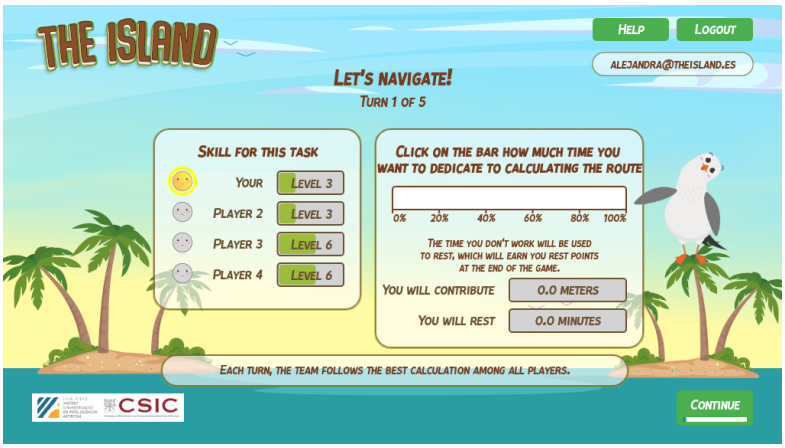


Figure C.8: Navigation mini-game (**disjunctive task**). Players devote time to performing calculations that help the team navigate. The team's progress each turn is determined by the highest individual contribution, reflecting a disjunctive task in which the most capable member drives team success.

C.2.8 Database

All experimental data are stored in an SQLite database. The schema captures the full hierarchy of the game logic, linking users, their avatars, group tasks, and turn-by-turn actions. The structure is summarised below:

- **Users** — stores participant login information, validation status, and session state. In-

cludes fields such as *email*, *estado* (current screen or phase), and *bot* to distinguish automated agents.

- **Players** — extends user data with immutable avatar parameters: skill levels (*fishing skills*, *building skills*, *navigating skills*), cumulative score, bonus, and task order.
- **Tasks** — represents each team mini-game instance, recording its type (additive, conjunctive, or disjunctive), current turn, accumulated group output, and performance thresholds.
- **Turns** — stores player actions within each task: contribution percentage, effective team input (*para_el_grupo*), rest points, and response times.
- **Player-Task** — linking table defining the many-to-many relation between players and tasks.
- **Questionnaires** — records demographic and self-reported background information collected at registration.
- **Chats** — logs player-administrator messages exchanged during the session.
- **Logs** — stores system-level debug and administrative events for experiment monitoring.

This schema enables complete reconstruction of game dynamics and facilitates both real-time supervision and post-experiment statistical and data analysis. Its level of granularity also allows the use of the same analytical techniques applied to the learned policies in Chapter 4.

C.2.9 Pilot test

Status and scope. The Island is currently a proof-of-concept platform. The purpose of the pilot deployment was to validate usability and instrumentation and to collect exploratory behavioural traces; it was not designed to provide statistically conclusive tests of the models' descriptive accuracy.

A brief pilot study was conducted to validate usability and to collect preliminary behavioural traces across the three task types. The study involved $N = 39$ participants, each playing an average of 3.38 games[2], with each game consisting of $H = 4$ turns. The pilot primarily served to assess usability and to support exploratory measurement.

[2]Game counts varied by participant, as new sessions began only when four players were matched in the queue.

Using the same proxies defined in Chapter 4, Subsection 4.5.3—team *anticipation* (β_1), *dynamic adjustment* ($\overline{D}$), and *role specialisation* (NMI)—we observed qualitative patterns consistent with the simulations: conjunctive tasks elicited stronger (albeit modest) coordination signals, while disjunctive tasks showed greater role differentiation. As expected, estimates were noisy due to the exploratory nature of the pilot and the fact that teammates' skills were visible in this version (which plausibly amplifies specialisation in disjunctive tasks). Accordingly, these results are reported as indicative rather than inferential.

Table C.1: **Summary of pilot coordination proxies** (median across episodes). Anticipation is the coefficient β_1 from the centred regression; $\overline{D}$ is the fraction of weighted negative shocks compensated; NMI indexes role specialisation ($\in [0, 1]$).

Task type	**Anticipation** (β_1)	**Adjustment** ($\overline{D}$)	**NMI**
Conjunctive (Building)	0.51	0.10	0.06
Additive (Fishing)	-2.17	0.01	0.07
Disjunctive (Navigation)	0.16	0.33	0.11

Note. Large variability (not shown) possibly reflects small sample size and early learning effects. Results are used here solely to illustrate that *The Island* reproduces the structural conditions needed to study emergent teamwork dynamics in human participants.

C.3 Conclusions

The Island was conceived as an empirical bridge between the theoretical and simulation-based frameworks developed in this dissertation. It provides a controlled and experimentally tractable environment in which human teamwork behaviour can be studied under the same structural constraints as those faced by artificial agents in *CollabQuest* and *TEMPO* (Chapter 3 and 4).

The game's logic and the platform's fully instrumented data collection make it suitable for fine-grained analysis of unenforced cooperation in the three canonical task types—conjunctive, additive and disjunctive. From the point of the experimenter, the platform is flexible by design—including configurable parameters, such as the observability of teammates' actions, the number of turns of each mini-game or the incentive schemes used to reward the team. We expect that such flexibility will enable systematic experimental manipulation of factors central to theories of collective intelligence and cooperation.

Preliminary pilot tests confirmed that the platform is functional and capable of eliciting interpretable behavioural patterns that qualitatively mirror the coordination profiles predicted by the simulations. Although these results remain exploratory, they demonstrate that the platform successfully reproduces the structural conditions required to study teamwork phenomena.

Next iterations of *The Island* will focus on scaling data collection to mitigate sample size effects and enable statistically sound inference, as well as refining the interface to facilitate experimental replication. In doing so, it will continue to serve as a key empirical component in the broader feedback loop between theory, simulation, and observation that underpins the methodological approach of this thesis. At present, it should be regarded as a pilot infrastructure rather than as empirical validation of the models.

D Agents, Firms and Social Identity

D.1 Derivations for the Identity-Salience Comparative Statics 266
D.2 Derivations for the Firm-Size Comparative Statics 269
D.3 Detailed computation of $\frac{\partial V_i^{SCT}}{\partial n}$ 271
D.4 Justification of Proposition 5.7 272
D.5 Simulation Kernel 274
D.6 Size and Utility in a Degenerated Case 275
D.7 Scenario A (i.i.d. Traits): Endogenous Identity without Network Homophily 276
D.8 Scenario B (Exogenous Identity): Salience Control via Sampled MCR 277

D.1 Derivations for the Identity-Salience Comparative Statics

This section details the intermediate algebraic steps underlying the derivation of Eq. (5.10) and the sign analysis of its denominator, as referenced in Section 5.5.1.

A. Derivation of $\frac{\partial F}{\partial e_i}$

Starting from

$$V_i^{\text{SCT}} = (1-\phi_i)\left[\alpha_i \ln y_i + (1-\alpha_i)\ln(1-e_i)\right] + \phi_i\left[\widetilde{\alpha}\ln\widetilde{y} + (1-\widetilde{\alpha})\ln(1-\bar{e}_{-i})\right]$$

and imposing equal sharing $y_i = \widetilde{y} = O(E)/n$ with total firm output $O(E) = aE + bE^{\beta}$ and $E = \sum_{j\in f} e_j$, the first-order condition is

$$F(e_i, \phi_i) = \frac{\partial V_i^{\text{SCT}}}{\partial e_i} = \frac{O'(E)}{O(E)}\left[(1-\phi_i)\,\alpha_i + \phi_i\,\widetilde{\alpha}\right] - (1-\phi_i)\frac{1-\alpha_i}{1-e_i} = 0$$

This expression has a natural interpretation as a balance between the marginal benefit and the marginal cost of effort.

Marginal benefit of effort. The first term,

$$\frac{O'(E)}{O(E)}\left[(1-\phi_i)\,\alpha_i + \phi_i\,\widetilde{\alpha}\right]$$

captures the marginal benefit of increasing e_i. The factor $\frac{O'(E)}{O(E)} > 0$ reflects how responsive output[1] is to changes in total effort E: when technology exhibits strong returns to effort or the firm is relatively small, a marginal increase in E translates into a larger proportional increase in income. The bracket

$$(1-\phi_i)\,\alpha_i + \phi_i\,\widetilde{\alpha}$$

is the effective marginal valuation of income under a hybrid evaluative frame. When ϕ_i is close to zero, the agent evaluates additional income almost entirely through her own preference parameter α_i. As salience rises and ϕ_i approaches one, this valuation shifts towards the firm-level normative parameter $\widetilde{\alpha}$. In other words, identity salience reweights the marginal benefit of effort away from purely individual preferences and towards the group's prototypical values.

Marginal cost of effort. The second term,

$$(1-\phi_i)\frac{1-\alpha_i}{1-e_i}$$

[1] And per-capita income under equal sharing.

represents the marginal cost of increasing e_i. The fraction $\frac{1-\alpha_i}{1-e_i}$ is the marginal disutility of effort in the individual frame. When effort is already high (so $1 - e_i$ is small), an additional unit of effort demands a steep loss of leisure and thus a large marginal utility cost. When effort is low, the same increase in e_i is comparatively less burdensome. This cost is multiplied by $(1 - \phi_i)$ because it arises only from the individual frame of evaluation. As identity salience increases and the social frame becomes more dominant (higher ϕ_i), the relative weight on this individual cost term diminishes: the agent's perceived marginal cost of effort becomes less anchored in her own lost leisure and more in the group norm against which behaviour is evaluated.

To simplify the sign analysis of $\frac{\partial F}{\partial e_i}$, it is useful to express the ratio $\frac{O'(E)}{O(E)}$ in elasticity form[2]. For the production function

$$O(E) = aE + bE^{\beta}$$

the elasticity of total output with respect to total effort is

$$z_3(E) \stackrel{\text{def}}{=} \frac{\partial O}{\partial E} \cdot \frac{E}{O(E)} = \frac{O'(E)E}{O(E)}$$

Equivalently,

$$\frac{O'(E)}{O(E)} = \frac{z_3(E)}{E}$$

so $\frac{O'(E)}{O(E)}$ can be interpreted as the elasticity of total output, scaled by the inverse of total effort. This representation will be convenient when differentiating $F(e_i, \phi_i)$ with respect to e_i.

Next, define the output shares of the linear and nonlinear components:

$$z_1(E) \stackrel{\text{def}}{=} \frac{aE}{O(E)}, \qquad z_2(E) \stackrel{\text{def}}{=} \frac{bE^{\beta}}{O(E)}$$

so that $z_1(E) + z_2(E) = 1$. These terms measure the fraction of total output generated by the linear part aE and by the nonlinear part bE^{β}, respectively. In terms of these shares, the elasticity $z_3(E)$ can be written as

$$z_3(E) = \frac{O'(E)E}{O(E)} = \frac{aE + b\beta E^{\beta}}{O(E)} = \frac{aE}{O(E)} + \beta\frac{bE^{\beta}}{O(E)} = z_1(E) + \beta z_2(E)$$

Thus $z_3(E)$ is a weighted average of the elasticities of the two components, lying between 1 and β. Substituting back into the derivative of the marginal-benefit term yields

$$\frac{\mathrm{d}}{\mathrm{d}E}\Big(\frac{O'(E)}{O(E)}\Big) = \frac{1}{E^2}\Big[\beta(\beta-1)z_2 - z_3^2\Big]$$

[2]Output elasticity represents how much output changes with labour input. More formally, it is given by the ratio of the percentage change in output by the percentage change in labour or effort.

which, together with $\frac{\partial E}{\partial e_i} = 1$, leads to the production/income component of $\frac{\partial F}{\partial e_i}$ reported in the main text:

$$\frac{\partial F}{\partial e_i} = \underbrace{\left[(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}\right]\frac{1}{E^2}\left[\beta(\beta-1)z_2 - z_3^2\right]}_{\text{(I) Production / income term}} - \underbrace{(1-\phi_i)\,\frac{(1-\alpha_i)}{(1-e_i)^2}}_{\text{(II) Leisure / marginal cost term}}$$

Term (II) is strictly positive and contributes negatively to $\frac{\partial F}{\partial e_i}$, reflecting the *increasing marginal cost of effort* in the individual frame. Term (I) can be positive or negative depending on the technology parameters; its sign is governed by the concave quadratic expression $\beta(\beta-1)z_2 - z_3^2$, which captures how the curvature of the production function affects the marginal return to effort as total effort increases.

B. Sign analysis of the production term

Recall that the production/income component of $\frac{\partial F}{\partial e_i}$ is proportional to

$$\beta(\beta-1)z_2 - z_3^2$$

where $z_2 \in [0,1]$ is the output share of the nonlinear component and $z_3 = 1 + (\beta-1)z_2$ is the elasticity of total output with respect to E. Substituting z_3 yields a concave quadratic in z_2:

$$g(z_2) \stackrel{\text{def}}{=} \beta(\beta-1)z_2 - z_3^2 = \beta(\beta-1)z_2 - \left[1 + (\beta-1)z_2\right]^2$$

The maximum of g over $\mathbb{R}$ is attained at

$$z_2^\star = \frac{\beta-2}{2(\beta-1)}$$

with value

$$g(z_2^\star) = \frac{\beta(\beta-4)}{4}$$

It follows that

$$g(z_2) < 0 \quad \text{for all } z_2 \in [0,1] \text{ whenever } 0 < \beta < 4$$

and $g(z_2) \leq 0$ when $\beta = 4$. Hence, for all $0 < \beta \leq 4$ the production term (I) in $\frac{\partial F}{\partial e_i}$ is non-positive, while the leisure term (II) is strictly negative. This implies

$$\frac{\partial F}{\partial e_i} < 0$$

throughout the relevant parameter space. Since the empirically calibrated range in the simulations is $\beta \sim \mathcal{U}(\frac{3}{4}, \frac{5}{4})$, the denominator of Eq. (5.10) is negative for all feasible parameter values considered in the main text. □

D.2 Derivations for the Firm-Size Comparative Statics

In what follows, the notation

$$R(n) \;\equiv\; \frac{O'(E)}{O(E)} \;=\; \frac{a + b\beta(\bar{e}n)^{\beta-1}}{a\bar{e}n + b(\bar{e}n)^{\beta}} \qquad E = \bar{e}n$$

is used to ease the discussion of the sign of the technological component in Eq. (5.11). Differentiating F with respect to n and allowing ϕ_i to depend on n yields

$$\frac{\partial F}{\partial n} = \underbrace{R'(n)\Big[(1-\phi_i)\,\alpha_i + \phi_i\widetilde{\alpha}\Big]}_{\text{(I) Marginal productivity shift}} + \underbrace{\frac{\partial \phi_i}{\partial n}\,R(n)\,(\widetilde{\alpha} - \alpha_i)}_{\text{(II) Salience-benefit channel}} + \underbrace{\frac{\partial \phi_i}{\partial n}\,\frac{1-\alpha_i}{1-e_i}}_{\text{(III) Salience-cost relief}} \tag{D.1}$$

A. Sign of term (I)

Recall that

$$R(n) = \frac{a + b\beta(\bar{e}n)^{\beta-1}}{a\bar{e}n + b(\bar{e}n)^{\beta}}$$

and that $R'(n)$ determines the sign of term (I) in Eq. (D.1). This subsection shows how the responsiveness of per-capita income to individual effort, as captured by $R(n) = O'(E)/O(E)$, decreases monotonically with firm size whenever $\beta \leq 4$, and how its derivative $R'(n)$ can only become non-negative for $\beta > 4$.

After differentiation and straightforward simplification, one obtains

$$\operatorname{sign}\big(R'(n)\big) = \operatorname{sign}\Big((a\bar{e} + b\bar{e}^{\beta}n^{\beta-1})\,(b\beta\,(\beta-1)\bar{e}^{\beta-1}n^{\beta-1}) - (a\bar{e} + b\beta\,\bar{e}^{\beta}n^{\beta-1})\,(a + b\beta\bar{e}^{\beta-1}n^{\beta-1})\Big)$$

Setting $z_4 = b\bar{e}^{\beta-1}n^{\beta-1}$, this expression simplifies to

$$\operatorname{sign}\big(R'(n)\big) = \operatorname{sign}\Big(\bar{e}\,(a + z_4)(\beta-1)\beta z_4 - \bar{e}\,(a + \beta z_4)^2\Big)$$

Since $\bar{e} > 0$, it can be omitted without loss of generality. Expanding and regrouping terms gives

$$\operatorname{sign}\big(R'(n)\big) = \operatorname{sign}\Big(-a^2 + a\beta b\bar{e}^{\beta-1}n^{\beta-1}(\beta-3) - b^2\bar{e}^{2(\beta-1)}n^{2(\beta-1)}\Big)$$

Introducing the shorthand $d = b\bar{e}^{\beta-1}$ and $z_5 = n^{\beta-1}$, the condition $R'(n) \geq 0$ can be written as the quadratic inequality in z_5:

$$\beta d^2 z_5^2 - a\beta d(\beta-3)z_5 + a^2 \leq 0$$

Since this quadratic is convex, finding a positive minimum $z_5^{\star}$ would imply that the quadratic is always non-negative, and thus $R'(n) \leq 0$ for all z_5. If, on the contrary, the minimum $g(z_5^{\star})$

is negative, then the quadratic crosses the z_5-axis (i.e., has real roots) and changes sign, which in turn implies that $R'(n) \geq 0$ for some n.

Computing the minimum of this quadratic yields

$$z_5^\star = \frac{a\,(\beta - 3)}{2d}$$

with the quadratic evaluated at that point being

$$g(z_5^\star) = a^2 \left[1 - \frac{\beta(\beta-3)^2}{4}\right]$$

This minimum is negative whenever

$$\beta(\beta-3)^2 - 4 > 0$$

Factoring the polynomial gives

$$\beta(\beta-3)^2 - 4 = (\beta-1)^2(\beta-4)$$

Hence, the condition for $R'(n)$ to be non-negative is

$$\boxed{\beta > 4}$$

For the parameter regime considered in this chapter ($\beta \sim \mathcal{U}(\frac{3}{4}, \frac{5}{4})$), this condition is never met. Consequently, $R'(n) < 0$ for all feasible n, confirming that the responsiveness of per-capita income to individual effort decreases monotonically with firm size under empirically relevant parameter values.

B. Signs of terms (II) and (III)

Both components depend on $\partial\phi_i/\partial n$. If identity salience *decreases* with size ($\partial\phi_i/\partial n < 0$; for instance, due to growing anonymity or weaker within-firm contrast), then term (III) is negative:

$$\text{(III)} = \frac{\partial\phi_i}{\partial n}\,\frac{1-\alpha_i}{1-e_i} < 0$$

so the relief in the marginal cost of effort is reduced as the firm grows. The sign of term (II) is given by

$$\text{sign}\left(\tfrac{\partial\phi_i}{\partial n}\right) \cdot \text{sign}(\widetilde{\alpha} - \alpha_i)$$

so that, with $\partial\phi_i/\partial n < 0$, it moves *against* the sign of $(\widetilde{\alpha} - \alpha_i)$: term (II) is negative when the collective norm values income more than the individual ($\widetilde{\alpha} > \alpha_i$), and positive when the

individual is more income-oriented than the group ($\widetilde{\alpha} < \alpha_i$). Intuitively, when salience falls, the marginal benefit channel is pulled back towards the individual preference α_i.

Conversely, if salience *increases* with size ($\partial\phi_i/\partial n > 0$), term (III) is positive and term (II) has the same sign as $(\widetilde{\alpha} - \alpha_i)$. When the social norm is at least as income-oriented as the individual ($\widetilde{\alpha} \geq \alpha_i$), both (II) and (III) then raise the incentive gradient, partially offsetting the negative technological effect in (I). When $\widetilde{\alpha} < \alpha_i$, the salience terms instead reinforce the tendency for effort to decline.

Overall, term (I) is first-order in $R'(n)$ and typically dominates unless $|\partial\phi_i/\partial n|$ is sufficiently large. Therefore, in empirically relevant regimes with $\beta \leq 4$ and moderate salience sensitivity, larger firms flatten individual incentive gradients and elicit lower equilibrium effort.

□

D.3 Detailed computation of $\frac{\partial V_i^{SCT}}{\partial n}$

Section 5.5.4 applied the Envelope Theorem to obtain an expression for the derivative of log-utility with respect to firm size

$$\frac{\partial V_i^{\text{SCT}}}{\partial n} = (1-\phi_i)\frac{\partial V_i^{\text{ind}}}{\partial n} + \phi_i \frac{\partial V_i^{\text{soc}}}{\partial n} + (V_i^{\text{soc}} - V_i^{\text{ind}})\frac{\partial \phi_i}{\partial n}$$

In what follows, the derivatives of the individual and social components in this expression are obtained. The term involving $\partial\phi_i/\partial n$ is left in reduced form, since the main text works directly with its sign.

Individual frame. The individual log-utility is

$$V_i^{\text{ind}} = \alpha_i \ln y_i + (1-\alpha_i)\ln(1-e_i^\star)$$

Since $e_i^\star$ is held constant by the Envelope Theorem,

$$\frac{\partial V_i^{\text{ind}}}{\partial n} = \alpha_i \frac{\partial \ln y_i}{\partial n} = \alpha_i \frac{(\beta-1)\, b\, \bar{e}^\beta\, n^{\beta-2}}{a\bar{e} + b\,\bar{e}^\beta\, n^{\beta-1}} \tag{D.2}$$

Social frame. The social log-utility is

$$V_i^{\text{soc}} = \widetilde{\alpha} \ln \bar{y} + (1 - \widetilde{\alpha}) \ln(1 - \bar{e}_{-i})$$

where $\bar{y} = a\bar{e} + b\,\bar{e}^{\beta}\,n^{\beta-1}$ is per-capita income under equal sharing and

$$\bar{e}_{-i} = \frac{n\bar{e} - e_i^{\star}}{n-1}$$

is the leave-one-out mean effort. Hence

$$\frac{\partial V_i^{\text{soc}}}{\partial n} = \widetilde{\alpha}\,\frac{\partial \ln \bar{y}}{\partial n} - (1 - \widetilde{\alpha})\frac{\partial \bar{e}_{-i}/\partial n}{1 - \bar{e}_{-i}} \tag{D.3}$$

with

$$\frac{\partial \ln \bar{y}}{\partial n} = \frac{(\beta - 1)\,b\,\bar{e}^{\beta}\,n^{\beta-2}}{a\bar{e} + b\,\bar{e}^{\beta}\,n^{\beta-1}}$$

and

$$\frac{\partial \bar{e}_{-i}}{\partial n} = \frac{e_i^{\star} - \bar{e}}{(n-1)^2}$$

Substituting into Eq. (D.3) yields

$$\frac{\partial V_i^{\text{soc}}}{\partial n} = \widetilde{\alpha}\,\frac{(\beta - 1)\,b\,\bar{e}^{\beta}\,n^{\beta-2}}{a\bar{e} + b\,\bar{e}^{\beta}\,n^{\beta-1}} - (1 - \widetilde{\alpha})\frac{e_i^{\star} - \bar{e}}{(1 - \bar{e}_{-i})(n-1)^2} \tag{D.4}$$

Identity salience. If identity salience ϕ_i depends on firm size, its contribution to $\partial V_i^{\text{SCT}}/\partial n$ is given by the term $(V_i^{\text{soc}} - V_i^{\text{ind}})\,\partial\phi_i/\partial n$ in the expression above. The main text therefore works directly with the sign and comparative-static implications of $\partial\phi_i/\partial n$, rather than specifying a particular functional relationship between ϕ_i and n. □

D.4 Justification of Proposition 5.7

Proposition 5.7 states that when the utility perceived under the social frame exceeds that under the individual frame, and identity salience decreases with firm size, workers' welfare is expected to decline with n. This section provides a justification for that result by examining each of the components of $\frac{\mathrm{d}U_i^{\text{SCT}\star}}{\mathrm{d}n}$ in turn

From Eq. (5.15) and Eq. (5.16),

$$\frac{\mathrm{d}U_i^{\mathrm{SCT}\star}}{\mathrm{d}n} = U_i^{\mathrm{SCT}\star}\Bigg[\underbrace{\Big[(1-\phi_i)\,\alpha_i + \phi_i\,\widetilde{\alpha}\Big]\frac{(\beta-1)\,b\,\bar{e}^{\beta}\,n^{\beta-2}}{a\bar{e}+b\,\bar{e}^{\beta}\,n^{\beta-1}}}_{\text{(I) Technological channel}}$$

$$\underbrace{-\,\phi_i(1-\widetilde{\alpha})\frac{e_i^{\star}-\bar{e}}{(1-\bar{e}_{-i})(n-1)^2}}_{\text{(II) Leisure channel}}$$

$$\underbrace{+\quad (V_i^{\mathrm{soc}}-V_i^{\mathrm{ind}})\frac{\partial\phi_i}{\partial n}}_{\text{(III) Identity-salience channel}}\Bigg]$$

Term (I) is technological and captures how changes in n affect utility through the production channel, by increasing or reducing per-capita income. The perceived impact of this technological variation is weighted by the worker's relative valuation of income under the two cognitive frames, $(1-\phi_i)\alpha_i + \phi_i\widetilde{\alpha}$. When $\beta > 1$, the income component in (I) is positive; when $\beta < 1$, it is negative. In the parameter regime used in the simulations, $\beta \sim \mathcal{U}(\frac{3}{4}, \frac{5}{4})$ (see Section 5.6), so the sign of (I) may vary across firms but its magnitude decreases with n

Term (II) represents the leisure channel. It contributes positively whenever the individual's effort is below the firm's mean effort ($e_i^{\star} < \bar{e}$), and negatively when $e_i^{\star} > \bar{e}$. Intuitively, if a worker exerts more effort than average, an increase in n lowers her welfare because her contribution becomes further diluted; conversely, if she exerts less than average, larger firm size increases her welfare. In both cases, the absolute magnitude of (II) decays as n grows, due to the $(n-1)^{-2}$ factor

Term (III) depends on the difference in utility between the social and individual frames and on how identity salience varies with n. When $V_i^{\mathrm{soc}} > V_i^{\mathrm{ind}}$, the worker derives more utility from evaluating her situation through the social frame than through the individual one. If, in addition, $\partial\phi_i/\partial n < 0$ so that salience falls as the firm grows, term (III) becomes negative and pulls welfare downward. Conversely, when $V_i^{\mathrm{soc}} < V_i^{\mathrm{ind}}$, a reduction in salience ($\partial\phi_i/\partial n < 0$) reallocates weight toward the individual frame and therefore increases welfare

Because the technological and leisure terms decay in magnitude with firm size, while the identity-salience term captures the re-weighting between cognitive frames, term (III) becomes the dominant component in the welfare response to firm growth in the regimes of interest. This yields the qualitative patterns summarised in Proposition 5.7

D.5 Simulation Kernel

This appendix reports the full pseudocode of the within-period update used in the computational model. The reader is reminded that $\varphi : \mathcal{A} \to \mathcal{F}$ is the assignment function mapping each agent $i \in \mathcal{A}$ with its firm $f \in \mathcal{F}$: $\varphi(i) = f$

Algorithm 1 Within-Period Update Loop

Activation
Draw $u_i \sim \mathcal{U}(0,1)$ for all i; set $\text{mask}_i \leftarrow [u_i < \pi_m]$ ▷ ≈4% active
Options per agent
for all agents i **do**
  $E_i^{\text{stay}} \leftarrow \textsc{SeenEffort}(i, \varphi(i))$
  $(e_i^{\star,\text{stay}}, U_i^{\text{stay}}) \leftarrow \textsc{BestResponse}(i, \varphi(i), E_i^{\text{stay}})$
  $(e_i^{\star,\text{stup}}, U_i^{\text{stup}}) \leftarrow \textsc{StartupBestResponse}(i)$
  $(f_i^{\text{fr}}, e_i^{\star,\text{fr}}, U_i^{\text{fr}}) \leftarrow \textsc{FriendBestResponse}(i)$ ▷ implemented via message-passing over G
end for
Choice (active only)
for all agents i **do**
  if $\text{mask}_i = 1$ **then**
    $(U_i^{\bullet}, e_i^{\bullet}, f_i^{\bullet}) \leftarrow \max\{(U_i^{\text{stay}}, e_i^{\star,\text{stay}}, \varphi(i)),\ (U_i^{\text{stup}}, e_i^{\star,\text{stup}}, \text{start-up}),\ (U_i^{\text{fr}}, e_i^{\star,\text{fr}}, f_i^{\text{fr}})\}$
    $e_i^{\star} \leftarrow e_i^{\bullet}; \quad f_i^{\star} \leftarrow f_i^{\bullet}$
  else
    $e_i^{\star} \leftarrow e_i^{\star,\text{stay}}; \quad f_i^{\star} \leftarrow \varphi(i)$
  end if
end for
Update (state transition)
$\textsc{AppendStartups}(\{i : f_i^{\star} = \text{start-up}\})$
$\textsc{ReassignAgents}(f_i^{\star})$; $\textsc{RecomputeSizesAges}()$
For each firm f: $E \leftarrow \sum_{i:f_i^{\star}=f} e_i^{\star}$; $O \leftarrow aE + bE^{\beta}$
$\textsc{PruneEmptyFirms}()$; $\textsc{ReindexIDs}()$
if identity salience is modelled endogenously **then**
  $\textsc{RecomputeMCRFromTraits}()$
end if
$\textsc{UpdateWithinFirmAlpha}(\text{median})$
Observation / Logging
$\textsc{LogFirmSnapshots}()$; $\textsc{LogAgentSnapshots}()$; $\textsc{LogMobility}()$

D.6 Size and Utility in a Degenerated Case

In Section D.3, the derivative of an agent's expected utility with respect to firm size n was expressed as:

$$\frac{\partial V_i^{\text{SCT}}}{\partial n} = (1-\phi_i)\frac{\partial V_i^{\text{ind}}}{\partial n} + \phi_i \frac{\partial V_i^{\text{soc}}}{\partial n} + (V_i^{\text{soc}} - V_i^{\text{ind}})\frac{\partial \phi_i}{\partial n}$$

To gain further intuition about the system's dynamics in the collective limit ($\phi \approx 1$), this section focuses on the social term of the expression, $\frac{\partial V_i^{\text{soc}}}{\partial n}$:

$$\frac{\partial V_i^{\text{soc}}}{\partial n} = \widetilde{\alpha}\,\frac{(\beta-1)\,b\,\bar{e}^{\beta}\,n^{\beta-2}}{a\bar{e} + b\,\bar{e}^{\beta}\,n^{\beta-1}} - (1-\widetilde{\alpha})\frac{e_i^{\star} - \bar{e}}{(1-\widetilde{e}_{-i})(n-1)^2} \tag{D.5}$$

The sign and magnitude of $\frac{\partial V_i^{\text{soc}}}{\partial n}$ depend critically on the technology parameter β:

- **If $\beta < 1$** (decreasing returns): The derivative is negative. Collective utility declines as the firm grows.
- **If $\beta = 1$** (constant returns): The derivative converges to zero. Firm size becomes neutral for collective welfare.
- **If $\beta > 1$** (increasing returns): The first term in Equation (D.5) dominates for large n. The numerator scales with $n^{\beta-2}$, while the denominator grows with $n^{\beta-1}$, yielding a positive term that decays more slowly than the cost, which scales as n^{-2}. Consequently, collective utility increases as the firm grows.

Under a fully collective regime ($\phi = 1$), agents evaluate firms solely in terms of collective welfare. Thus:

- Firms with $\beta < 1$ cease to retain members and stabilise at finite size.
- Firms with $\beta > 1$ generate persistently positive marginal social utility, and those who enter them are more easily retained.
- Over time, the population converges towards firms with the highest β—or splits evenly among those tied at the maximum value.

This analytical limit provides a clear boundary condition for the model: When $\phi = 1$, the system behaves as a *fully collective economy*, whose long-run structure is governed by technological increasing returns. This degenerate social case illustrates an endogenous emergence of monopolies when social identity fully dominates individual incentives.

D.7 *Scenario A (i.i.d. Traits): Endogenous Identity without Network Homophily*

In this robustness experiment, agents' attributes are drawn independently and identically distributed (i.i.d.) from the population distribution, removing the network-based homophily that, in the baseline scenario, amplifies inter-intra group contrast.

Without homophily, both within-group similarity and between-group distinctiveness decline. Empirically, the average of ϕ slightly decreases relative to the homophilic baseline, as shown in the simulation diagnostics of Fig. D.1.

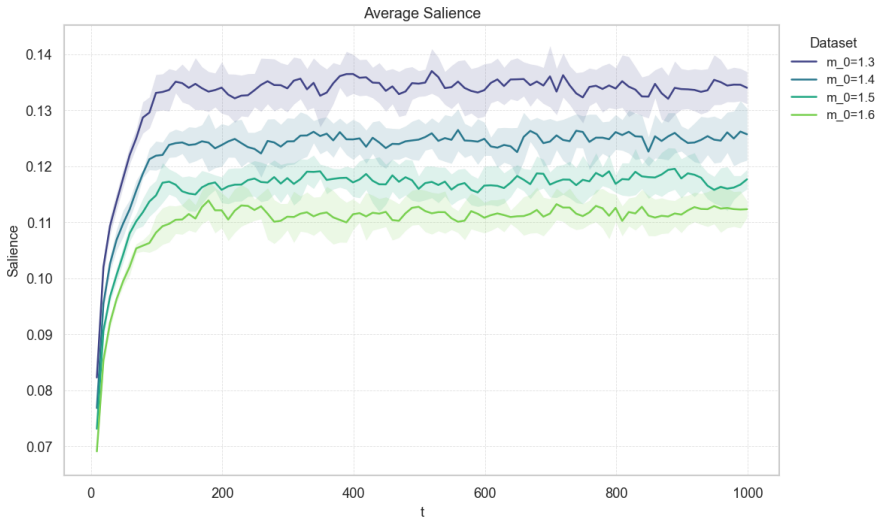


Figure D.1: **Average salience over time for different normal trait distributions** in Scenario A (i.i.d. traits). Removing homophily reduces the overall level and variability of ϕ compared to the baseline.

The direction of all effects remains identical to the baseline. At the macro level, the economy retains a heavy-tailed firm-size distribution, confirming that the absence of homophily affects identity salience but not the structural scaling regularity. Thus, endogenous identity formation—even without network reinforcement—reproduces qualitatively the dynamics of the baseline case.

D.8 Scenario B (Exogenous Identity): Salience Control via Sampled MCR

In this complementary scenario, identity salience is controlled exogenously by sampling the meta-contrast ratio MCR from a log-normal distribution. The mean of the distribution is varied between 1.0 and 2.0 while keeping $M_0 = 1.5$, thereby generating distinct regimes of contrast strength. This procedure allows direct inspection of how exogenously imposed salience shapes behaviour and macro outcomes.

D.8.1 Salience response

As shown in Fig. D.2, increasing the log-normal mean induces a steep, monotonic rise in mean salience. When $\mu = 1.0$, average ϕ stabilises around 0.1–0.2; for $\mu = 1.6$ it climbs beyond 0.5; and for $\mu = 2.0$ it approaches saturation near 1. This confirms that even modest changes in the mean of the underlying MCR distribution produce disproportionately large shifts in the salience field due to the logistic transformation in Eq. (5.3).

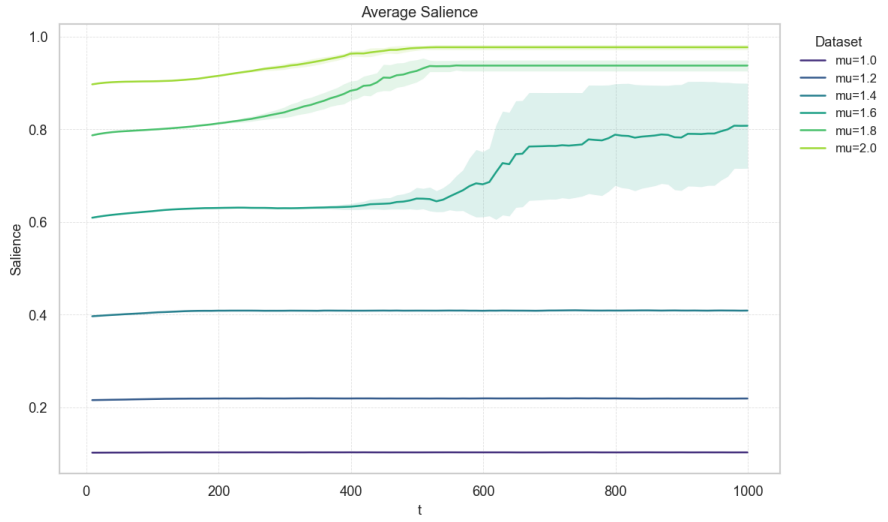


Figure D.2: **Average salience over time for different log-normal means μ in the sampled-MCR scenario**. Higher mean values of MCR sharply increase salience, with $\mu = 2.0$ leading to near-saturation.

D.8.2 *Behavioural and meso-level effects*

Non-linearities are also present as far as free-riding and firm turnover concerns. An initial increase in μ for moderately-low values ($\mu \in [1.0 - 1.4]$) *reduces* free-riding. However, as shown in Fig. D.3, the fraction of free-riders grows sharply—reaching more than 80%—once it reaches sufficiently high values ($\mu = 1.6$), leading to almost full shirking for $\mu = 2.0$. Mobility (Fig. D.4) declines as salience increases, indicating that over-identification locks individuals into firms even when coordination collapses. This pattern mirrors the theoretical findings in Section D.2: when $\phi \approx 1$, the system behaves as a *fully collective economy* converging towards one or very few mega-firms (Fig. D.5). Under those circumstances, once an agent joins a company, they rarely leave it. At the same time, and according to the results in Section 5.5.2, an increasing headcount generally inhibits contributions, which explains the rise in free-riding.

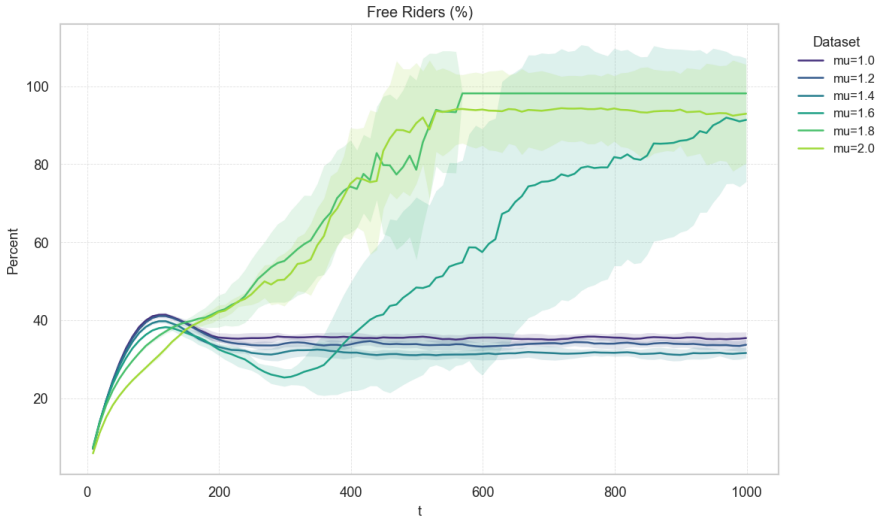


Figure D.3: Share of free-riders over time for different μ values in the sampled-MCR scenario. High-salience regimes ($\mu \geq 1.6$) produce extensive shirking as conformity suppresses individual effort.

D.8.3 *Macro structure*

Despite these behavioural shifts, the aggregate economy preserves its heavy-tailed firm-size distribution whenever social identity salience is not excessively high and size distribution is not biased by the presence of mega-firms (Fig. D.6). In particular, it remains robust to whether identity is endogenous or exogenous.



Figure D.4: **Job changes over time**. As salience rises, worker mobility declines, reflecting over-identification and entrenchment within firms.

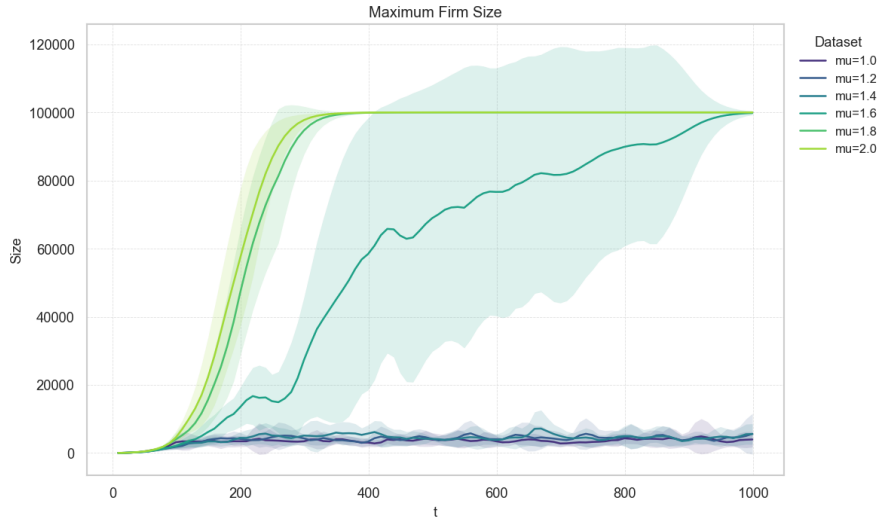


Figure D.5: Maximum firm size over time for different values of the log-normal mean, μ. High-salience regimes generate the emergence of macro-firms dominating the economy.

The sampled-MCR configuration reveals a clear map of behavioural regimes that concur with the theoretical results derived from the comparative statics in Section 5.5. Moderate contrast enhances cooperation and productivity, whereas excessive contrast leads to conformity, stagnation, and large but *lazy* firms. Overall, salience ϕ acts as a mediator linking individual behaviour and macro structure: low ϕ weakens coordination, moderate ϕ sustains it, and

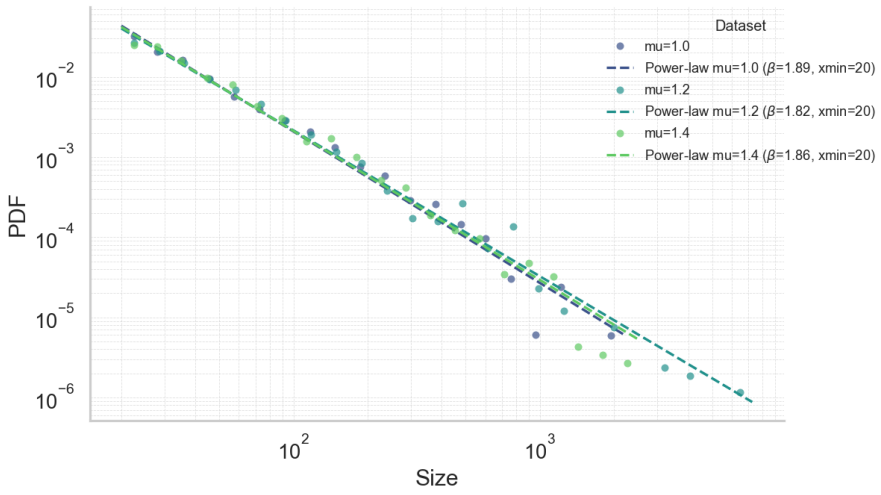


Figure D.6: **Firm-size distribution under exogenous** MCR **control**. A power-law fit yields $\beta \simeq 1.85$ for $N_0 = 20$ in the non-degenerate cases, confirming that the macro-scaling structure is preserved.

excessively high ϕ can invert incentives.

Bibliography

Daron Acemoglu and Martin Kaae Jensen. Aggregate comparative statics. *Games and Economic Behavior*, 81:27–49, 2013.

John P Agapiou, Alexander Sasha Vezhnevets, Edgar A Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, DJ Strouse, Michael B Johanson, Sukhdeep Singh, Julia Haas, Igor Mordatch, Dean Mobbs, and Joel Z Leibo. Melting pot 2.0. *arXiv preprint arXiv:2211.13746*, 2022.

George A. Akerlof and Rachel E. Kranton. Economics and Identity*. *The Quarterly Journal of Economics*, 115(3):715–753, 08 2000a.

George A Akerlof and Rachel E Kranton. Economics and identity. *The quarterly journal of economics*, 115(3):715–753, 2000b.

George A. Akerlof and Rachel E. Kranton. Identity and the economics of organizations. *Journal of Economic Perspectives*, 19(1):932, March 2005. doi: 10.1257/0895330053147930. URL https://www.aeaweb.org/articles?id=10.1257/0895330053147930.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

Stefano V. Albrecht, Filippos Christianos, and Lukas Schäfer. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*. MIT Press, 2024. URL https://www.marl-book.com.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.

Ewa Andrejczuk, Juan A Rodriguez-Aguilar, and Carles Sierra. A concise review on multiagent teams: contributions and research opportunities. In *European Conference on Multi-Agent Systems*, pages 31–39. Springer, 2016.

Qazi Mohammad Areeb, Mohammad Nadeem, Shahab Saquib Sohail, Raza Imam, Faiyaz Doctor, Yassine Himeur, Amir Hussain, and Abbes Amira. Filter bubbles in recommender systems: Fact or fallacya systematic review. *WIREs Data Mining and Knowledge Discovery*, 13(6):e1512, 2023. doi: https://doi.org/10.1002/widm.1512. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1512.

Holly Arrow, Joseph McGrath, and Jennifer Berdahl. *Small Groups As Complex Systems: Formation, Coordination, Development, and Adaptation*. 01 2000. ISBN 9780803972308. doi: 10.4135/9781452204666.

Kenneth J Arrow. A difficulty in the concept of social welfare. *Journal of political economy*, 58 (4):328–346, 1950.

Kenneth J Arrow. *Social choice and individual values*, volume 12. John Wiley & Sons, 1951.

Blake E. Ashforth and Fred Mael. Social identity theory and the organization. *The Academy of Management Review*, 14(1):20–39, 1989.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Robert Axelrod and William D Hamilton. The evolution of cooperation. *science*, 211(4489): 1390–1396, 1981.

R. Axtell, Center on Social, Economic Dynamics, Brookings Institution, and Johns Hopkins University. *The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions*. Working paper (Center on Social and Economic Dynamics). Center on Social and Economic Dynamics, 1999. URL https://books.google.es/books?id=p_h50AEACAAJ.

Robert Axtell. Chapter 3 - endogenous firm dynamics and labor flows via heterogeneous agents. In Cars Hommes and Blake LeBaron, editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, pages 157–213. Elsevier, 2018. doi: https://doi.org/10.1016/bs.hescom.2018.05.001. URL https://www.sciencedirect.com/science/article/pii/S1574002118300133.

Robert L. Axtell. Non-cooperative dynamics of multi-agent teams. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 3*, AAMAS '02, page 10821089, New York, NY, USA, 2002. Association for Computing Machinery. ISBN 1581134800. doi: 10.1145/545056.545076. URL https://doi.org/10.1145/545056.545076.

Robert L Axtell. 120 million agents self-organize into 6 million firms: A model of the us private sector. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 806–816, 2016a.

Robert L Axtell. 120 million agents self-organize into 6 million firms: A model of the us private sector. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 806–816, 2016b.

Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 63(1):197–287, 2025.

Michael Bacharach. Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics*, 53(2):117–147, 1999. ISSN 1090-9443. doi: https://doi.org/10.1006/reec.1999.0188. URL https://www.sciencedirect.com/science/article/pii/S1090944399901886.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

Bowen Baker. Emergent reciprocity and team formation from randomized uncertain social preferences. *Advances in neural information processing systems*, 33:15786–15799, 2020.

Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015. doi: 10.1126/science.aaa1160. URL https://www.science.org/doi/abs/10.1126/science.aaa1160.

Daniel Balliet and Paul AM Van Lange. Trust, conflict, and cooperation: a meta-analysis. *Psychological bulletin*, 139(5):1090, 2013.

Albert Bandura and Richard H Walters. *Social learning theory*, volume 1. Englewood cliffs Prentice Hall, 1977.

Adriana Alventosa Baños and Penélope Hernández Rojas. Monitoring teamwork: how do punishment motivations change with group size. *Economía industrial*, (413):51–59, 2019.

Martin Beckenkamp. A game theoretic taxonomy of social dilemmas. Technical report, Preprints aus der Max-Planck-Projektgruppe Recht der Gemeinschaftsgüter, 2002.

Theodore C Bergstrom, Lawrence E Blume, and Hal R Varian. Uniqueness of nash equilibrium in private provision of public goods: an improved proof. 1992.

Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53:659–697, 2015.

Lutz Bornmann. Skewed distributions of scientists productivity: A research program for the empirical analysis. *Scientometrics*, 129(4):2455–2468, 2024.

Samuel Bowles and Herbert Gintis. *The Origins of Human Cooperation*. 10 2003.

Rodica Branzei, Dinko Dimitrov, and Stef Tijs. *Models in cooperative game theory*, volume 556. Springer Science & Business Media, 2008.

Steven J Breckler and Anthony G Greenwald. Motivational facets of the self. 1986.

Wolfgang Buchholz and Dirk Rübbelke. *The Theory of Externalities and Public Goods*. Springer, 2017.

Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008.

Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997. ISSN 0022-0531. doi: https://doi.org/10.1006/jeth.1997.2319. URL https://www.sciencedirect.com/science/article/pii/S002205319792319X.

Colin F Camerer. Behavioural game theory. In *The New Palgrave Dictionary of Economics*, pages 867–874. Springer, 2018.

Michael A Campion, Gina J Medsker, and A Catherine Higgs. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel psychology*, 46(4):823–847, 1993.

Stuart K Card. *The psychology of human-computer interaction*. Crc Press, 2018.

Jeffrey Carpenter, Samuel Bowles, Herbert Gintis, and Sung-Ha Hwang. Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behavior & Organization*, 71 (2):221–232, 2009.

C.H. Casal, F.M. Alonso, Á.G. Jiménez, and J.F.M. Domínguez. *PSICOLOGÍA DE LOS GRUPOS*. Colección Grado. UNED, 2011. ISBN 9788436261677. URL https://books.google.es/books?id=2R1vtwAACAAJ.

Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. *Computational aspects of cooperative game theory*. Springer Nature, 2022.

Archie C Chapman, David S Leslie, Alex Rogers, and Nicholas R Jennings. Convergent learning algorithms for unknown reward games. *SIAM Journal on Control and Optimization*, 51(4): 3154–3180, 2013.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009. doi: 10.1137/070710111. URL https://doi.org/10.1137/070710111.

Philip R Cohen and Hector J Levesque. Confirmations and joint action. In *IJCAI*, pages 951–959, 1991a.

Philip R Cohen and Hector J Levesque. Teamwork. *Nous*, 25(4):487–512, 1991b.

Vincent Conitzer and Caspar Oesterheld. Foundations of cooperative ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15359–15367, 2023.

Luis C Corchón. Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences*, 28(3):151–165, 1994.

Luis C Corchón. Aggregative games. *SERIEs*, 12(1):49–71, 2021.

Richard Cornes. Dyke maintenance and other stories: some neglected types of public goods. *The Quarterly Journal of Economics*, 108(1):259–271, 1993.

Richard Cornes. *Voluntary Contribution Model of Public Goods*, pages 1–9. Palgrave Macmillan UK, London, 2016.

Richard Cornes and Roger Hartley. Aggregative public good games. *Journal of Public Economic Theory*, 9(2):201–219, 2007a.

Richard Cornes and Roger Hartley. Weak links, good shots and other public good games: Building on bbv. *Journal of Public Economics*, 91(9):1684–1707, 2007b.

Richard Cornes and Todd Sandler. *The theory of externalities, public goods, and club goods*. Cambridge University Press, 1996.

John G. Cross. A stochastic learning model of economic behavior. *The Quarterly Journal of Economics*, 87(2):239–266, 1973. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/1882186.

Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R McKee, Joel Z Leibo, Kate Larson, and Thore Graepel. Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*, 2020.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Cooperative ai: machines must learn to find common ground. *Nature*, 593(7857):33–36, 2021.

Robyn M Dawes. Social dilemmas. *Annual review of psychology*, 1980.

David De Cremer. Trust and fear of exploitation in a public goods dilemma. *Current Psychology*, 18:153–163, 1999.

David De Cremer and Mark Van Vugt. Social identification effects in social dilemmas: A transformation of motives. *European journal of social psychology*, 29(7):871–893, 1999.

Lauro Langosco Di Langosco, Jack Koch, Lee D Sharkey, Jacob Pfau, and David Krueger. Goal misgeneralization in deep reinforcement learning. In *International Conference on Machine Learning*, pages 12004–12019. PMLR, 2022.

Alan Dix. Human–computer interaction, foundations and new paradigms. *Journal of Visual Languages & Computing*, 42:122–134, 2017.

Edgar A Duéñez-Guzmán, Ramona Comanescu, Yiran Mao, Kevin R McKee, Ben Coppin, Suzanne Sadedin, Silvia Chiappa, Alexander S Vezhnevets, Michiel A Bakker, Yoram Bachrach, et al. Perceptual interventions ameliorate statistical discrimination in learning agents. *Proceedings of the National Academy of Sciences*, 122(25):e2319933121, 2025.

Barbara Dunin-Keplicz and Rineke Verbrugge. Teamwork in multi-agent systems: A formal approach. *Teamwork in Multi-Agent Systems: A Formal Approach*, 05 2010. doi: 10.1002/9780470665237.

JR Eiser. Cooperation and competition between individuals. *Introducing Social Psychology. New York: Penguin*, pages 151–175, 1978.

Edith Elkind and Jörg Rothe. Cooperative game theory. *Economics and computation: an introduction to algorithmic game theory, computational social choice, and fair division*, pages 135–193, 2016.

J. Espinosa, Javier Lerch, and Robert Kraut. Explicit vs. implicit coordination mechanisms and task dependencies: One size does not fit all. *Team Cognition: Understanding the Factors That Drive Process and Performance*, 01 2004. doi: 10.1037/10690-006.

Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

Deborah L Feltz, Norbert L Kerr, and Brandon C Irwin. Buddy up: the köhler effect applied to health games. *Journal of Sport and Exercise Psychology*, 33(4):506–526, 2011.

Stephen M. Fiore and Eduardo Salas. Why we need team cognition. In Eduardo Salas and Stephen M. Fiore, editors, *Team cognition: Understanding the factors that drive process and performance*, pages 235–248. American Psychological Association, 2004. doi: 10.1037/10690-011. URL https://doi.org/10.1037/10690-011.

Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 21452153, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

Srihari Govindan, Philip J Reny, Arthur J Robson, et al. A short proof of harsanyi's purification theorem. *Games and Economic Behavior*, 45(2):369–374, 2003.

Barbara J Grosz. Collaborative systems (aaai-94 presidential address). *AI magazine*, 17(2):67–67, 1996.

Barbara J Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

Barbara J. Grosz and Candace L. Sidner. Plans for discourse. In *Intentions in Communication*. The MIT Press, 06 1990. ISBN 9780262270540. doi: 10.7551/mitpress/3839.003.0022. URL https://doi.org/10.7551/mitpress/3839.003.0022.

Zihao Guo, Shuqing Shi, Richard Willis, Tristan Tomilin, Joel Z Leibo, and Yali Du. Socialjax: An evaluation suite for multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:2503.14576*, 2025.

Richard A Guzzo and Marcus W Dickson. Teams in organizations: Recent research on performance and effectiveness. *Annual review of psychology*, 47(1):307–338, 1996.

Gymnasium. Vector normalize wrapper, 2023. URL https://gymnasium.farama.org/v0.29.0/_modules/gymnasium/wrappers/normalize/.

Roger Hartley and Richard Cornes. Joint Production Games And Share Functions. Keele Department of Economics Discussion Papers (1995-2001) 2000/07, Department of Economics, Keele University, October 2000. URL https://ideas.repec.org/p/kee/keeldp/2000-07.html.

Jack Hirshleifer. From weakest-link to best-shot: The voluntary provision of public goods. *Public choice*, 41(3):371–386, 1983.

Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

James Hollan, Edwin Hutchins, and David Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(2):174–196, 2000.

Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems*, 31, 2018.

David E Hyatt and Thomas M Ruddy. An examination of the relationship between work group characteristics and performance: Once more into the breech. *Personnel psychology*, 50 (3):553–585, 1997.

Daniel R Ilgen, John R Hollenbeck, Michael Johnson, and Dustin Jundt. Teams in organizations: from input-process-output models to imoi models. *Annual review of psychology*, 56:517543, 2005. ISSN 0066-4308. doi: 10.1146/annurev.psych.56.091103.070250.

Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro A Ortega, DJ Strouse, Joel Z Leibo, and Nando de Freitas. Intrinsic social motivation via causal influence in multi-agent rl. 2018.

Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, DJ Strouse, Joel Z Leibo, and Nando De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International conference on machine learning*, pages 3040–3049. PMLR, 2019.

Julian Jara-Ettinger. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29:105–110, 2019.

Nicholas R Jennings. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial intelligence*, 75(2):195–240, 1995.

Nick R Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *The knowledge engineering review*, 8(3):223–250, 1993.

Martin Kaae Jensen et al. Aggregative games. *Handbook of game theory and industrial organization*, 1:66–92, 2018.

Daniel Kahneman. Econ ometrica i ci. *Econometrica*, 47(2):263–291, 1979.

Daniel Kahneman. Maps of bounded rationality: A perspective on intuitive judgement and choice. 2002.

Tatsuya Kameda, Mark F Stasson, James H Davis, Craig D Parks, and Suzi K Zimmerman. Social dilemmas, subgroups, and motivation loss in task-oriented groups: In search of an" optimal" team size in division of work. *Social Psychology Quarterly*, pages 47–56, 1992.

Steven Karau and Kipling Williams. *Understanding individual motivation in groups: The collective effort model*, pages 113–141. 01 2001.

Steven J Karau and Kipling D Williams. Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, 65(4):681, 1993.

Harold H. Kelley and John W. Thibaut. Interpersonal relations: A theory of interdependence. 1978. URL https://api.semanticscholar.org/CorpusID:145761013.

H.H. Kelley. *Personal Relationships: Their Structures and Processes*. Distinguished Lecture Series. Taylor & Francis Group, 1979. ISBN 9780898594706. URL https://books.google.es/books?id=BRMevgAACAAJ.

Norbert L Kerr. Motivation losses in small groups: A social dilemma analysis. *Journal of personality and social psychology*, 45(4):819, 1983.

Norbert L Kerr and R Scott Tindale. Group performance and decision making. *Annu. Rev. Psychol.*, 55:623–655, 2004.

Sebastian Krakowski. Human-ai agency in the age of generative ai. *Information and Organization*, 35(1):100560, 2025.

János Kramár, Tom Eccles, Ian Gemp, Andrea Tacchetti, Kevin R McKee, Mateusz Malinowski, Thore Graepel, and Yoram Bachrach. Negotiation and honesty in artificial intelligence methods for the board game of diplomacy. *Nature Communications*, 13(1):7214, 2022.

Roderick M Kramer, Benjamin A Hanna, Steven Su, and Jane Wei. Collective identity, collective trust, and social capital: Linking group identification and group cooperation. In *Groups at work*, pages 173–196. Psychology Press, 2014.

Nikolai S Kukushkin. A fixed-point theorem for decreasing mappings. *Economics Letters*, 46 (1):23–26, 1994.

James R Larson Jr. *In search of synergy in small group performance*. Psychology Press, 2013.

Joel Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas, 2017. arXiv:1702.03037.

David S. Leslie and E. J. Collins. Individual q-learning in normal form games. *SIAM Journal on Control and Optimization*, 44(2):495–514, 2005. doi: 10.1137/S0363012903437976.

Hector J Levesque, Philip R Cohen, and José HT Nunes. *On acting together*, volume 485. SRI International Menlo Park, CA 94025-3493, 1990.

Kurt Lewin. Resolving social conflicts; selected papers on group dynamics. 1948.

Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994.

Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

Alejandra López de Aberasturi Gómez, Carles Sierra, and Jordi Sabater-Mir. Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach. *Artificial Intelligence*, 341:104307, 2025. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2025.104307. URL https://www.sciencedirect.com/science/article/pii/S0004370225000268.

Udari Madhushani, Kevin R McKee, John P Agapiou, Joel Z Leibo, Richard Everett, Thomas Anthony, Edward Hughes, Karl Tuyls, and Edgar A Duéñez-Guzmán. Heterogeneous social value orientation leads to meaningful diversity in sequential social dilemmas. *arXiv preprint arXiv:2305.00768*, 2023.

Setareh Maghsudi and Ekram Hossain. Multi-armed bandits with application to 5g small cells. *IEEE Wireless Communications*, 23(3):64–73, 2016.

Michelle A. Marks, John E. Mathieu, and Stephen J. Zaccaro. A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, 26(3):356–376, 2001. URL http://www.jstor.org/stable/259182.

Michael Maschler, Shmuel Zamir, and Eilon Solan. *Game theory*. Cambridge University Press, 2020.

J Maynard Smith and G. R Price. The logic of animal conflict. 1973.

John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

Joseph Edward McGrath. Social psychology: A brief introduction. 1964.

Kevin R McKee, Ian Gemp, Brian McWilliams, Edgar A Duéñez-Guzmán, Edward Hughes, and Joel Z Leibo. Social diversity and social preferences in mixed-motive reinforcement learning. *arXiv preprint arXiv:2002.02325*, 2020.

Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for extensive form games. *Experimental economics*, 1(1):9–41, 1998.

Ryan Murphy, Kurt Ackermann, and Michel Handgraaf. Measuring social value orientation. *Judgment and Decision Making*, 6:771–781, 01 2023. doi: 10.1017/S1930297500004204.

David P Myatt and Chris Wallace. Evolution, teamwork and collective action: Production targets in the private provision of public goods. *The Economic Journal*, 119(534):61–90, 2009.

John F Nash. Non-cooperative games. In *The Foundations of Price Theory Vol 4*, pages 329–340. Routledge, 2024.

Kamal Ndousse, Doug Eck, Sergey Levine, and Natasha Jaques. Emergent social learning via multi-agent reinforcement learning. *Internation Conference on Machine Learning (ICML)*, 2021.

Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46 (5):323–351, 2005.

Bernard A Nijstad. *Group performance*. Psychology Press, 2009.

Martin A Nowak. Five rules for the evolution of cooperation. *science*, 314(5805):1560–1563, 2006.

Mancur Olson Jr. *The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix*, volume 124. harvard university press, 1971.

Elinor Ostrom. Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation. Technical report, 2003a. URL http://about.jstor.org/terms.

Elinor Ostrom. Toward a behavioral theory linking trust, reciprocity, and reputation. *Trust and reciprocity: Interdisciplinary lessons from experimental research*, 6:19–79, 2003b.

Julien Perolat, Joel Z Leibo, Vinicius Zambaldi, Charles Beattie, Karl Tuyls, and Thore Graepel. A multi-agent reinforcement learning model of common-pool resource appropriation. *Advances in neural information processing systems*, 30, 2017.

Canice Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, 37(1): 7–63, 1999. ISSN 00220515. URL http://www.jstor.org/stable/2564725.

David V Pynadath, Milind Tambe, Nicolas Chauvat, and Lawrence Cavedon. Toward team-oriented programming. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 233–247. Springer, 1999.

Anatol Rapoport and Albert M Chammah. *Prisoner's dilemma: A study in conflict and cooperation*, volume 165. University of Michigan press, 1965.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning, 2018. URL https://arxiv.org/abs/1803.11485.

Paul B Reverdy, Vaibhav Srivastava, and Naomi Ehrich Leonard. Modeling human decision making in generalized gaussian multiarmed bandits. *Proceedings of the IEEE*, 102(4):544–571, 2014.

Ramón Rico, Miriam Manzanares, Francisco Gil, and C. Gibson. Team implicit coordination processes: A team knowledge-based approach. *The Academy of Management Review (AMR)*, 33:163184, 01 2008. doi: 10.5465/AMR.2008.27751276.

Ramón Rico, Carlos-Maria Alcover, and Carmen Tabernero. Work team effectiveness, a review of research from the last decade (1999-2009). *Psychology in Spain*, 15:57–79, 06 2011.

E Frank Ritter, D Gordon Baxter, and F Elizabeth Churchill. *Foundations for designing user-centered systems: What system designers need to know about people*. Springer, 2014.

Herbert Robbins. Some aspects of the sequential design of experiments. 1952.

Stuart Russell. *Human compatible: AI and the problem of control*. Penguin Uk, 2019.

Francisco C Santos, Jorge M Pacheco, and Tom Lenaerts. Cooperation prevails when individuals adjust their social ties. *PLoS computational biology*, 2(10):e140, 2006a.

Francisco C Santos, Jorge M Pacheco, and Tom Lenaerts. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences*, 103(9):3490–3494, 2006b.

Thomas C. Schelling. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. *The Journal of Conflict Resolution*, 17(3):381–428, 1973. ISSN 00220027, 15528766. URL http://www.jstor.org/stable/173406.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*, 2015.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Reinhard Selten. *Preispolitik der Mehrproduktenunternehmung in der statischen Theorie*. Springer-Verlag, 2013.

Lloyd S Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953a.

Lloyd S Shapley. A value for n-person games. *Contribution to the Theory of Games*, 2, 1953b.

Yoav Shoham. Agent-oriented programming. *Artificial intelligence*, 60(1):51–92, 1993.

Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

Yoav Shoham and Moshe Tennenholtz. On social laws for artificial agent societies: off-line design. *Artificial intelligence*, 73(1-2):231–252, 1995.

Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial intelligence*, 171(7):365–377, 2007.

Carles Sierra, Jordi Sabater, Jaume Agustí, and Pere Garcia. Integrating evolutionary computing and the sadde methodology. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, page 11161117, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136838. doi: 10.1145/860575.860823. URL https://doi.org/10.1145/860575.860823.

Herbert A Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

Christine Smith. Task performance in groups. In *Oxford Research Encyclopedia of Psychology*. 2022.

Ivan Dale Steiner. *Group process and productivity*. Academic press New York, 1972.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

Henri Tajfel and John C Turner. The social identity theory of intergroup behavior. In *Political psychology*, pages 276–293. Psychology Press, 2004.

Milind Tambe. Towards flexible teamwork. *Journal of artificial intelligence research*, 7:83–124, 1997.

Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1):145–156, 1978. ISSN 0025-5564. doi: https://doi.org/10.1016/0025-5564(78)90077-9. URL https://www.sciencedirect.com/science/article/pii/0025556478900779.

John W. Thibaut and Harold H. Kelley. *Interpersonal Relations: A Theory of Interdependence*. Wiley, New York, 1959.

Michel Tokic and Günther Palm. Value-difference based exploration: adaptive control between epsilon-greedy and softmax. In *Annual conference on artificial intelligence*, pages 335–346. Springer, 2011.

Amy J. C. Trappey, Ching-Hung Lee, Zuoxu Wang, and Fan Li. Exploring advanced human-ai hybrid intelligence solutions now and beyond: a synthesis of the special issue portfolio. *Journal of Engineering Design*, 0(0):1–6, 2026. doi: 10.1080/09544828.2025.2610794. URL https://doi.org/10.1080/09544828.2025.2610794.

J.C. Turner. *Social Categorization and Self-Concept: A Social Cognitive Theory of Group Behavior*, pages 77–121. JAI Press, Greenwich, Connecticut, 1985.

J.C. Turner. *Rediscovering the Social Group: A Self-categorization Theory*. B. Blackwell, 1987. ISBN 9780631148067. URL https://books.google.co.uk/books?id=C-yZQgAACAAJ.

John C. Turner. *Social identification and psychological group formation*, page 518538. European Studies in Social Psychology. Cambridge University Press, 1984.

John C Turner and S A Haslam. *Social identity, organizations and leadership*, volume 0, pages 25–65. Lawrence Erlbaum Associates, first edition, 2001. ISBN 0805820795.

Karl Tuyls and Ann Nowé. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review*, 20(1):63–90, 2005.

Karl Tuyls and Simon Parsons. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7):406–416, 2007. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2007.01.004. URL https://www.sciencedirect.com/science/article/pii/S0004370207000082. Foundations of Multi-Agent Learning.

Karl Tuyls, Dries Heytens, Ann Nowe, and Bernard Manderick. Extended replicator dynamics as a key to reinforcement learning in multi-agent systems. In *European Conference on Machine Learning*, pages 421–431. Springer, 2003.

Karl Tuyls, Ann Nowe, Tom Lenaerts, and Bernard Manderick. An evolutionary game theoretic perspective on learning in multi-agent systems. *Synthese*, 139(2):297–330, 2004.

John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior, 2nd rev. 1947.

Ruth Wageman. The meaning of interdependence. In *Groups at work*, pages 197–217. Psychology Press, 2014.

Yufeng Wang, Yuan Xu, Anastasia Nikolova, Yuxuan Wang, Jianyu Wang, Chongyang Wang, and Xin Tong. How do we research human-robot interaction in the age of large language models? a systematic review. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400722783. doi: 10.1145/3772318.3790920. URL https://doi.org/10.1145/3772318.3790920.

Gerhard Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press, 1999.

K. D. Williams, Karau and S. J. Karau. Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology*, 61(4): 570–581, 1991.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.

Richard Willis, Yali Du, Joel Z Leibo, and Michael Luck. Quantifying the self-interest level of markov social dilemmas. *arXiv preprint arXiv:2501.16138*, 2025.

David H Wolpert and Kagan Tumer. An introduction to collective intelligence. *arXiv preprint cs/9908014*, 1999.

David H Wolpert and Kagan Tumer. Optimal payoff functions for members of collectives. *Advances in Complex Systems*, 4(02n03):265–279, 2001.

Michael Wooldridge. *An introduction to multia-gent systems*. John Wiley & Sons, 2009.

Charley M Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D Nelson, and Björn Meder. Generalization guides human exploration in vast decision spaces. *Nature human behaviour*, 2 (12):915–924, 2018.

Jangho Yang, Torsten Heinrich, Julian Winkler, François Lafond, Pantelis Koutroumpis, and J. Doyne Farmer. Measuring productivity dispersion: a parametric approach using the lévy alpha-stable distribution. *Industrial and Corporate Change*, 34(1):79–117, 05 2024. ISSN 0960-6491. doi: 10.1093/icc/dtae021. URL https://doi.org/10.1093/icc/dtae021.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624, 2022.

Kaiqing Zhang, Zhuoran Yang, and Tamer Basar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, 18: 321, 2021.

Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C Parkes, and Richard Socher. The ai economist: Improving equality and productivity with ai-driven tax policies. *arXiv preprint arXiv:2004.13332*, 2020.

Illustration credits

The artwork in this thesis was created by **elartgentino**.

©2025. Reproduced with permission of the artist.