Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

Marco LiCalzi and Roland Mühlenbernd

**Categorization and cooperation
across games**

**Working Paper n. 14/2018
November 2018**

ISSN: 2239-2734

This Working Paper is published under the auspices of the Department of Management at Università Ca' Foscari Venezia. Opinions expressed herein are those of the authors and not those of the Department or the University. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional nature.

Categorization and cooperation across games*

MARCO LICALZI

[licalzi@unive.it]

ROLAND MÜHLENBERND

[roland.muhlenbernd@unive.it]

Department of Management, Università Ca' Foscari Venezia

(November 2018)

Abstract We study a model where agents face a continuum of two-player games and categorize them into a finite number of situations to make sense of their complex environment. Each agent can cooperate or defect, conditional on the perceived category. Agents may not share the same categorization. The games are fully ordered by one parameter, interpreted as the temptation to break joint cooperation by defecting. We prove that in equilibrium agents must share the same categorization. Most equilibria achieve less cooperation than it would be possible if agents could fully discriminate games. All the equilibria are evolutionarily stable, but the only stochastically stable profile leads to defection everywhere, destroying all opportunities for cooperation. We then study agents' social learning when they imitate successful players over similar games, but lack any information about the categorizations of other players. We show how imitation leads to a shared categorization that achieves higher cooperation than under full discrimination.

Keywords: categorization, cooperation, evolutionary stability, learning by imitation, prisoners' dilemma, stag hunt.

JEL Classification Numbers: C72, D91, C73, D83.

Correspondence to:

Roland Mühlenbernd	Department of Management Università Ca' Foscari Venezia San Giobbe, Cannaregio 873 30121 Venezia, Italy
Fax:	[+39] 041-234-7444
E-mail:	roland.muhlenbernd@unive.it

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 732942.

1 Introduction

The standard approach to rational decisions presumes that the agent knows the states of the world, her available actions, and the outcomes over which he has a preference ordering. This knowledge is very difficult to attain, because an agent needs to make sense of the environment in order to distill states, actions, and outcomes. The agent’s cognitive processes mediate between the sensory inputs and the formulation of a decision problem: according to Denzau and North (1994: 4), “the mental models are the internal representations that individual cognitive systems create to interpret the environment.”

The interplay between agents’ mental models and the formulation of an interactive decision problem is a quagmire, because agents’ perceptions may differ and possibly collide. Hoff and Stiglitz (2016: 26) argue that the “social context creates the set of mental models upon which individuals can draw, and affects the circumstances that prime alternative mental models.” This paper studies a simple model of how agents’ cognition affects their strategic interaction over time.

Our approach keeps distinct the perception of the environment from the choice of actions. The environment is a space of 2×2 symmetric games that differ only over the value of a specific payoff, modeled as an interval $[0, 1]$. The agent cannot perceive each point in this continuum of values: she reduces the complexity of the environment to a finite number of categories.¹ Each category defines a *strategic situation*: the agent is confronted with a continuum of games, but recognizes and processes only a finite number of situations. Conditional on the perceived situation, the agent has correct expectations of the payoffs and maximizes her expected payoff; see Gibbons, LiCalzi and Warglien (2017).

Given his mental model, the agent chooses an action in each of the situations he faces. Because the same situation can bundle different games, this suggests that a failure to perceive differences across games may be a reason² for the behavioral spillovers reported in the literature; see Bednar et al. (2012) or Cason, Savikhin and Sheremeta (2012). The game space studied in this paper contains 50% of variants of stag hunt and 50% of variants of prisoners’ dilemma. All the games are fully ordered by the value of the temptation to destroy joint cooperation with a unilateral defection. In each stag hunt game joint cooperation and joint defection are two equilibria; instead, only joint defection is a (strictly dominant) equilibrium in the prisoners’ dilemmas.

We restrict players to binary categorizations: each agent perceives two situations, but agents may have different categorizations. A situation bundles all the games from an interval in $[0, 1]$. If her perceived situation bundles only prisoners’ dilemmas, the only rationalizable action for an agent is defection. However, if she classifies stag hunts and prisoners’ dilemmas in the same category, both cooperation and defection may be rationalizable. The agent’s categorization directly affects her choice over many distinct games (perceived as one situation).

We allow agents to have different categorizations. However, because agents do not have full control on their cognitive processes, we do not assume that they can choose *a priori*

¹ See Cohen and Lefebvre (2005) for extensive surveys on the cognitive primacy of categorization processes.

² See Steiner and Stewart (2008) for a formal model of spillovers due to learning across games.

which games are to be bundled in the same category. This is an important assumption that differentiates our work from Heller and Winter (2016), who explicitly consider the joint choice of categorizations and actions across a space of games. Mengel (2012) deals with the related question of how agents learn across games when they jointly experiment over categorizations and actions using reinforcement learning. The theme of her work is the closest to ours in the literature. However, we assume that the game space is a fully ordered interval and force agents to learn interval categories, imposing a notion of conformity between the ordinal structures of the underlying environment and of the agents' mental models. Mengel (2012) considers only a finite and unstructured game space, imposing no such restrictions on conformity.

We inquire about three main questions. The first question concerns how binary categorizations affect the strategic choice between cooperation and defection. The main result is that in equilibrium agents share the same categorizations. In our limited setup, a shared understanding of the world is not an assumption but a consequence of equilibration. We also find that agents may get stuck in equilibria where they cooperate less than it would be possible. However, the range of equilibrium payoffs under binary categorizations is the same as if agents could discriminate each game.

The second question concerns what categorizations emerge when agents grope for a better understanding of the environment. Because agents lack control of their cognitive processes, we consider evolutionary dynamics where random mutations in the binary categorization compete for success. Under binary categorizations, the range of evolutionarily stable strategies matches the range of equilibrium strategies. However, a modicum of persistent mutation leads with high probability to a state where stag hunts and prisoners' dilemmas are conflated in the same category, and agents persistently defect almost all the time. When agents grope in the dark, they are very likely to end up (as if distrusting everybody and) playing the max-min strategy of defection everywhere.

The third question, which occupies most of the paper, concerns agents' social learning of categories based on the imitation of successful players over similar games. Because an agent perceives only situations, he cannot discriminate *ex ante* across games in the same category. However, after observing the play of a different pair of agents, he may learn their actions and their payoffs, making inferences about how a similar game in the same category might or should be played. These inferences drive his search for a better understanding and thus affects how categorizations evolve. We compare different models of social imitation and find which characteristics support the emergence of self-confirming shared categorizations where agents cooperate more than in equilibrium. This self-confirming outcome arises because agents' opportunities for learning by imitation are limited by which games they actually realize and by what they perceive as similar. Categorization conceals different games under one situation: under imitation, this form of ignorance is blissful.

2 The model

2.1 The game space

We consider a space $\mathcal{G}$ of 2×2 symmetric games that are instances of either the prisoners' dilemma or the stag hunt. The strategic form for a game in $\mathcal{G}$ has two players (named 1 and

2), who choose between C (for cooperation) and D (for defection). The payoffs to $i = 1, 2$ are as follows: i) if both players cooperate, each receives $1/2$; ii) if both players defect, each receives 0 ; iii) if i cooperates unilaterally, i receives $-1/2$; iv) if i defects unilaterally, i receives a payoff x in $(0, 1)$ which varies across the games in $\mathcal{G}$.

A typical 2×2 game in $\mathcal{G}$ is shown below. The action D is always risk-dominant and the profile (D, D) is always a strict Nash equilibrium.

$$\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
 \begin{array}{c} C \\ D \end{array} & \begin{array}{|cc|} \hline \frac{1}{2}, \frac{1}{2} & -\frac{1}{2}, x \\ \hline x, -\frac{1}{2} & 0, 0 \\ \hline \end{array}
 \end{array} \tag{1}$$

The payoff x received by a unilateral defector is called the *temptation*. The temptation payoff x parameterizes the games in $\mathcal{G}$: as it increases from 0 to 1 , playing C becomes a less attractive reply than D to the opponent's choice of C . If $x > 1/2$, the temptation is high: a player is lured away from C by the strictly dominant strategy D and we have a prisoners' dilemma. If $x < 1/2$, the temptation is low and the profile (C, C) is payoff-dominant: we have a stag hunt with the two strict Nash equilibria (D, D) and (C, C) .

Because the games in $\mathcal{G}$ differ only in the value x of the temptation, we identify each game G_x in $\mathcal{G}$ with the point x in $(0, 1)$. Hence, the game space $\mathcal{G}$ is completely ordered: if $x > y$, then the game G_x has a higher temptation than the game G_y . We assume that the temptation x is a random variable X with uniform distribution on $(0, 1)$.

2.2 Discriminating games

We are interested in how the ability of players to perceive and discriminate games affects how they play them. In general, let K be a partition of $\mathcal{G}$ and let k denote a cell (or situation) of K . Each cell is a *similarity class*: two games in the same cell are not perceived as distinct before being played; see LiCalzi (1995). We say that an agent perceives the same *situation* when he is facing two games from the same cell.

Given a partition K , a (pure) strategy is a mapping from the situations of K to the set of actions $\{C, D\}$. The payoffs attributed to a cell k match expected values across all games in k ; that is, conditional on his ability to discriminate, an agent has correct expectations. Agents choose strategies that maximize expected utility. After a game is played, the agents receive their actual payoffs. In short, *ex ante* the games are perceived and categorized as situations; *ex post*, the payoffs from the actual game played are realized and received.

In our model, the only feature that differentiates two games is the temptation x in $(0, 1)$; therefore, with a slight abuse of notation, we identify $\mathcal{G}$ with $(0, 1)$ and consider partitions over the interval $(0, 1)$.

Under full discrimination, a player perceives each game in $\mathcal{G}$ as distinct. If $x > 1/2$, he perceives a prisoners' dilemma where D is the dominant strategy. Otherwise, he perceives a stag hunt because C is payoff-dominant and D is both the risk-dominant and the max-min strategy.³ Thus, his best reply is to play D if $x > 1/2$ and C or D otherwise. If both players

³ This remains true at the null event $x = 1/2$, when Player 1 gets the same payoff under (C, C) and (D, C) .

fully discriminate, the equilibria in pure strategies are (D, D) for $x > 1/2$ and one of the two profiles (C, C) or (D, D) for $x \leq 1/2$. The expected equilibrium payoffs range from 0 (if they defect for any x) to $1/4$ (if they defect only for $x > 1/2$).

Under no discrimination, a player cannot discriminate two games in $\mathcal{G}$: all games are perceived as the same situation. Because the expected value for X is $1/2$, the payoff matrix perceived by the agent for this situation is

	C	D
C	$\frac{1}{2}, \frac{1}{2}$	$-\frac{1}{2}, \frac{1}{2}$
D	$\frac{1}{2}, -\frac{1}{2}$	$0, 0$

The strategy D weakly dominates C . Under no discrimination, (D, D) is the only equilibrium in weakly dominant strategies, with an expected payoff of 0.

This paper studies a population of agents that simultaneously learn to partially discriminate games and to play equilibrium strategies adapted to their discriminating ability. The key cognitive process is *categorization*: an agent partitions the space $\mathcal{G}$ in a finite number of cells. The agent perceives different situations, but cannot discriminate games within a situation. Because the game space $\mathcal{G} = (0, 1)$ is completely ordered (by the temptation value x), the cells of a partition are intervals and we can associate any (finite) partition with a set of *thresholds* $0 \leq t_1 < t_2 < \dots < t_n < 1$ that define $n + 1$ contiguous cells.⁴

In particular, any threshold t in $(0, 1)$ defines a binary (interval) partition with two situations named ℓ (low) and h (high). When $x < t$, the agent perceives a situation with low temptation that we denote ℓ ; similarly, if $x > t$, he perceives a situation with high temptation denoted h . (What happens at the null event $x = t$ is immaterial.) Given his threshold t , the agent has correct expectations: he has an expected value for X that is $E(X|X < t) = t/2$ when the temptation is low and $E(X|X > t) = (1 + t)/2$ when it is high. Depending on t and whether he perceives a situation with low or high temptation, the agent faces one of the two payoff matrices in Table 1.

Table 1: The two situations under a binary categorization with threshold t .

	C	D		C	D
C	$\frac{1}{2}, \frac{1}{2}$	$-\frac{1}{2}, \frac{t}{2}$	C	$\frac{1}{2}, \frac{1}{2}$	$-\frac{1}{2}, \frac{1+t}{2}$
D	$\frac{t}{2}, -\frac{1}{2}$	$0, 0$	D	$\frac{1+t}{2}, -\frac{1}{2}$	$0, 0$
low temptation (ℓ)			high temptation (h)		

⁴ Technically, because cells overlap at the boundary, thresholds define partitions up to zero-measure sets.

2.3 Strategies and binary categorizations

We focus on learning across games when agents use binary (interval) categorizations. Any binary categorization over $(0, 1)$ is identified by a single threshold t in $(0, 1)$. We allow for the special case of no discrimination by adding the convention that for $t = 0$ all the games are bundled in the unique category $\{x \in (0, 1) : x > 0\}$ and for $t = 1$ that they are bundled in the category $\{x \in (0, 1) : x < 1\}$. Therefore, we assume that the threshold t is in $[0, 1]$.

Given the binary categorization defined by a threshold t in $(0, 1)$, a player perceives two situation (ℓ and h) and has two possible actions (C and D). Hence, she has only four possible strategies: CC (always play C); DD (always play D); CD (play C when ℓ and D when h); DC (play D when ℓ and C when h). If $t = 0$ or $t = 1$, the player cannot discriminate at all and must always play the same action: we associate the strategy DD with $t = 0$ and CC with $t = 1$. We emphasize that the agent cannot perceive x and does not consciously use the threshold t to discriminate between situations with a low or a high temptation. The threshold is only a modeling device to compactly describe the categories used by an agent. We write $A_t B$ to denote the case where the threshold is t and the agent plays action A when $x < t$ and action B when $x > t$. For instance, the strategy $C_{0.3}D$ cooperates when $x < 0.3$ and defects when $x > 0.3$.

Table 1 shows that, for any t in $(0, 1)$, a situation ℓ is perceived as a stag hunt game and a situation h as a prisoners' dilemma. Clearly, any agent with a binary categorization prefers to plays D whenever she perceives a situation h . The only strategic uncertainty concerns how she should play in the ℓ situation. Thus, given any t in $(0, 1)$, the only two rationalizable strategies are $C_t D$ or $D_t D$. In the special cases when $t = 0$ or $t = 1$ and the agent cannot discriminate situations, our convention write $C_0 D$ for “always D ” and $C_1 D$ for “always C ”. Both “always D ” and “always C ” are rationalizable strategies; however, because “always D ” weakly dominates “always C ”, this latter is not evolutionary stable. Without loss of generality, we restrict the search for optimal strategies to the class $C_t D$ for some t in $[0, 1]$.

The rest of the paper focuses on the dynamics by which agents evolve or learn their understanding of situations (and hence their strategies). Given the mapping between binary categorizations and thresholds, this can be done by studying the evolution of agents' thresholds in the set $[0, 1]$.

3 Equilibria over thresholds

In our model, the viable strategies can be written as $C_t D$ for some t in $[0, 1]$. Heller and Winter (2016) suppose that agents can choose how they partition the space of games, and thus act as if they can select their own threshold t . In their setup, our first result is that the only equilibria in pure strategies require agents to share the same categorization, for some threshold t in $[0, 1/2]$. Intuitively, if agents start off with different categorizations, there are games where they miscoordinate and play (D, C) , making one of them a “sucker” who scores $-1/2$ and regrets how his categorization led him to play.

Proposition 1. *Assume binary categorizations with t in $[0, 1]$. If agents have full control on the thresholds, the only equilibria in pure strategies are symmetric: they share a common*

threshold t in $[0, 1/2]$ and play $C_t D$.

Proof. Suppose $t_1 \leq t_2$. If 1 plays $C_{t_1} D$ and 2 plays $C_{t_2} D$, then the pair plays (C, C) with probability t_1 , plays (D, C) with probability $t_2 - t_1$ and plays (D, D) with probability $(1 - t_2)$. The expected utility is $(t_1 - t_1^2 + t_2^2)/2$ for Player 1 and $(2t_1 - t_2)/2$ for Player 2. In particular, if $t_1 = t_2 = t$, each player scores $t/2$. The payoff submatrix of Player 1 for two arbitrary strategies $t_1 \leq t_2$ is

$$\begin{array}{cc}
 & C_{t_1} D & C_{t_2} D \\
 C_{t_1} D & \begin{array}{|c|} \hline \frac{t_1}{2} \\ \hline \end{array} & \begin{array}{|c|} \hline \frac{t_1 - t_1^2 + t_2^2}{2} \\ \hline \end{array} \\
 C_{t_2} D & \begin{array}{|c|} \hline \frac{2t_1 - t_2}{2} \\ \hline \end{array} & \begin{array}{|c|} \hline \frac{t_2}{2} \\ \hline \end{array}
 \end{array} \tag{2}$$

Clearly, if 2 is using $C_{t_1} D$, then playing $C_{t_2} D$ gives 1 a strictly lower expected payoff than $C_{t_1} D$; hence, it is never optimal for an agent to have a strictly higher threshold than his opponent. On the other hand, suppose that 2 uses $C_{t_2} D$: if $t_2 \leq 1/2$, then 1 maximizes his expected payoff when $t_1^* = t_2$; if $t_2 > 1/2$, then 1 maximizes his expected payoff when $t_1^* = 1/2$. Combining these two conditions yields that the equilibria in pure strategies are those where players share the same threshold t^* in $[0, 1/2]$. $\square$

Importantly, and differently from Heller and Winter (2016), we assume that agents have no full control on their own categorization. This implies that, if two agents share the same threshold t in $[0, 1]$, then their strategies $C_t D$ are perfectly coordinated: they play (C, C) when $x < t$ and (D, D) when $x > t$. Because they never miscoordinate (except perhaps at the null event $x = t$), they find themselves at an equilibrium for any t in $[0, 1]$. In particular, for $t > 1/2$, each agent receives an expected payoff $t/2$: the “non-equilibrium” threshold t is not upset because agents are unaware that they can profitably “deviate” to another categorization with a lower threshold.

When agents can consciously change their binary categorizations, Proposition 1 implies that the range of equilibrium payoffs is $[0, 1/2]$. On the other hand, when they lack full control, any shared categorization t in $[0, 1]$ may survive and thus the range of equilibrium payoffs expands to $[0, 1]$. When agents happen to share a threshold $t > 1/2$, they achieve a higher payoff: here, ignorance is bliss. But, if agents start with different thresholds, it is not clear how they come to a shared categorization and which ones might emerge. The rest of this paper explores the social dynamics of agents who have no full control on their perception, but use concurrent play from other players to draw inferences and revise their categorizations.

4 Evolutionary and stochastic stability

Even if agents do not have full control on their categorizations, they may grope for a better understanding of their environment. We model an agent who changes his binary categorization to a smaller (resp., greater) ℓ cell by a shift to a lower (higher) threshold. If t decreases, the agent is more likely to perceive a prisoners’ dilemma situation, and thus plays D more often; otherwise, if t increases, he is more likely to perceive a stag hunt situation where C is

rationalizable. Therefore, agents who increase their thresholds may learn to cooperate more often. In particular, when $t > 1/2$, some prisoners' dilemma games ($x > 1/2$) are bundled under a stag hunt situation and cooperation may spill over to them.

This section draws an analogy between the biological evolution of genes and the cultural evolution of categorizations. We imagine that, occasionally, an agent experiments with a different categorization and becomes a "mutant". If her new threshold is sufficiently successful in the current environment, it replicates and possibly displaces the predominant threshold. The random mutations model a search process over categorizations by *trial and error*, because agents cannot pursue purposeful selections on how they perceive the environment.

The natural first step is to consider the evolutionary stability of agents' threshold-based strategies. Abusing notation, we refer to a strategy C_tD by the value t in $[0, 1]$ of the threshold. Note that in our model the interaction among agents may be viewed both as a two-player symmetric game and as a population game.

A strategy t is *evolutionarily stable* (ES) if, for every strategy $t' \neq t$, there is $\delta > 0$ such that

$$u(t, \varepsilon t' + (1 - \varepsilon)t) > u(t', \varepsilon t' + (1 - \varepsilon)t)$$

for any $0 < \varepsilon < \delta$. This definition tests the evolutionary stability of the strategy C_tD against one mutant $C_{t'}D$ at a time; in particular, δ does not need to be uniform across all $t' \neq t$. The first result of this section is that all the Nash equilibria in Proposition 1 are evolutionarily stable.

Proposition 2. *Assume binary categorizations with t in $[0, 1]$. The only evolutionarily stable strategies are C_tD with t in $[0, 1/2]$.*

Proof. Consider any strategy t_1 in $[0, 1)$ and a mutant $t_2 > t_1$. The 2×2 payoff matrix over t_1 and t_2 is the same as in (2). Then $C_{t_1}D$ is a strict Nash equilibrium and thus it is evolutionary stable against any mutation to a higher threshold. Similarly, consider a strategy t_2 in $(0, 1/2]$ and a mutant $t_1 < t_2$. Using again the 2×2 payoff matrix in (2), we see that $C_{t_2}D$ is a strict Nash equilibrium and thus it is evolutionary stable against any mutation to a lower threshold. However, if $t_2 > 1/2$, then this strategy is no longer a best reply against a mutant $t_1 < t_2$. Combining all the conditions above yields the result. $\square$

Note that our model is based on a unidimensional continuous strategy space, because of the mapping from C_tD to t in $[0, 1]$. This allows for stronger notions of stability that control for success over a neighborhood of mutants; see f.i. Cressman and Apaloo (2018). However, none of the standard notions of stability selects a unique strategy: once we include deviations to other strategies in the interval between t and t' , the fine details of the process may help or hinder the evolutionary success of a mutant t' . Nonetheless, we appeal to long-run stochastic stability (SS) and argue that C_tD with $t = 0$ is very likely to prevail most of the time among all ES strategies.

Suppose that agents undergo a process of stochastic strategy revision over a set S of strategies C_tD (or just t for simplicity), based on a Darwinian selection dynamics and independent random mutations as in Kandori, Mailath and Rob (1993). We assume that S has a finite nonempty intersection with the set of ES strategies and denote by t^* the minimum

threshold associated with a strategy in S . Our main result is that the strategy $C_{t^*}D$ is the only long run stochastically stable strategy in S . In particular, if $t^* = 0$, defection everywhere will be persistently and recurrently played as the rate of mutation approaches zero.

Proposition 3. *Assume binary categorizations with t in $[0, 1]$. If t^* is the minimum threshold in S , then the only stochastically stable strategy is $C_{t^*}D$.*

Proof. Consider the game where two players can choose any strategy t in S . Its strategic form associates with any two strategies $t_1 \leq t_2$ the payoffs given in the submatrix (2). For any ES strategy C_tD , the strategic profile (t, t) is a symmetric equilibrium with payoff $t/2$. We say that the symmetric equilibrium (t, t) is $1/2$ -dominant if C_tD is a strict best reply against any mixed strategy placing at least probability $1/2$ on C_tD ; see Morris, Rob and Shin (1995). Then the proof follows from Corollary 1 in Ellison (2000: 27), provided we show that t^* is $1/2$ -dominant. But the strategy t^* is $1/2$ -dominant if, for any t in S ,

$$\frac{1}{2}u(t^*, t^*) + \frac{1}{2}u(t^*, t) > \frac{1}{2}u(t, t) + \frac{1}{2}u(t, t^*)$$

or, equivalently, if $t^* + (t^* - t^2 + t^2) > t + (2t^* - t)$. This holds if and only if $t^* < t$, establishing the claim. $\square$

5 Social learning for categorizers

Section 4 shows that the random mutations underlying evolutionary stability yield low levels of cooperation. A strategy C_tD is evolutionary stable only if $t \leq 1/2$. In any finite set that includes at least one of the ES strategies, the only stochastically stable strategy is at the lowest threshold t^* and cooperation is minimal. If $t = 0$ is available and the search for a shared categorization is driven by random mutations over a finite support, after a sufficiently long wait agents would be expected to always defect, effectively unlearning to discriminate stag hunt situations from prisoners' dilemmas.

We argue that the dynamic process underlying evolutionary stability or its variants may not be a realistic model of the search for useful categorizations. Roughly speaking, the evolutionary dynamics tests whether random mutants perform better than the current population and, if so, favor their reproduction. This is a plausible model when the agents of change have no understanding of the environment and thus their search is undirected: an organism cannot know which mutations are more promising, and hence we assume that they occur randomly.

On the other hand, although an agent has no full control of his cognitive processes, he is aware that his own categories affect his choice and hence his performance. Because it takes considerable effort to change how one perceives the world, an agent is unlikely to experiment (and "mutate") randomly. We posit that the agent tries to amass sufficient cause for attempting a change, and that such changes tend to occur in a direction that he hopes to be favorable. In short, we expect that changes should not be purposeless, but instead abide by some form of bounded rationality.

We use two well-known mechanisms to model this bounded rationality. First, an agent initiates an attempt to change his categorization only if his performance in a situation is

sufficiently low. Second, an agent tries to imitate successful categorizations. These two mechanisms are usually called aspiration levels and learning by imitation. When dealing with categorizations, their application requires special care.

Consider the aspiration level. An agent compares her payoff in (say) situation ℓ with others' payoffs in situation ℓ . Because an agent perceives different games as the same situation, she is actually comparing her payoff in a game G_x in ℓ against others' payoffs in different games G_y in ℓ . Moreover, if two agents have different categorizations and $\ell_1 \neq \ell_2$, they will make their comparisons based on different sets of games.

Concerning imitation, agents do not know others' categorization thresholds and hence cannot imitate how others categorize the world. But they can use observed play to make inferences about how others view the world and try to approximate what seems to be a more successful viewpoint. Recall our assumption that each agent plays a strategy $C_t D$ for some t in $[0, 1]$. Consider an agent with threshold $t = 0.7$ who faces a game G_x with $x = 0.6$, categorizes it as ℓ and plays C . Unknown to him, his opponent has a threshold $t = 0.5$ and plays D . The first player receives a payoff -0.5 while the second player scores $x = 0.6$; see the payoff matrix in (1). After observing payoffs and actions, the first player learns that $t_2 < x < t_1$. He can infer that the second player has a smaller ℓ category than his ($t_2 < t_1$) but does not know what t_2 is and cannot imitate it; however, because he observes x and $t_2 < x < t_1$, he may review his own categories so as to shift x from ℓ to h (and play D in the future). That is, he can "imitate" the second player by updating his t towards x .

In the following, we present a few variants for a model of social learning where agents use aspiration levels to initiate change and learn by imitation. We show that, even under the limitations imposed by their categorical perceptions, agents learn to cooperate much more than a mere argument of evolutionary stability suggests. Within our modest scope, this is consistent with the general argument that cultural evolution may achieve different outcomes than biological evolution; see f.i. Boyd and Richerson (2009).

At a high level, the main features of our model for learning categorizations by imitation are the following. We assume that agents play only strategies $C_t D$ for some threshold value t in $[0, 1]$. The threshold is a modeling device: an agent may not be consciously using a threshold t to classify games – it suffices that he perceives the two (interval) categories ℓ and h and that he plays C when ℓ and D when h . Agents do not know and cannot observe others' thresholds. However, for simplicity, we speak of agents who learn their threshold value t instead of agents who learn binary categorizations.

There is a population of n agents, who are randomly matched in each period. Each pair may play a different game G_x . Any agent categorizes the game according to his threshold and plays accordingly. Agents do not review their categorization after each interaction. Once in a while, an agent who realizes a performance below his aspiration level attempts to review his categorization. He compares his performance against the performance of others who played games that he puts in the same situation. Drawing on his observations, he infers the direction in which he should move his threshold and adjust it accordingly.

5.1 Learning by imitation

This section describes the model and some variants in detail, using a pseudo-code notation. We use two classes ($\mathcal{A}$ for agents and $\mathcal{G}$ for games) as well as a few functions. We eschew a formal definition for the simple functions *random_pairs*, *random_member* and *random_number* because their meaning is obvious.⁵

There is a finite population $\mathcal{A}$ of n agents. The state of an agent a in $\mathcal{A}$ is described by four features:

1. *a.threshold*: the current threshold that separates a 's categories ℓ and h ;
2. *a.action*: the last action (C or D) played by a ;
3. *a.xValue*: the x value of the last game G_x played by a ;
4. *a.payoff*: the last payoff received by a after playing G_x .

An agent gathers information about other agents' categorizations by observing their actions and their payoffs. If a player considers a different agent a who has played a game G_x , she can observe what a chose (*a.action*), what game a played (*a.xValue*), and which payoff a received (*a.payoff*). The player ignores the threshold value of agent a (*a.threshold*), because a 's mental model cannot be observed. In short, a player can use three of the four features that describe the state of an agent in order to imitate the fourth one.

Figure 1 provides the pseudo-code describing the structure of the *Threshold Imitation* (TI) algorithm that implements the simulation model. The input parameters are in lines 1–4, the initialization is in line 5, the agents' interaction is in lines 7–18, and the learning is in lines 19–22.

The input parameters are self-explanatory: we provide more details in the next subsection. Each agent is initialized with her own threshold (line 5). Agents are matched and play games at each simulation round, until a break condition B is reached (line 6). At each round of interaction, agents' features are updated (lines 7-18) as follows:

line 7: the agents in $\mathcal{A}$ are randomly matched in pairs;

lines 8–10: each pair of agents (a_j, a_k) plays a game G_x from $\mathcal{G}$ where x is an independent uniform draw from $(0, 1)$;

lines 13–16: for each agent a ,

- the agent's game *a.xValue* is set equal to x (line 12);
- the agent's played action *a.action* is set by the strategy $C_t D$ to be C if *a.xValue* $< a.threshold$ and D otherwise (lines 13-16);

lines 17–18: the two paired agents play their chosen action in the game G_x and the payoff *a.payoff* is computed for each of them.

⁵An implementation of the model (using python) is available at the following GitHub repository:
https://github.com/muehlenbernd/temptation_threshold_imitation

TI Algorithm	
1	Input: set of n agents $A = \{a_1, a_2, \dots, a_n\}$,
2	initialization function <i>initialize_thresholds</i> ,
3	break condition B ,
4	imitator agent choice rule R_I
5	<i>initialize_thresholds</i> (A)
6	while not B :
7	$P = \text{random_pairs}(A)$
8	for $(a_j, a_k) \in P$:
9	$x = \text{random_real}(0,1)$
10	$G = \text{Temptation_Game}(x)$
11	for $a \in \{a_j, a_k\}$:
12	$a.xValue = x$
13	if $a.xValue < a.threshold$
14	$a.action = 'C'$
15	else:
16	$a.action = 'D'$
17	$a_j.payoff = G.utility(a_j.action, a_k.action)$
18	$a_k.payoff = G.utility(a_k.action, a_j.action)$
19	$a_i = R_I(A)$
20	$a_e = \text{random_member_with_higher_payoff}(A, a_i)$
21	if $a_i.action \neq a_e.action$ and <i>is_in_same_situation</i> (a_i, a_e) :
22	$a_i.threshold = a_e.xValue$

Figure 1: The TI algorithm (Threshold Imitation).

No agent updates her threshold value during these rounds. Learning by imitation takes place separately, at the end of a round of interaction (lines 19-22). The learning part works as follows. The choice rule R_I selects an agent, labeled as the *imitator agent* a_i , who attempts to change his categorization (line 19). The choice rule R_I captures how bounded rationality affects the willingness of an agent to undertake this laborious effort. We tested a few different choice rules that are presented in the next section.

The imitator a_i wants to imitate only successful agents (f.i., someone who scored better in the same situation as a_i is). The agent who is targeted for imitation is labeled as the *exemplar agent* a_e (line 20). An imitator applies sequentially three criteria to select his exemplar: 1) she must have scored better than a_i ; 2) she must have played a different action from a_i ; 3) she must have played in a game that a_i categorizes as the same situation that a_i himself played.⁶ These three criteria are implemented respectively by the function *random_member_with_higher_payoff* described in Appendix A.1, by the test $action \neq a_e.action$, and by the function *is_in_same_situation* given in Appendix A.2.

For instance, if a target agent a plays the same action as a_i in the same situation (according

⁶ One could of course change the order in which these criteria are applied. We tested the six possible combinations and all yield high levels of cooperation. The order in the main text provides the highest levels.

to a_i) but gets a higher payoff, a_i does not take him into consideration for imitation because he (correctly) attributes a 's higher payoff to noise: she must have either faced a different opponent's action or another game (categorized by a_i under the same situation).

The three criteria may be checked using only information about a_e that is *ex post* observable by a_i , namely $a_e.\text{payoff}$, $a_e.x\text{Value}$ and $a_e.\text{action}$. When all three criteria are met, the imitator a_i decides that the higher payoff received by the exemplar a_e is due to playing a different action than a_i when facing the same situation. This prompts a_i to initiate an effort to imitate a_e .

The imitator a_i , however, cannot just imitate a_e 's categorization, because $a_e.\text{threshold}$ is not observable. He must be subtler and use available information to move his categorization closer to a_e 's. This is possible because the game space is fully ordered. Two cases are possible:

- Suppose a_i faced an ℓ situation and thus played C . Because a_e played the different action D , he knows that a_e must have played a game that a_e perceives as an h situation. Thus a_i infers $a_e.x\text{Value} > a_e.\text{threshold}$. On the other hand, because a_e played a game that a_i perceives as an ℓ situation, a_i knows that $a_e.x\text{Value} < a_i.\text{threshold}$. Combining the two inequalities gives $a_e.\text{threshold} < a_e.x\text{Value} < a_i.\text{threshold}$. Therefore, a_i knows that *decreasing* her threshold from $a_i.\text{threshold}$ to $a_e.x\text{Value}$ moves her categorization closer to a_e 's.
- Similarly, suppose a_i faced an h situation and played D . Because a_e played C , he knows that a_e faced a game that a_e perceives as an ℓ situation. Then a_i infers $a_e.x\text{Value} < a_e.\text{threshold}$. And, because a_e played a game that a_i perceives as an h situation, a_i knows that $a_e.x\text{Value} > a_i.\text{threshold}$. This yields $a_e.\text{threshold} > a_e.x\text{Value} > a_i.\text{threshold}$. Hence, a_i knows that *increasing* her threshold to $a_e.x\text{Value}$ makes her categorization closer to a_e 's.

In either case, a_i moves her threshold closer to a_e 's by making the observed game $a_e.x\text{Value}$ the boundary between her two new categories (line 22). Such reliance on the observed game takes care of the recognition problem: the agent needs to experience a game before using it to categorize other games; see Ellison and Holden (2014).

This imitation dynamics also implies that the imitator a_i may only decrease his threshold after playing in an ℓ situation, or increases it after playing in an h situation. This property plays an important role in the analysis of the dynamics given in Section 5.3. Next, we present the output for some variants of the TI algorithm.

5.2 Simulations

The first three of the four input parameters for the TI algorithm are set as follows. (1) We let the size of the population of agents be $n = 50$. Testing higher values for n produced no notable impact on the results of the simulation. (2) The threshold initialization rule follows an *equidspaced allocation*, where agent i 's threshold value is set to $\frac{i-1}{n}$, for $i = 1, 2, \dots, n-1, n$. Using a *random allocation*, where each agent is initialized with a random threshold value, yields similar results; we prefer the first rule that avoids injecting additional noise in the simulations. (3) The break condition B that terminates a simulation is to stop when the

distance between the maximum and the minimum value for the current agents' thresholds dropped strictly below 0.01. All our experiments eventually reached this condition so the population came to share (approximately) the same threshold.

The fourth and last input parameter is the choice rule R_I for the imitator a_i , stating under what conditions an agent decides to attempt a revision of his categorization. Our experiments show that the choice rule has a clear impact on the threshold value that the population eventually learns to use. Therefore, we present three different lines of experiments that differ solely in the choice rule R_I .

5.2.1 Experiment I: only suckers imitate

A “sucker” is an agent who plays C against an opponent who plays D : the “sucker” receives a payoff of $-1/2$, which is by far the worst possible outcome. The first line of experiments assumes that only “suckers” attempt to update their thresholds. Equivalently, one may assume that players initiate change only if they receive a strictly negative payoff. (Except for the sucker's payoff, all the games in $\mathcal{G}$ have positive payoffs.)

Accordingly, the choice rule (called “suckers imitate”) is

$$R_I(\mathcal{A}) = \text{random_member}(\mathcal{A}'), \text{ where } \mathcal{A}' = \{a \in \mathcal{A} \mid a.\text{payoff} < 0\}$$

We made 100 simulation runs, all of which reached the break condition B : $\max(T) - \min(T) < 0.01$, with T being the current set of agents' thresholds. In each run, the population of agents always approximates $t^* = 0$ and thus learns to defect over the whole game space $\mathcal{G}$. This dismal outcome matches the stochastically stable strategy discussed in Section 4.

5.2.2 Experiment II: all imitate

The second line of experiments lets everyone attempts to update his threshold. The choice rule (called “ALL imitate”) is

$$R_I(\mathcal{A}) = \text{random_member}(\mathcal{A})$$

Again, we made 100 simulation runs: all reached the break condition B : $\max(T) - \min(T) < 0.01$. We recorded the value $\hat{t} = \frac{\min(T) + \max(T)}{2}$ for the (approximate) shared threshold learned by the population. The datapoints for $\hat{t}$ from 100 simulations are reported using the boxplot labeled [II (ALL)] in Figure 2. (See keys on the right.)

The results of Experiment II are that the population of agents converges to an average threshold value of 0.497. Half of the datapoints are between 0.496 (first quartile) and 0.498 (third quartile), with a median at 0.497. The dispersion is so minute that we conclude that agents come to a shared categorization where the common threshold value is very close to the payoff-best ES value ($t = 0.5$). Agents categorize games so as to play C over almost all stag hunts (G_x with $x < 1/2$) and to play D over all prisoners' dilemmas (G_x with $x > 1/2$). There is no cooperation spillover from the stag hunts to the prisoners' dilemmas.

Figure 2: Boxplots for the shared threshold from Experiments II and III.

5.2.3 Experiment III: agents imitate when below average

“Suckers imitate” restrict imitation to a very specific case. “ALL imitate” opens it to everybody. The third line of experiments postulates an intermediate choice rule. We assume that an agent considers imitation only if he has scored below the average payoff achieved by players who (from the agent’s viewpoint) faced the same situation as he did. Accordingly, the choice function (called “AVG imitate”) is

$$R_I(\mathcal{A}) = \text{random_member}(\mathcal{A}'), \text{ where } \mathcal{A}' = \text{below_average_same_situation}(\mathcal{A})$$

The function *below_average_same_situation* is described in Appendix A.3.

We had 100 simulations runs and all reached the break condition $B: \max(T) - \min(T) < 0.01$. We recorded the value $\hat{t} = \frac{\min(T) + \max(T)}{2}$ for the (approximate) shared threshold learned by the population. The datapoints for $\hat{t}$ from 100 simulations are reported using the boxplot labeled [III (AVG)] in Figure 2.

The results of Experiment III are that the population of agents converges to an average threshold value of 0.567. Half of the datapoints are between 0.545 (first quartile) and 0.591 (third quartile), with a median at 0.573. The dispersion is much higher than in Experiment II, but the average threshold achieved by the population is now clearly above the payoff-best

ES value ($t = 0.5$). In fact, more than 90% of the simulation runs produced a value $\hat{t}$ above 0.5. Experiment III frequently exhibits a cooperation spillover from the stag hunts to the prisoners' dilemmas.

A visual summary of the main result for the three experiments is in Figure 3. Under

Figure 3: The (categorized) strategy learned by agents: C is in light gray, D is in dark gray.

“suckers imitate”, agents learn the threshold $t = 0$: they defect over the whole space $\mathcal{G}$ and achieve the lowest equilibrium payoff of 0. Under “ALL imitate”, agents learn a threshold very close to $t = 1/2$: approximately, they cooperate over the stag hunts and defect over the prisoners' dilemma, achieving close to the highest equilibrium payoff of $1/4$. Finally, under “AVG imitate” agents usually learn a threshold higher than $t = 1/2$: on average, they cooperate over the stag hunts as well as over about 13% of the prisoners' dilemmas, achieving a payoff of about 0.284. The third (intermediate) imitation dynamics achieves an outcome that is socially superior to either of the other two extreme cases. This outcome is supported by a strategy that is not evolutionarily stable: the random emergence of those specific mutants that could displace it is kept under control by the purposefulness of the rules that initiate imitation. The following section explains this mechanism in detail.

5.3 Analysis

The choice rule used to select imitators is a reduced form for the aspiration level that prompt agents to reconsider their categorization. Under “suckers imitate”, only suckers attempt to review their mental model. Under “ALL imitate”, everybody tries. Under “AVG imitate”, only a subset of lower-performing agents initiate imitation. Which behavioral rule prevails has a remarkable impact on the shared categorization learned by the population. This section takes a closer look at each choice rule and its impact on the dynamics of the process.

Every 2×2 game has four possible outcomes. After each pairwise interaction, an agent may be in one of four different states. Accordingly, we refer to an agent as follows:

sucker (S): a player is an S -agent if he plays C and his opponent plays D ;

rewarded (R): a player is an R -agent if both he and his opponent play C ;

punished (P): a player is a P -agent if both he and his opponent play D ;

tempted (T): a player is a T -agent if he plays D and his opponent plays C .

Consider what happens if only one of these types may attempt imitation and change his threshold. Recall that simulations are initialized assuming equispaced thresholds to reduce exogenous noise.

Experiment I concerns the case where only S -agents (suckers) may imitate: the entire population learns to share the threshold $t^* = 0$. This follows from two simple reasons. First, an S -agent plays C when facing an ℓ situation; as discussed in Section 5.1, imitation at an ℓ situation may only decrease the threshold. Second, an S -agent receives the lowest possible payoffs so any agent who is not a sucker is a potential exemplar. In sum, under “suckers imitate”, a player who experience the pain of being a sucker expands her h category, increasing the chance to defect later and making it more likely for other (more cooperative) agents to turn out as suckers themselves.

Suppose that only R -agents are allowed to imitate. Then imitation stops once all agents’ thresholds lie between 0 and $1/2$. The reason is again twofold. First, an R -agent plays C when facing an ℓ situation: then imitation can only decrease his threshold. Moreover, the realized game G_x must have a temptation value smaller than the imitator’s threshold t_i , so $x < t_i$. Second, because an R -agents receives the payoff $1/2$, his only viable exemplars have a payoff above $1/2$. Any of these exemplars must be a T -agent who scored against an S -agents with a threshold $t > x > 1/2$. Once all agents’ thresholds are below $1/2$, no R -agent finds viable exemplars to imitate and learning stops.

Assume that only P -agents imitate. Then the population learns to share the highest threshold in the initial population. First, a P -agent plays D when facing an h situation: then imitation can only increase her threshold. Second, a P -agents receives the payoff 0, so R -agents (who score $1/2$) are viable exemplars. As threshold values increase, cooperation becomes more likely and thus there are more and more R -agents that P -agents try to imitate.

Finally, suppose that only T -agents imitate. Then imitation stops once all agents’ thresholds are above $1/2$. First, a T -agent plays D when facing an h situation: imitation can only increase his threshold. Second, because a T -agents receives a payoff above 0, his only viable exemplars are R -agents who score $1/2$. (A T -agent never imitate another T -agent because they both have played the same action.) Once all agents’ thresholds are above $1/2$, then a T -agent defects only when facing a game G_x with $x > 1/2$ and receives a payoff $x > 1/2$, what is higher than what any R -agent receives. Therefore, a T -agents finds no agents to imitate and learning stops.

Figure 4 summarizes this analysis, showing both the direction in which thresholds are updated and their basins of attraction for the four single-type choice rules.

We are ready to consider the dynamics for the two choice rules “ALL imitate” and “AVG imitate”. Figure 5.(a) shows the cumulated number of threshold updates made by each of the four types of agents (R -, S -, T -, P -) for the first 500 update steps of one simulation run for “ALL imitate”. Initially, updates distribute roughly in the same amount across the four types. Over time, however, agents’ thresholds get closer and thus they choose similar actions over more games. This implies that the joint cooperation (C, C) and joint defection (D, D) are more frequent: hence, R - or P -agents are more likely. Updates from the R -agents drive thresholds down towards $1/2$, whereas updates from P -agents push them up; see Figure 4. These two countervailing forces balance at a threshold value $\hat{t}$ around $1/2$.

Figure 5.(b) has the analogous plot for “AVG imitate”. The R -agents never imitate

Figure 4: Updating thresholds by imitation for the four single-type choice rules.

Figure 5: Cumulated percentages of updates from R -, S -, T -, P -agents.

because they score $1/2$ whereas the average payoff of agents in their same situation cannot be higher than $1/2$. On the other hand, S -agents and P -agents, who score respectively $-1/2$ and 0 , are very likely to attempt imitation. Simultaneously, as thresholds get closer, P -agents become more frequent than S -agents. The prevalence of P -agents eventually drives the updates pushing thresholds up and often above $1/2$; see Figure 4.

6 Stability for learning by imitation

6.1 Long-term behavior

Proposition 3 shows that a small rate of mutation makes low thresholds the most likely long-run outcome under an evolutionary dynamics. This section tests how “ALL imitate” and “AVG imitate” fare under a constant invasion of random mutants. We initialize the TI algorithm with randomly chosen thresholds and let it run until the population settles around a threshold $\hat{t}_1$. (We use the break condition B to adjudicate convergence.)

Once a shared categorization is found, a random portion of the population is removed and replaced with an equally sized group of invaders who are given random thresholds. The replacement concern a percentage randomly chosen at a level between 1% and 50% of the population. We restart the algorithm with such new population and run it until they find another common categorization $\hat{t}_2$; once this occurs, we iterate the above and a new invasion occurs.

After 1,000 iterations, we record the values of the common thresholds $\hat{t}_i$ and count how frequently they fall in one of the ten equispaced intervals $[z, z + 0.1)$, for $z = 0, 1, \dots, 0.9$. The output of these experiments are presented in Figure 6 for both “ALL imitate” (on the left) and “AVG imitate” (on the right).

Figure 6: Distribution of common thresholds under invasions of random mutants.

The frequency histogram for “ALL imitate” shows that the population converge on a common threshold in the interval $[0.4, 0.5)$ in almost 70% of the cases, and otherwise to a common threshold in $[0.5, 0.6)$. Threshold values slightly short of $1/2$ not only are most likely to emerge, but they recur even after random invasions of mutants. This stability is due to the counterbalancing effects from T -agents and R -agents around $1/2$, adjusted for the stronger impact of R -agents in the final steps as described in the previous section.

The frequency histogram for “AVG imitate” shows that the population converge on a common threshold that has a roughly bell-shaped distribution around 0.7, with 97% of the datapoints above $1/2$. The long-term behavior under constant invasion sustains common thresholds with substantial cooperation spillovers: when the common threshold is 0.7, agents cooperate in around 40% of all prisoners’ dilemmas. The output of Experiment III had common thresholds mostly between 0.55 and 0.6. This discrepancy follows because R -agents never imitate: the predominant imitators are P -agents, who tend to push the threshold up towards 1. However, as thresholds increase, defection becomes less likely and P -agents rarer, making room for imitation by S -agents or T -agents that decrease thresholds.

6.2 Resilience after invasions

Under “AVG imitate”, the population learns to share a categorization associated with a strategy that is not evolutionarily stable. This section explains what prevents random mutations

to take over and destroy the cooperation spillovers achieved under “AVG imitate”.

Generally speaking, most evolutionary dynamics (e.g., the replicator dynamics) share three features: 1) a population playing an ES strategy will resist invasion by a sufficiently small group of mutants; 2) if both the current strategy and the invading strategy are ES, the latter may displace the former when the size of the mutant group overcomes the invasion barrier for the current strategy; 3) if the population plays a strategy that is not ES, this will eventually be displaced by some ES strategy. Neither the incumbent strategy nor the mutant change their threshold value.

The stability of a pure strategy t depends on whether t is ES or not. If t is not ES, it will be displaced by an ES mutant. If t is ES, and a mutant t' overcomes its invasion barrier, it will be displaced by the mutant. If t' does not overcome the invasion barrier, it will get extinct. Thus, under most evolutionary dynamics, invasion is an all-or-nothing event: either the invaders take over completely, or they disappear.

The dynamics under “AVG imitate”, instead, leads a successful imitator to change his threshold t . An incumbent strategy t responds to an invasion from another strategy t' by changing the value of t , and so does t' . Intuitively, evolutionary stability deals with two fixed strategies, while “AVG imitate” transforms the values for both t and t' and make them closer. There are no two strategies fighting for the survival of the fittest, but two different categorizations blending into each other.

The strength with which t and t' respond to each other is stronger the greater their initial distance. Moreover, all else being equal, t is more responsive to $t' > t$ than to $t' < t$. Figure 7 illustrates these features. Denote a strategy $C_t D$ by the threshold t . We color the threshold of $C_t D$ in light gray if the strategy is ES, and in dark gray otherwise. Thick arrows represent the displacement of a population’s strategy t . Shifts in the invading strategy are shown as thin arrows for small invasions and as regular arrows for large invasions.

Figure 7: Dynamics under “AVG imitate”.

The relative success of t versus t' is no longer an all-or-nothing event, and it is fairly independent of the evolutionary stability of either strategy. The amount by which they

change depends on the relation of the values t and t' , and on the size of the groups using t or t' . Moreover, the same amount of invaders leads to a stronger (weaker) push to t' , when $t' > t$ ($t' < t$). The dynamics under “AVG imitate” leads almost always to learn shared categorizations with a threshold value above $1/2$; see Figure 6.

Let us review in detail what prevents an ES strategy from invading a strategy that is not ES. First, note that once a population achieves a shared categorization, all agents behave similarly. They either cooperate pairwise or they defect pairwise: there are only R -agents and P -agents. Assume for instance that the shared categorization is not ES, so $t > \frac{1}{2}$. This population is more likely to see R -agents than P -agents.

Consider an invasion from a small group of mutants who play an ES strategy $t' \leq \frac{1}{2}$. There will be some T -agents, when the mutants interact with incumbent agents over a game G_x with $t' < x < t$. The mutants with a low t' , who imitate R -agents, increase their threshold. Incumbent agents decrease their threshold when they happen to be S -agents or R -agents. The R -agents are more frequent, because the probability that incumbents interact within their group is higher than across the mutant group. But R -agents never imitate; see Figure 5.(b). Because T -agents imitate R -agents (but not vice versa), the incumbent population with $t > \frac{1}{2}$ resist invasion from mutants with $t' < t$ at the cost of a small change in their new shared threshold.

7 Discussion

A lot of research about the emergence of cooperation in social dilemmas has paid attention to the prisoners’ dilemma (from Hobbes, 1651/1968 to Nowak and Highfield, 2011) and to the stag hunt game (from Rousseau, 1755 to Skyrms, 2004). Experimental evidence shows that humans are often more cooperative in a social dilemma than the theory predicts the *homo oeconomicus* should be. Some studies suggest that this may a behavioral spillover (Bednar et al., 2012; Cason, Savikhin and Sheremeta, 2012). This work is a first step in exploring a cognitive foundation for this effect. We consider the emergence of cooperation in a space that includes prisoners’ dilemmas and stag hunt games. We provide a formal framework and a computational model of social learning that generate cooperation spillovers.

A standard evolutionary analysis suggests that the most likely strategy to emerge in the long run is defection not only over the prisoners’ dilemmas, but also over all the stag hunts. Considering instead social learning by imitation, we find that “AVG imitate” (based on an intermediate aspiration level) produces spillovers in cooperation: agents cooperate even in those prisoners’ dilemmas that have payoffs closer to the stag hunt games. These spill-over strategies resist displacement by ES mutants because imitation makes them “adapt” and blend with the invaders. A sobering lesson from our study is that a bit of compromise with (receptive) invaders is a more promising avenue to foster cooperation than an all-out battle for survival between stubborn antagonists.

A Functions used in the main algorithm

A.1 random_member_with_higher_payoff

```
1  Input:  set of  $n$  agents  $A = \{a_1, a_2, \dots, a_n\}$ ,  
2           imitator agent  $a_i$   
3   $O = \{\}$ :  
4  for  $a \in A$ :  
5      if  $a_i.\text{payoff} < a.\text{payoff}$ :  
6           $O = O \cup \{a\}$   
7  return  $\text{random\_member}(O)$ 
```

A.2 is_in_same_situation

```
1  Input:  imitator agent  $a_i$ ,  
2           example agent  $a_e$   
3   $c = \text{false}$   
4  if  $a_i.x\text{Value} < a_i.\text{threshold}$  and  $a_e.x\text{Value} < a_i.\text{threshold}$ :  
5       $c = \text{true}$   
6  if  $a_i.x\text{Value} > a_i.\text{threshold}$  and  $a_e.x\text{Value} > a_i.\text{threshold}$ :  
7       $c = \text{true}$   
8  return  $c$ 
```

A.3 below_average_same_situation

```
1  Input:  set of  $n$  agents  $A = \{a_1, a_2, \dots, a_n\}$ ,  
2   $M = \{\}$ :  
3  for  $a \in A$ :  
4       $\text{accumulated\_payoff} = 0.0$   
5       $\text{num\_opponents} = 0$   
6      for  $a' \in A$ :  
7          if  $\text{is\_in\_same\_situation}(a, a')$ :  
8               $\text{accumulated\_payoff} = \text{accumulated\_payoff} + a'.\text{payoff}$   
9               $\text{num\_opponents} = \text{num\_opponents} + 1$   
10      $\text{avg\_payoff} = \text{accumulated\_payoff} / \text{num\_opponents}$   
11     if  $a.\text{payoff} < \text{avg\_payoff}$ :  
12          $M = M \cup \{a\}$   
13 return  $M$ 
```

References

- [1] J. Bednar, Y. Chen, T.X. Liu and S. Page (2012), “Behavioral spillovers and cognitive load in multiple games: An experimental study”, *Games and Economic Behavior* **74**, 12–31.

- [2] R. Boyd and P.J. Richerson (2009), “Culture and the evolution of human cooperation”, *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 3281–3288.
- [3] T.N. Cason, A.C. Savikhin and R.M. Sheremeta (2012), “Behavioral spillovers in coordination games”, *European Economic Review* **56**, 233–245.
- [4] H. Cohen and C. Lefebvre (2005), eds., *Handbook of Categorization in Cognitive Science*, Amsterdam: Elsevier.
- [5] R. Cressman and J. Apaloo (2018), “Evolutionary game theory”, in: T. Başar and G. Zaccour (eds.), *Handbook on Dynamic Game Theory*, Cham (CH): Springer., 461–510.
- [6] A. Denzau and D. North (1994), “Shared mental models: Ideologies and institutions”, *Kyklos* **47**, 3–31.
- [7] G. Ellison (2000), “Basins of attraction, long-run stochastic stability, and the speed of step- by-step evolution”, *Econometrica* **67**, 17–45.
- [8] G. Ellison and R. Holden (2014), “A theory of rule development”, *Journal of Law, Economics and Organization* **30**, 649–682.
- [9] R. Gibbons, M. LiCalzi and M. Warglien (2017), “What situation is this? Coarse cognition and behavior over a space of games”, Working paper 9/2017, Department of Management, Università Ca’ Foscari Venezia, September.
- [10] Y. Heller and E. Winter (2016), “Rule rationality”, *International Economic Review* **57**, 997–1026.
- [11] T. Hobbes (1651/1968). *Leviathan*. Baltimore: Penguin Books.
- [12] K. Hoff and J.E. Stiglitz (2016), “Striving for balance in economics: Towards a theory of the social determination of behavior”, *Journal of Economic Behavior and Organization* **126**, 25–57.
- [13] M. Kandori, G. Mailath, and R. Rob (1993), “Learning, mutation and long run equilibrium in games”, *Econometrica* **61**, 29–56.
- [14] M. LiCalzi (1995), “Fictitious play by cases”, *Games and Economic Behavior* **11**, 64–89.
- [15] F. Mengel (2012), “Learning across games”, *Games and Economic Behavior* **74**, 601–619.
- [16] S. Morris, R. Rob and H. Shin (1995), “ p -dominance and belief potential”, *Econometrica* **63**, 145–157.
- [17] M. Nowak and R. Highfield (2011), *Supercooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. New York: Free Press.
- [18] J.J. Rousseau (1755), *Discours sur l’origine et les fondements de l’inégalité parmi les hommes*. English translation (2011): *Discourse on the origin and foundations of inequality among men*, New York: Bedford/St. Martins.
- [19] B. Skyrms (2004), *The Stag Hunt and the Evolution of Social Structure* . Cambridge (UK): Cambridge University Press.
- [20] J. Steiner and C. Stewart (2008), “Contagion through learning”, *Theoretical Economics* **3**, 431–458.