The evolution of taking roles[*]

Florian Herold[†]and Christoph Kuzmics[‡]

10 March 2020

Abstract

Individuals are randomly matched to play an ex-ante symmetric hawk-dove game. Individuals assume one of a finite set of observable labels and condition their action choice on their opponent's label. We study the evolutionary stability of chosen labels and their social interaction structure. Evolutionarily stable social structures differ for games in which a dove player prefers the opponent to play hawk (anti-coordination games), and those in which everyone prefers their opponent to play dove (conflict games). Non-trivial hierarchical social structures can only emerge in anti-coordination games. Egalitarian social structures can emerge in both, but are more fragile in conflict games.

Keywords: Evolution, Hawk-Dove Games, Roles, Social Structure
JEL-Codes: C72, C73

[*] We thank Karl Schlag and participants at seminars in Nottingham and Bonn, at the young researcher workshop in Munich, at the annual meeting VfS in Frankfurt and at the Annual Meeting of the EEA in Toulouse for helpful comments and suggestions. Florian Herold gratefully acknowledges funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement PCIG11-GA-2012-322253.

[†] Department of Economics and Social Sciences, University of Bamberg, Feldkirchenstr. 21, 96045 Bamberg, Germany, florian.herold@uni-bamberg.de

[‡] Department of Economics, University of Graz, Universitätsstrasse 15, 8010 Graz, Austria, christoph.kuzmics@uni-graz.at

1 Introduction

We study the endogenous evolution of social structures that govern who does what in economically motivated human or animal interactions. Individuals in our model can freely adopt any publicly observable social role (or label, as we call them in the model) before they interact. When they interact, individuals can then choose more or less assertive or attractive actions. Following the literature, we call the more assertive actions "hawkish" and the less assertive actions "dovish" in our model. Individuals can make this choice contingent on the social roles that they and the individuals with whom they interact adopt. The more well-balanced the mix of actions among the interacting group of individuals, the higher the joint economic success they provide; those who choose the more assertive actions, however, gain relatively more than those who do not.

A social structure combines two things: a distribution of adopted social roles, and a convention describing how individuals behave as a function of the adopted social roles. We study the question how non-trivial social structures can evolve (be evolutionarily stable), even when all social roles can be freely adopted, and, if they do, how efficient these could be.

We have a wide range of situations of human or animal interaction in mind. Foremost, we think of social structures on a small scale. Consider, for instance, the case of two (or more) animals competing over a resource, as in Maynard Smith and Price (1973). Fighting is likely detrimental to all who engage in it. If one animal is prepared to fight (i.e. act hawkishly) and the other is not (i.e. act dovishly), they would (efficiently) resolve the conflict without a fight. Or consider the case of animals hunting together in a pack to hunt for bigger prey. Each pack member needs to perform a distinct action to guarantee the overall success of the hunt. One would expect that not all actions are equally attractive to each pack member; some actions are probably riskier or require more energy than others. In both examples, the question of which animal adopts which action is central to the conflict. In either problem, would we expect a non-trivial social structure to evolve among animals of equal ability (or strength)?

In the human world, these interactions could represent the internal task allocation among members of a team (see, e.g., Forsyth (2016, Chapter 12) for a textbook treatment of team dynamics). In particular Forsyth (2016, p. 391) refers to a "task conflict" for such problems. See also Forsyth (2016, p. 262-266) for the similar problem of who should adopt the leadership role (provided one is required). Consider, for instance, two members of a team discussing the best way to carry out a joint project. A team member who acts more assertively than the other can profit by pushing the project in his or her preferred direction. But if both members act very assertively, the team spirit may suffer, which hurts both. Who takes on the leadership role in such a discussion may depend on simple social cues, such as clothing, which are not directly relevant to the interaction except that they can influence social expectations. In some social environments, a convention may develop suggesting that the person with the more formal dressing style is expected to take on the leadership role. Presumably, in such a hierarchical social structure, a formal dress code will become very popular. In other social environments, a more complicated convention might develop in which someone in formal clothes takes the lead when meeting someone in working clothes; the person in working clothes takes the lead when meeting someone wearing sports clothes; but the person in sports clothes takes the lead when interacting with someone in formal clothes. Such a social structure would presumably result in a more equitable mix of outfits and leadership positions. We want to understand which of these social structures are stable under which circumstances and for which class of interactions.

On a larger scale, the results of our analysis may also shed some light on social structures in entire societies. Specialization is a hallmark of economic activity. Different societies have developed different social structures that coordinate which role or specialization people take. Historically, such social systems were often rather hierarchical. The Indian caste system "allowed" people of higher castes to perform the more profitable (hawkish) tasks and assigned members of lower castes the less profitable (dovish) tasks (see, e.g., Brown (1965, p. 107)). Similarly, the social hierarchy of "the three estates" in medieval Europe prescribed different tasks to different estates: The first estate (clergy) performed (the more hawkish) religious tasks including preserving religious tradition by studying; the second estate (nobility) took over the (also hawkish) tasks of governing and protecting the people; and the third estate (peasants and the bourgeoisie with its social sub-structure of different guilds) carried out their respective (relatively more dovish) tasks of production. The successful overall production requires the right number of people to perform each of these activities. But how efficiently did these hierarchical social systems allocate human resources? Some societies developed less hierarchical social systems. After the Enlightenment movement with its call for egalité, for instance, some societies prospered economically under more egalitarian social systems with checks and balances and a separation of powers. Can an egalitarian social system enhance economic efficiency?

Our model is far too simplistic to do justice to any particular situation or society, and we do not attempt to model each situation separately with models closely tailored to the specific problem. This model, however, does allow us to identify a key distinction that partitions the set of such interactions into two categories and plays a key role in determining the likelihood of the emergence (or evolution) of a non-trivial social structure and, if applicable, the kind of social structure we would be likely to see. This finding allows us to provide a possible explanation for certain social structures (or lack thereof) observed in animals and humans.

We model all these situations as simple 2-by-2 hawk-dove games, in which individuals can adopt either a more (H for hawkish) or a less (D for dove) assertive action, with payoffs given as follows.[1]

$$\begin{array}{c|cc} & H & D \\ \hline H & c & a \\ D & b & d \end{array}$$

We assume that $b > c$ and $a > d$ in order to capture that H and D are best responses to each other. Furthermore, we assume throughout that $a > b$, capturing the conflict that H fares better than D in the asymmetric equilibria.

The key distinction that we identify in this paper and that governs the likelihood of the emergence of a non-trivial social structure is whether or not $d < b$ or $d > b$. When $d < b$ then a D strategist prefers his or her opponent to choose H. The (H, D) outcome is then Pareto-better than the (D, D) outcome (and in all games better than the (H, H) outcome). We refer to such situations as anti-coordination games and interpret these games as models of task allocation and job specialization in joint production. The examples of animals hunting in a pack as well as the examples from the human sphere involve task allocation and are intended as examples of anti-coordination games.

When $d > b$ then a D strategist prefers his or her opponent to choose D. We consider examples of animals competing for a resource as typically falling into this category, as both animals playing D could often be interpreted as sharing the resource; this would presumably be better than being the one animal that, because of the opponent's hawkish behavior, withdraws from the conflict with nothing. We refer to such situations as conflict games.[2]

How can we model a social structure? First note that this class of games is also the canonical example in the evolutionary game theory literature for games, in which the evolutionarily stable strategies in the one-population model differ greatly from those in the two-population model. In the one-population model, both players playing the game are drawn from the same population, and the unique evolutionarily stable strategy is the symmetric mixed Nash equilibrium. In the two-population model, the mixed Nash equilibrium is not even a neutrally stable (and thus neither evolutionarily stable) strategy, and the only evolutionarily stable strategies are the two asymmetric equilibria. The cause for this drastic difference in results is that in the one-population model players cannot make their play contingent on

[1] As the game is symmetric only the row-player's payoffs are given. The column players payoffs can be obtained by taking the transpose of the payoff matrix.

[2] We are not completely satisfied with the somewhat generic term "conflict game" for the class of hawk-dove games with $d > b$. Various names for such games have been used in the literature, such as "hawk-dove" "snowdrift", "chicken", "brinkmanship" games or "war of attrition". No classification terminology seems to be universally accepted, especially when it comes to the boundaries and overlaps between different such classes.

their player-position and it is, thus, impossible to play an asymmetric mixed strategy profile.

One interpretation of this finding is that evolutionary forces would probably induce players to exploit any noticeable differences between players to play asymmetrically. But if this is so, one would expect players to try and adopt (if possible) an appearance that allows them to play hawk against dove in as many encounters as possible. One could call such an adopted appearance a role, as it is consistent with the way social psychologists interpret this term (see e.g., Brown (1965, Chapter 4)).[3] Alternatively, one could call it a type or, perhaps better, a label.[4] One could imagine many models of such role adoption. For instance, one model could exogenously provide certain observable appearances; this is in essence dealt with in Selten (1980) and addressed in more detail below. Some roles may also be easier or harder to adopt by different degrees by different people, an idea that is based on that of costly signalling as described by Spence (1973).

In this paper we study a benchmark model in which players can adopt roles completely freely. We identify the kinds of role distributions that could emerge together with the kinds of social (interaction) structures that could emerge between these roles. To do so, we add an arbitrary fixed set of (completely payoff-irrelevant) publicly observable types or labels to the description of the game, assuming that any player must first adopt some label and then choose an action-function that describes the action he or she would choose in response to all possible labels an opponent might have.

Thus, while evolutionary competition is still evaluated across the entire (single) population, both the actions that depend on labels and the distribution of labels evolve. We "solve" the "meta-game" of first choosing a label and then an action in a hawk-dove base game using standard evolutionary stability concepts.[5] More precisely, we consider neutrally stable strategies (NSS) and evolutionarily stable strategies (ESS) of this meta-game, both refinements of symmetric Nash equilibrium.[6] These notions of

[3] One key aspect of role theory in social psychology (see e.g., Brown (1965, Chapter 4)), is that people generally behave differently in the different roles they may find themselves in, such as when they act in the kinship role of a mother versus the occupational role of a doctor.

[4] Note that types play a somewhat different role here as compared to that described in the literature on the evolution of preferences under observability in which a type implies some subjective preferences and, therefore, some particular strategic behavior (compare, e.g., Dekel, Ely, and Yilankaya (2007) and Herold and Kuzmics (2009)). Here types (or labels) are payoff-irrelevant.

[5] Thus, we implicitly make the standard assumption that individuals in this single population are recurrently and randomly matched to play the meta-game.

[6] In Appendix A we investigate further evolutionary notions of stability as a robustness check: evolutionarily stable sets introduced by Thomas (1985), CURB sets introduced by Basu and Weibull (1991), limit-ESS introduced by Selten (1980), and its refinements of uniform limit-ESS, and strict limit-ESS introduced by Heller (2014). The key insights do not change.

	conflict			general			anti-coordination	
	H	D		H	D		H	D
H	0	3	H	0	3	H	0	3
D	1	d=2	D	1	d	D	1	d=0

Table 1: Examples of hawk-dove games

evolutionary stability, introduced by Maynard Smith and Price (1973) and Maynard Smith (1982), are elegant shortcuts that can be used to analyse the long-term evolutionary outcome of behavior. See, e.g., Weibull (1995, Propositions 3.10 and 3.12), who states that every ESS (NSS) is asymptotically (Lyapunov) stable under the replicator dynamics of Taylor and Jonker (1978). The converse is not generally true. One can, thus, argue that an ESS is evolutionarily stable in a strong sense. This comes at the cost of some games not having an ESS. This is also reflected in our setting. We identify an important class of games, however, in which we always find an ESS, and thus an outcome that is evolutionarily stable in a strong sense. But it is also true that not all our meta-games have an ESS. Nevertheless, our meta-games always have a strategy that satisfies the slightly weaker NSS notion of evolutionary stability, which is typically interpreted as stability in the medium run of evolutionary dynamics.

1.1 Summary of Main Results

Our results are best summarized in the final Proposition 4 of this paper. To best explain this here, consider the class of hawk-dove games parameterized by only one parameter, the payoff d that is accrued by both players if both choose D, with payoffs given in Table 1. The game on the left, with $d = 2$, is a conflict game, the game on the right, with $d = 0$ is an anti-coordination game.

There are two key kinds of social structure that can potentially emerge. One is hierarchical, and labels are (socially) ordered. This means that whenever two players with different labels meet, where one label is higher ranked than the other (in this social order), the player with the higher-ranked label plays hawk and the player with the lower-ranked label plays dove. The other structure is egalitarian (or approximately egalitarian when the number of types is even). In such a social structure, each label[7] plays hawk against some labels and dove against others in such a way that the overall frequency of hawk played by any type of label is the same. For three labels one can imagine such a structure by considering three people sitting around a table, with each one playing hawk against their right-hand neighbor and dove

[7] We will sometimes write “. . . a label plays . . . ” instead of “ . . . a player with a specific label plays . . . ”.

against their left-hand neighbor.

As an example, consider the case with three labels, denoted T, M, and B. The hierarchical and egalitarian label structures, respectively, can be represented in terms of the induced "label game" given by the following two matrices, where $u^* = \frac{3}{4-d}$ is the payoff in the symmetric mixed strategy equilibrium of the hawk-dove game.

$$\begin{array}{c|ccc} & \text{T} & \text{M} & \text{B} \\ \hline \text{T} & u^* & 3 & 3 \\ \text{M} & 1 & u^* & 3 \\ \text{B} & 1 & 1 & u^* \end{array} \qquad \begin{array}{c|ccc} & \text{T} & \text{M} & \text{B} \\ \hline \text{T} & u^* & 3 & 1 \\ \text{M} & 1 & u^* & 3 \\ \text{B} & 3 & 1 & u^* \end{array}$$

hierarchical egalitarian

For the hierarchical label structure, T is the "top" label in the social hierarchy and plays H against all other labels. Label M is the "middle" label, who plays D against label T and H against label B. Label B is the "bottom" label, who plays D against all other labels. In the egalitarian label structure, all labels play H against half of the other labels and D against the remaining half of the other labels. Thus, they are all in equal or "egalitarian" positions.

The evolutionary stability results, as well as their welfare consequences, depend on d, and we now review them gradually taking d from $+\infty$ to $-\infty$ for this class of games. These results are generalized for all games considered in this paper in Proposition 4.

Suppose first that $d \geq 3$. This case is not considered in this paper as it is a trivial case. Then action D is strictly dominant (weakly if $d = 3$) and, in any equilibrium, every player plays D with probability 1.

If d is such that $\bar{d} = 2.5 < d < 3$ (conflict game), then any NSS has only one label in its support and earns the expected payoff $u^* = \frac{3}{4-d}$. This is efficient (among all pre-stable equilibria of the meta-game).[8] No ESS exists in this case.

If $\frac{a+b}{2} = 2 < d < \bar{d}$ (conflict game), then any NSS has only one label in its support and earns the expected payoff $u^* = \frac{3}{4-d}$. This NSS is inefficient and achieves the lowest payoff among all pre-stable equilibria of the meta-game. No ESS exists in this case.

Suppose $b = 1 \leq d < \frac{a+b}{2} = 2$ (conflict game). Then multiple NSS exist with distinct payoffs: there are at least two NSS when the label space has an odd number of labels. The NSS with only one label in its support gives the

[8] The meta-game has many symmetric Nash equilibria. A symmetric Nash equilibrium of the meta-game is pre-stable if, whenever two different labels meet, they anti-coordinate with one label playing hawk and the other dove; when two players using the same label meet, they play the mixed equilibrium of the hawk-dove game. There are many pre-stable equilibria, and any NSS and ESS must be pre-stable, see Lemma 2.

base game	parameters	single label NSS	multiple-label NSS
conflict	$d > \bar{d} = \frac{a^2+b^2-c(a+b)}{a+b-2c}$	exists and efficient	do not exist
	$d \in \left(\frac{a+b}{2}, d\right)$	exists and inefficient	
	$d \in \left(b, \frac{a+b}{2}\right)$		exist, are ESS, egalitarian efficient
anti-coordination	$d < b$	does not exist	

Table 2: Summary of key findings

minimum payoff among all pre-stable equilibria. If there is an odd number of possible labels, then an egalitarian NSS (which is also ESS) exists that gives the maximum possible expected payoff.

Finally, if $d < b = 1$ (anti-coordination game) multiple ESS (and thus also multiple NSS) exist. An egalitarian (for an odd number of labels) and approximately egalitarian (for an even number of labels) ESS are among these, which yield the maximum possible expected payoff. Hierarchical ESS also exist and are payoff-dominated by the egalitarian or approximately egalitarian ESS.

In addition to this result, in Section 4.2 we discuss the robustness of these results regarding the potential emergence of additional labels. In this robustness discussion, we find that the results for anti-coordination games are somewhat robust to the introduction of additional labels, while adding more labels in conflict games can easily lead to the collapse of any (also egalitarian) social structure (that could emerge in some conflict games, i.e. when $b = 1 \leq d < \frac{a+b}{2} = 2$).

More generally, our results are roughly summarized in Table 2.

1.2 Related Literature

Hawk-dove games were among the first games analyzed in evolutionary game theory, starting with the seminal work by Maynard Smith and Price (1973) and Maynard Smith (1982). They analyzed these games in the single-population setting and established the evolutionary stability of the only symmetric (and mixed) equilibrium in these games. Selten (1980) demonstrates that, if players have different roles, only the asymmetric pure strategy equilibria are evolutionarily stable. See Oprea, Henwood, and Friedman (2011) for an experiment that provides strong empirical support for these findings. See also Benndorf, Martinez-Martinez, and Normann (2016) for an exogenously fixed "in-between" model between a single and a double population evolutionary model and experiment. While in Selten (1980) these different roles are given exogenously, we are interested here in examining their endogenous evolution.

Evolutionary papers on cheap-talk games are most closely related to this paper. We can redefine payoff irrelevant labels as cheap talk messages and search for the NSS or ESS of these games. An early paper that describes an anti-coordination game with a specific type of cheap talk messages is Farrell (1987). Farrell allows only for a specific type of communication that corresponds to our hierarchical structure in anti-coordination games and analyzes the corresponding Nash equilibria.

Most of this related cheap-talk literature, including Robson (1990), Sobel (1993), Blume, Kim, and Sobel (1993), Schlag (1993), Wärneryd (1993), Schlag (1995), Kim and Sobel (1995), Bhaskar (1998), Banerjee and Weibull (2000), and Hurkens and Schlag (2003), focusses on coordination games and investigates how far cheap talk does - or does not - support selection against inefficient equilibria. The most closely related formal setup to our work is that described in Hurkens and Schlag (2003) and Banerjee and Weibull (2000). While both papers focus on coordination games, the work by Hurkens and Schlag (2003) also has a section on a task allocation game which falls into our sub-class of anti-coordination games.

For this task-allocation game, they find necessary conditions for ESS that correspond to our conditions (a), (b) and (d) in Lemma 2. Our lemma adds to their necessary condition, beyond the added full generality, by providing a full characterization of ESS. In their proofs of the lowest and highest payoffs in any ESS, they also use constructions corresponding to what we call hierarchical label structure and, respectively, egalitarian or approximately egalitarian label structures. They also conjecture that these results extend to a larger class of anti-coordination games. Our Propositions 1, 2, and 3, which also cover, among others, also all anti-coordination games, confirm this conjecture. More importantly, in Propositions 1, 2, and 3 we analyze all 2×2 games with the best response structure of hawk-dove games. These propositions cover conflict games as well as anti-coordination games, and identify the key distinction between these two (mutually exclusive and jointly exhaustive) subclasses of hawk-dove games, which has not been discussed in the literature before.

We return to the relation of our work to that of Farrell (1987), Banerjee and Weibull (2000), and Hurkens and Schlag (2003) and describe it in more detail in Section 4 after we have derived our results.

2 Model

This paper studies a special class of symmetric, two-player, two-strategy games with a pre-game cheap-talk phase. We call the two-by-two game the *base game*, and the base game plus the cheap-talk phase the *meta-game* as in Banerjee and Weibull (2000).

2.1 The Base Game

The *base game* is a symmetric 2x2 game given by the payoff matrix

$$\begin{array}{c|cc} & H & D \\ \hline H & c & a \\ D & b & d \end{array},$$

with the following restrictions.[9] We assume throughout that $a > b$.

The key restrictions we impose on our base game are $b > c$ and $a > d$. These last two restrictions imply that the best response to H is D and to D is H. This means we rule out dominant strategy games and coordination games and this is all we rule out.[10] We shall, thus, call the base games we consider here (general 2×2) hawk-dove base games or simply *H-D base games.*

The results presented in this paper differ crucially for two (disjoint and jointly exhaustive) subclasses of the class of H-D base games. The crucial distinction is how b compares to d. When $b \leq d$, a player always prefers the opponent to play the dove strategy D, independently of his or her own choice of action. We shall call such games *conflict games.* In contrast, when $b > d$, a player who did commit to playing action D would actually prefer the opponent to play action H. We call such games *anti-coordination games.* The following lemma collects a few immediate and mostly well-known facts about this class of games, which are useful for the further analysis of the meta-game.

Lemma 1. *An H-D base game (with parameters a, b, c, d satisfying $a > b > c$ and $a > d$) has the following properties.*

1. *There are exactly two pure strategy Nash equilibria. These are asymmetric. One player plays H and the other D.*

2. *The game has a unique symmetric equilibrium which is in mixed strategies with probability x^* placed on H, where $x^* = \frac{a-d}{a-d+b-c}$.*

3. *The expected payoff (to both players) in the symmetric (mixed strategy) equilibrium is given by $u^* = \frac{ab-cd}{a-d+b-c}$.*

4. *The payoff in the symmetric equilibrium, u^*, is lower than b, the low payoff in the asymmetric equilibria, if and only if $d < b$ (i.e. if and only if the game is an anti-coordination game).*

[9] Throughout the paper we ignore the possibilities of payoff-ties. Generically there are no payoff-ties and this simplifies the exposition without affecting the main message.

[10] Symmetric 2x2 games are typically classified by the best responses into four categories: two classes of dominant strategy games (efficient dominant strategy games and prisoners dilemma games), coordination games, and hawk-dove (also chicken) games. Compare, e.g., Weibull (1995) or Eshel, Samuelson, and Shaked (1998). Dominant strategy games are of no interest for our purpose. In such games in our model, evolution will always favor those who play the dominant action. Players may send messages but these will not impact play. Coordination games are of interest in our context, but have already been subjected to a thorough analysis by Banerjee and Weibull (2000) and Hurkens and Schlag (2003), among others.

5. *There exists a strategy that limits the opponent's expected payoff to at most* $\min\{u^*, b\}$. *In anti-coordination games, this is achieved by playing* x^*, *in conflict games by playing* H.

6. *Given parameters* a, b, c, *the mixed equilibrium payoff* u^* *strictly increases in the parameter* d, *with* $\lim_{d\to-\infty} u^* = c$ *and* $\lim_{d\to a} u^* = a$.

7. *Keeping* a, b, c *fixed, there is a unique cutoff value* $\bar{d} \in (\frac{a+b}{2}, a)$ *for which* $u^*(\bar{d}) = \frac{a+b}{2}$, *specifically* $\bar{d} = \frac{a^2+b^2-c(a+b)}{a+b-2c}$.
The payoff in the symmetric equilibrium u^* *is higher than the average of the two payoffs in an asymmetric equilibrium* $\frac{1}{2}(a+b)$ *if and only if* $d > \bar{d}$.

Points 1-3 of this Lemma are commonly known and their proofs omitted. The remaining points are not usually emphasized. Their straightforward proofs are given in Appendix B.1. In particular Point 4 is important for the evolutionary analysis of the meta-game.

2.2 The meta-game

Let $G = (A, u)$ be any two-player H-D base game (with $a > b > c$ and $a > d$). Before players play the base game they can freely, i.e. without cost, adopt one of finitely many (commonly observable) *labels*. One could also call these roles or types or (commonly distinguishable) messages. To avoid confusion we only use the terms *label* or *type of label* throughout the main Sections 2 and 3 of this paper. The finite set of labels is given by *label space* Θ. Labels are, therefore, payoff-irrelevant, but perfectly observable and players can condition their play on the opponent's label $\theta' \in \Theta$.

The formal setup is, thus, almost identical to that of Banerjee and Weibull (2000) and Hurkens and Schlag (2003), except that we study the entire class of H-D base-games and focus on the resulting social structures, while their focus was on studying equilibrium selection in coordination games and in an anti-coordination game.[11]

After adopting a label, a player chooses an action function $f : \Theta \to A$. This function is a mapping from the set of labels to the set of actions; namely $f(\theta')$ is the action that a player using action function f chooses against an

[11] There is one formal, but non-substantive, difference between the way we define pure strategies in the meta-game and the way this is done in Banerjee and Weibull (2000). They allow players to condition on both their opponent's as well as their own label. We prefer to reduce the number of strategies, without losing anything, by allowing players to condition only on their opponent's label. We thus follow Schlag (1993), Schlag (1995) and Hurkens and Schlag (2003), in this respect. For a discussion of this issue, see pages 11-12 in Banerjee and Weibull (2000). One advantage of using this reduced form approach is that it helps to clarify when a failure of evolutionary but not neutral stability is simply due to a large number of equivalent strategies or due to a more fundamental problem intrinsic to the game under analysis.

opponent of label θ'. Let $F = \{f : \Theta \to A\}$ the (finite) set of all such action functions.

Define $S = \Theta \times F$ as the (finite) set of pure strategies of the meta-game. Correspondingly, let $\Delta(S)$ be the set of mixed strategies of the meta-game and u the appropriately expanded payoff function. Thus $\Gamma = (S, u)$ defines the finite meta-game.

A mixed strategy $\sigma \in \Delta(S)$ induces both a probability distribution over adopted labels as well as, for each adopted label, a probability distribution over actions. For the purpose of stating (and proving) our results it is useful to have formal expressions of these distributions.

We define $\sigma(\theta) \equiv \sum_{f \in F} \sigma(\theta, f)$, the marginal probability of a player, using mixed strategy $\sigma \in \Delta(S)$, adopting label $\theta \in \Theta$. Furthermore we denote the (conditional) probability that a player of label θ, given strategy $\sigma \in \Delta(S)$ with $\sigma(\theta) > 0$ plays H against an opponent of label θ' by $x_\theta(\theta') = \frac{\sum_{f \in F, f(\theta')=H} \sigma(\theta, f)}{\sigma(\theta)}$.[12]

Note that any $\sigma \in \Delta(S)$ uniquely determines $\sigma(\theta)$ for every $\theta \in \Theta$ and $x_\theta(\theta')$ for all $\theta, \theta' \in \Theta$. The converse is not generally true.[13] However, in order to compute the expected payoff $u(\sigma, \tilde{\sigma})$ which a player with strategy σ obtains against an opponent with strategy $\tilde{\sigma}$, it is sufficient to know $\sigma(\theta)$ and $\tilde{\sigma}(\theta)$ for every $\theta \in \Theta$ and $x_\theta(\theta')$ and $\tilde{x}_\theta(\theta')$ for all $\theta, \theta' \in \Theta$:[14]

$$\begin{aligned} u(\sigma, \tilde{\sigma}) \quad = \textstyle\sum_{\theta, \theta' \in \Theta} \sigma(\theta)\tilde{\sigma}(\theta') \quad & \left[x_\theta\left(\theta'\right) \tilde{x}_{\theta'}\left(\theta\right) c \right. \\ & + x_\theta\left(\theta'\right)\left(1 - \tilde{x}_{\theta'}\left(\theta\right)\right) a \\ & + \left(1 - x_\theta\left(\theta'\right)\right) \tilde{x}_{\theta'}\left(\theta\right) b \\ & \left. + \left(1 - x_\theta\left(\theta'\right)\right)\left(1 - \tilde{x}_{\theta'}\left(\theta\right)\right) d \right]. \end{aligned}$$

2.3 The Solution Concept

We can now use standard concepts such as Evolutionarily Stable Strategy (ESS) and Neutrally Stable Strategy (NSS) from evolutionary game theory

[12] We should perhaps indicate the dependence of $x_\theta(\theta')$ on σ by writing $x_\theta^\sigma(\theta')$. The context should be sufficient for clarity. We shall, for instance, have σ and σ' and then correspondingly $x_\theta(\theta')$ and $x'_\theta(\theta')$.

[13] Consider for instance a meta-game with $\Theta = \{T, B\}$ and the corresponding set of action functions $F = \{f_{HH}, f_{HD}, f_{DH}, f_{DD}\}$, where f_{a_T, a_B} is the action function with $f(T) = a_T$ and $f(B) = a_B$ for $a_T, a_B \in \{H, D\}$. Then the two strategies $\sigma = \frac{1}{2}(T, f_{HH}) + \frac{1}{2}(T, f_{DD})$ and $\tilde{\sigma} = \frac{1}{2}(T, f_{HD}) + \frac{1}{2}(T, f_{DH})$ which are different from each other but lead to the same $\sigma(\theta) = \tilde{\sigma}(\theta)$ for every $\theta \in \Theta$ and $x_\theta(\theta') = \tilde{x}_\theta(\theta')$ for all $\theta, \theta' \in \Theta$.

[14] We could, thus, call two strategies $\sigma \in \Delta(S)$ and $\hat{\sigma} \in \Delta(S)$ *equivalent* if $\sigma(\theta) = \hat{\sigma}(\theta)$ for all $\theta \in \Theta$ and $x_\theta(\theta') = \hat{x}_\theta(\theta')$ for all $\theta, \theta' \in \Theta$. We could then define the corresponding equivalent classes. It turns out, however, that all strategies that satisfy any of our necessary conditions for neutral stability or evolutionary stability are unique in their equivalent class and we do not further need to worry about this issue for our results.

and apply them to our meta-game.[15] One way to define these concepts is as follows.

Definition 1. *A strategy of the meta-game $\sigma \in \Delta(S)$ is a* neutrally stable strategy (NSS) *if and only if the following two conditions hold:*

$$
\begin{array}{lll}
(1) & u(\sigma,\sigma) \geq u(\sigma',\sigma) & \forall \sigma' \in \Delta(S) \\
(2) & u(\sigma,\sigma) = u(\sigma',\sigma) \quad \Rightarrow u(\sigma,\sigma') \geq u(\sigma',\sigma') & \forall \sigma' \neq \sigma.
\end{array}
$$

Strategy $\sigma \in \Delta(S)$ is an evolutionarily stable strategy (ESS) *if and only if the same two conditions hold and the last inequality is strict.*

We refer to condition (1) as the first-order condition or FOC and condition (2) as the second-order condition or SOC. Note that any ESS is also an NSS.

3 Results

3.1 Preliminary results that hold for all H-D base games

One additional definition is useful in our discussion of preliminary results. We call a strategy $\sigma \in \Delta(S)$ a *full-label-support NSS* if σ is an NSS and $\sigma(\theta) > 0$ for all $\theta \in \Theta$. The following lemma provides a full characterization of ESS and full-label-support NSS for any H-D base game.

Lemma 2. *Let $|\Theta| \geq 2$. A strategy $\sigma \in \Delta(S)$ of the meta-game of any H-D base game is an ESS if and only if conditions (a) to (e) are satisfied. It is a full-label-support NSS if and only if conditions (a) to (d) and (e') are satisfied.*

(a) For all $\theta \in \Theta$: $x_\theta(\theta) = x^*$.

(b) For all $\theta, \theta' \in \Theta$, $\theta \neq \theta'$:
$x_\theta(\theta') = 1 - x_{\theta'}(\theta) \in \{0,1\}$.

(c) For all $\theta \in \Theta$: $\sigma(\theta) > 0$ (all labels are played with positive probability).

(d) All strategies in the support of σ earn the same payoff: $u(s,\sigma) = u(\sigma,\sigma)$ $\forall s \in Supp(\sigma)$.

(e) $\frac{a+b}{2} > d$.

(e') $\frac{a+b}{2} \geq d$.

[15] See, e.g. Chapter 2 of Weibull (1995) for a textbook treatment of these definitions and concepts.

The detailed proof of Lemma 2 is given in Appendix B.2. For a specific anti-coordination game (task allocation game) Hurkens and Schlag (2003, Lemma 2(ii)) provided corresponding necessary conditions for evolutionary stability and showed that, in any ESS, all labels must be played with positive probability.

The necessity of conditions (a) and (b) for an NSS and ESS implies that, in any NSS and ESS of the meta-game, every two different labels which are chosen with positive probability must anti-coordinate on $\{H, D\}$ or $\{D, H\}$ when matched against each other. When matched with their own label, the mixed symmetric equilibrium of the base game must be played. This part of the result is in some sense well-known. We know from Maynard Smith (1982) (see, e.g. Weibull (1995, pp. 40-41) for a textbook treatment) that the only evolutionarily stable outcome is the symmetric mixed equilibrium in the single-population case (i.e. here, whenever two individuals of the same label meet). We known from Selten (1980) that the only evolutionarily stable outcome must be a strict, and, hence, pure and possibly asymmetric equilibrium in the multiple-population model (i.e. here, whenever two individuals of different labels meet).

The intuition behind why condition (e), respectively condition (e'), is necessary, is somewhat involved. Suppose we have a full-label-support equilibrium σ, satisfying conditions (a) to (d). Conditions (a) and (b) then imply that, for each label θ, there are actually two pure strategies in the support of σ: They both prescribe the same behavior as the other against all labels other than θ, but one prescribes action H and the other action D against their own label θ. We could call them the hawk and dove varieties of label θ (as used in σ). One can then identify a mutant strategy, let us denote it by μ, which outperforms the incumbent strategy σ exactly when $\frac{a+b}{2} < d$. This mutant strategy μ is as follows: It puts the same probability as σ does on all hawk varieties of all labels θ (as used in σ), and the remaining probability μ places on the dove-variety of a single θ (as used in σ). By condition (d) we have that $u(\mu, \sigma) = u(\sigma, \sigma)$, i.e. the first-order condition for ESS (and NSS) holds with equality. For σ to be an ESS we, thus, need that $u(\sigma, \mu) > u(\mu, \mu)$, i.e. we can focus on the second-order condition for ESS. To see that this inequality is satisfied if and only if $d < \frac{a+b}{2}$, we need a few steps. Note first that, conditional on σ and μ realizing in any hawk variety of any label, both strategies σ and μ are identical and, thus, provide the same payoff in this case. Perhaps a bit harder to see, but nevertheless also true, conditional on the opponent strategy μ realizing in any hawk-variety of any label and both strategies σ and μ realizing in a dove-variety of some label, both μ and σ again yield the same payoff. The potential success of the mutant strategy, thus, depends on how well it does relative to the incumbent strategy against a mutant strategy in the case when all these strategies realize in a dove-variety of some label. In this case, as the mutant strategy places all (remaining) probability on one such label, it yields a payoff of d. The incumbent strategy, on the other hand, as it attaches positive probability to dove varieties for all labels, yields a payoff that is a convex combination of d - when the realized label in σ coincides with the single mutant label (of a dove variety) - and $\frac{a+b}{2}$ - on average when it uses a different label.

Note that Lemma 2, in its characterization of NSS, is mute about strategies σ without full label-support. The next lemma gives a necessary condition for such a strategy to be an NSS as well as a sufficient condition for the existence of such an NSS with certain given properties.

Lemma 3. *Consider a meta-game of any H-D base game with set of labels Θ. For any strategy σ of this game let Θ_S denote the set of labels $\theta \in \Theta$ with $\sigma(\theta) > 0$. Let $\sigma|\Theta_S$ denote the strategy σ restricted to the set of labels Θ_S.*

- *(a) If σ is an NSS of the meta-game with set of labels Θ, then $\sigma|\Theta_S$ is a (full-label-support) NSS of the meta-game with the same H-D base game with the set of labels Θ_S and except for the knife-edge case of $\frac{a+b}{2} = d$ it is even an ESS of that game.*
- *(b) Let $\tilde{\sigma}$ be a strategy of the meta-game with set of labels Θ with support only on $\tilde{\Theta} \subset \Theta$, with $|\tilde{\Theta}| \geq 2$, and $\tilde{\sigma}|\tilde{\Theta}$ is a (full-label-support) NSS of the game restricted to set of labels $\tilde{\Theta}$. Then there exists a strategy σ that is an NSS of the meta-game with the same H-D base game with $\sigma(\theta) = 0$ for all $\theta \notin \tilde{\Theta}$, $\sigma(\theta) = \tilde{\sigma}|\tilde{\Theta}(\theta)$ for all $\theta \in \tilde{\Theta}$ and identical $x_\theta(\theta')$ for all $\theta, \theta' \in \tilde{\Theta}$.*

Lemma 2 is the key result that enables us to identify all full-label-support evolutionarily and neutrally stable strategies. Note that there are no ESS without full label-support. The only remaining cases are NSS without full label-support. The analysis of such strategies, by force of Lemma 3, can be reduced to identifying full-label-support NSS (which are then typically also ESS) in a restricted meta-game, where unused labels were removed, which then is already covered by Lemma 2.

Note that Lemmas 2 and 3 are silent regarding the distribution of labels in Θ in an NSS. Understanding this is where the main contribution of this paper lies and this is what we investigate in Section 3.2. To do so the following definition is helpful.

Definition 2. *Consider a meta-game with an H-D base game and a set of labels Θ. Given a meta-game strategy $\sigma \in \Delta(S)$, we call the induced label-type behavior x, with $x_\theta(\theta') \in \Delta(A)$ the behavior of label θ when meeting label θ' as defined in Section 2.2, the induced* label structure.

- *(i) A* pre-stable label structure *is a label structure that satisfies the following conditions:*

(a) For all $\theta \in \Theta$: $x_\theta(\theta) = x^*$.

(b) For all $\theta, \theta' \in \Theta$, $\theta \neq \theta'$: $x_\theta(\theta') = 1 - x_{\theta'}(\theta) \in \{0, 1\}$.

(ii) The induced label game *of a pre-stable label structure is a 2-player normal-form game with $|\Theta| \times |\Theta|$ payoff-matrix T defined by $T_{\theta\theta} \equiv u^*$ for all $\theta \in |\Theta|$ and for all $\theta' \neq \theta$: $T_{\theta\theta'} = a$ if $x_\theta(\theta') = 1$ and $T_{\theta\theta'} = b$ if $x_\theta(\theta') = 0$.*

For any given pre-stable label structure we can now investigate how the composition of labels evolves in the corresponding reduced form "label game." First, we investigate which distribution of labels leads to a Nash equilibrium in this label game. Then it is straightforward to check whether the corresponding strategies are evolutionarily stable in the meta-game. The relationship is summarized in the following lemma:

Lemma 4. *Consider the meta-game of any H-D base game with finite set of labels Θ with $|\Theta| \geq 2$.*

(a) An ESS $\sigma \in \Delta(S)$ of the meta-game with a given label structure exists, if and only if this label structure is pre-stable, the corresponding label game has a full support Nash equilibrium, and $\frac{a+b}{2} > d$.[16]

(b) A full-label-support NSS $\sigma \in \Delta(S)$ of the meta-game with a given label structure and $\sigma(\theta) > 0$ for all $\theta \in \Theta$ exists, if and only if this label structure is pre-stable, the corresponding label game has a full support Nash equilibrium, and $\frac{a+b}{2} \geq d$.

This result follows immediately from Lemma 2. Furthermore, if σ is an ESS of the meta-game, then it must be the unique ESS with this label structure. To see this note that, if σ is an ESS of the meta-game, then the corresponding strategy must also be an ESS of the corresponding label game. In the label game, it is a full support ESS and must, therefore, be unique.[17] But then no other strategy of the meta-game with the same label game can form an ESS.

To check whether NSS of the meta-game with a given pre-stable label structure without full label-support exist, we can look at the Nash equilibria of the label game (without full support), yet we need to check that all best responses in the meta-game perform weakly worse against themselves than the NSS strategy performs against this mutant strategy.

[16] Within the label game the ESS condition is only $\frac{a+b}{2} > u^*$. Still, for evolutionary stability in the meta-game, we need the more restrictive condition $\frac{a+b}{2} > d$: H-D base-games exist with $\frac{a+b}{2} < d < \bar{d}$ for which egalitarian structures (defined later) form no NSS, for example $a = 3, b = 1, c = 0$, and $d = 2.2$. Then $\frac{a+b}{2} = 2$, $\bar{d} = 2,5$, $x^* = \frac{4}{9}$, and $\sigma = \frac{1}{27}(4, 4, 4, 5, 5, 5)$ (where we restrict σ to the mixtures of pure best responses, and write first the three optimal pure strategies playing H against its own label, and then the three optimal pure strategies playing D against its own label). Then, e.g., the mutant strategy $\mu = \frac{1}{27}(4, 4, 4, 15, 0, 0)$ violates the SOC of NSS (and thus also for ESS).

[17] See, e.g. Weibull (1995, p. 41).

3.2 Main Results

The following definitions prove useful for our further analysis.

Definition 3. *For any H-D base game, consider a meta-game strategy $\sigma \in \Delta(S)$ and an induced pre-stable label structure.*

(a) If there is an order of labels $\succ$ such that $x_\theta(\theta') = 1$ (plays H) if $\theta \succ \theta'$ and $x_\theta(\theta') = 0$ (plays D) if $\theta' \succ \theta$, then x is called a hierarchical label structure. *Given a hierarchical label structure, we call the unique label $\theta \in \Theta$ that satisfies $x_\theta(\theta') = 1$ (plays H) for all labels $\theta' \in \Theta$, $\theta' \neq \theta$ the* top label.

(b) Suppose $|\Theta|$ is odd. If, for every label θ, $x_\theta(\theta') = 1$ (plays H) for exactly half of all labels $\theta' \neq \theta$ and $x_\theta(\theta') = 0$ (plays D) for the other half of all labels $\theta' \neq \theta$, then x is called an egalitarian label structure.

(c) Suppose $|\Theta| \geq 4$ is even, i.e. there is a natural number $k > 1$ such that $|\Theta| = 2k$. If for exactly k labels $x_\theta(\theta') = 1$ (plays H) for exactly half of all labels $\theta' \neq \theta$ and $x_\theta(\theta') = 0$ (plays D) for the other $k-1$ of all labels $\theta' \neq \theta$, and if for the remaining k labels $x_\theta(\theta') = 1$ (plays H) for exactly $k-1$ of all labels $\theta' \neq \theta$ and $x_\theta(\theta') = 0$ (plays D) for the other k of all labels $\theta' \neq \theta$ (and if the resulting label game has a full support Nash equilibrium), then x is called an approximately egalitarian label structure.

Note that these definitions are not empty, meaning that we can construct a strategy $\sigma \in \Delta(S)$ with a hierarchical label structure and also construct one with an egalitarian label structure, provided the number of labels in Θ is odd.[18] If the number of labels in Θ is even, we can construct a strategy $\sigma \in \Delta(S)$ with an approximately egalitarian label structure.[19] See the Introduction for examples of hierarchical and egalitarian label structures for the case of three types, i.e. $|\Theta| = 3$.

The next two propositions investigate the evolutionary stability properties of hierarchical and egalitarian label structures in our two classes of games, games of anti-coordination and conflict games.

Proposition 1. *Let $|\Theta| \geq 2$.*

(a) An ESS of the meta-game with a hierarchical label structure exists if and only if the base game is an anti-coordination game, i.e. $d < b$.

[18] An egalitarian label structure can be visualized in several ways. For instance, one could arrange labels in Θ on a circle such that each label θ plays H against the $(n-1)/2$ labels located clockwise from θ and plays D against all other labels.

[19] One construction can again be visualized by arranging labels in Θ on a circle such that the first k of the $2k$ types θ plays H against the k labels located clockwise from θ and plays D against the other labels. Each label θ' from the remaining $k+1$ to $2k$ labels plays H against the $k-1$ labels located clockwise from θ' and D against the others.

This ESS is the unique, up to a permutation of labels, symmetric full-label-support equilibrium with hierarchical label structure.

(b) *An NSS of the meta-game with a hierarchical label structure exists for all H-D base-games. For anti-coordination games this NSS is the unique, up to a permutation of labels, ESS with hierarchical label structure with full label-support as in (a). For conflict games this hierarchical NSS has only the top label in its support.*

For conflict games, the unique symmetric Nash equilibrium of the hierarchical label game puts all weight on the top label strategy. The corresponding strategy cannot be an ESS of the meta-game but is still an NSS.

Proposition 2. *Let $n \equiv |\Theta| \geq 3$ be an odd number. For any H-D base-game, a strategy of the meta-game exists that induces a symmetric equilibrium with an egalitarian label structure and has full label-support. In this egalitarian equilibrium each strategy receives an average payoff of*

$$v_n \equiv \frac{u^*}{n} + \frac{n-1}{n}\frac{a+b}{2}. \tag{3}$$

(a) *If $d < \frac{a+b}{2}$ (i.e. all anti-coordination games and some conflict games) then such a strategy inducing an egalitarian label structure forms an ESS (and thus also an NSS) of the meta-game.*

(b) *If $d > \frac{a+b}{2}$ (i.e. the game must be conflict game) then such a strategy inducing an egalitarian label structure is not an NSS (and thus also not an ESS) of the meta-game.*

Note that the condition $d > \frac{a+b}{2}$ in Proposition 2.b implies that the game at hand is a conflict game (as $d > b$) and that such conflict games (with $d > \frac{a+b}{2}$) do not have an egalitarian ESS or NSS, while the condition $d < \frac{a+b}{2}$ in Proposition 2.a covers all anti-coordination games ($d < b$) and some conflict games (with $b < d < \frac{a+b}{2}$) and states that all these have an egalitarian ESS.

The proofs of Proposition 1 and 2 are relegated to Appendices B.5 and B.6. They both follow the same steps. First, we compute the unique symmetric full-label-support equilibrium (if it exists). Evolutionary stability - or instability - then follows from Lemma 2.

Note that, the egalitarian equilibrium payoff v_n lies strictly between u^* and $\frac{a+b}{2}$. It tends to $\frac{a+b}{2}$ as n tends to infinity - from below if $u^* < \frac{a+b}{2}$. In case of a base game with $u^* > \frac{a+b}{2}$, v_n approaches $\frac{a+b}{2}$ from above, but note that by Lemma 1, point 7, we know that in this case $d > \bar{d} > \frac{a+b}{2}$ and hence, by Lemma 2, this full-label-support egalitarian equilibrium cannot be an NSS.

Proposition 3. *Let $n \equiv |\Theta| \geq 4$ be an even number. If $d < b$ (i.e. anti-coordination base game), then an ESS of the meta-game of any H-D base-game exists that induces an approximately egalitarian label structure and has full label-support.*

For relative small numbers of labels (we consider $|\Theta| \leq 6$) we can derive all ESS for conflict games explicitly:

Remark 1. *Consider a base game of conflict with $d \in (b, \frac{a+b}{2})$:*

$|\Theta| = 2$*: No ESS of the meta-game exists.*

$|\Theta| = 3$*: The egalitarian structure is the unique (modulo relabeling) ESS of the meta-game.*

$|\Theta| = 4$*: No ESS of the meta-game exists.*

$|\Theta| = 5$*: The meta-game has exactly two ESS (modulo relabeling). One is the egalitarian one, the other contains an egalitarian subgroup of three labels. This subgroup forms a rock-scissors-paper like structure with the remaining two labels (for details see Appendix D.2).*

$|\Theta| = 6$*: For some parameters no ESS of the meta-game exists. For other parameters there exists an (approximately egalitarian) ESS (for details see Appendix D.3).*

In Appendices D.1 and D.3 we explain these results for $|\Theta| \in \{4, 5, 6\}$ in more detail. For a larger number of labels complex structures and sub-structures can emerge that we cannot fully describe. For the case of a subset of labels that is treated equally by all other labels in a pre-stable label structure, we provide some further interesting condition for ESS and NSS (See Appendix D).[20]

3.3 Welfare

Lemma 5. *Consider the average payoff in a label game induced by a pre-stable label structure.*

(a) If $u^* < \frac{a+b}{2}$, then the average payoff is maximized by an equal distribution over all types and minimized by having all weight on one label only.

(b) If $u^* > \frac{a+b}{2}$, then the average payoff is maximized by a distribution that has only one label in its support and is minimized by an equal distribution over all labels.

[20] For example, the different guilds of the third estate in medieval Europe might be considered as such a sub-group.

The next proposition characterizes which NSS and which ESS (if they exist) are efficient among all possible distributions over pre-stable structures (we call this pre-stable efficient) and which are at least Pareto dominating any other NSS. Note that in our setting at least one NSS always exists, and that $a > \bar{d} > \frac{a+b}{2} > b$ is always guaranteed by Lemma 1.

Proposition 4. *Welfare properties of NSS and ESS:*

(a) Assume $d > \bar{d}$: any NSS has only one label in its support and earns the expected payoff u^*, which is pre-stable efficient (since $u^* > \frac{a+b}{2}$). No ESS exists.

(b) Assume $\bar{d} > d > \frac{a+b}{2}$: any NSS has only one label in its support and earns the expected payoff u^*. This NSS is inefficient (since $u^* < \frac{a+b}{2}$) and achieves the lowest payoff in any pre-stable equilibrium. Note that $u^* \in (b, \frac{a+b}{2})$. No ESS exists.

(c) Assume $\frac{a+b}{2} > d \geq b$: For odd $|\Theta| \geq 3$, multiple (at least two) NSS with distinct payoffs exist. The NSS with only one label in its support gives the minimum payoff among pre-stable equilibria. The egalitarian NSS (which is also ESS) gives the maximum expected payoff. Note that the payoff is in $[u^*, v_n] \subset [b, \frac{a+b}{2})$.

(d) Assume $d < b$: In these anti-coordination games, multiple ESS (and thus also multiple NSS) exist: Egalitarian ESS (which exist for odd $|\Theta|$) give the maximum expected payoff. For even $|\Theta|$, approximately egalitarian ESS exist. Hierarchical ESS also exist and are payoff dominated by the egalitarian or approximately egalitarian ESS.

4 Discussion

4.1 Relation and contribution to the cheap talk literature

4.1.1 Cluster points in payoff space

Banerjee and Weibull (2000) study NSS of the meta-game when the base game is a coordination game. Denote by U_n the set of ex-ante expected payoffs in an NSS of the meta-game when the set of labels has n elements. Banerjee and Weibull (2000) show that the union of all these payoffs sets $\bigcup_{n=0}^{\infty} U_n$ has a unique cluster point, which is the Pareto efficient Nash equilibrium payoff.

In contrast, for anti-coordination games, we can show that the set of possible NSS payoffs has multiple cluster points. For instance, as every meta-game has a hierarchical NSS by Proposition 1, there is a cluster point at b (the lower payoff in the asymmetric pure strategy equilibrium of the base game). This follows immediately from Lemma 7.

However, every anti-coordination meta-game with an odd number of labels has also an egalitarian NSS by Proposition 2, which implies that there is another cluster point at $\frac{a+b}{2}$.

Also, for conflict games with $d < \frac{a+b}{2}$, we have at least two cluster points. One cluster point at $\frac{a+b}{2}$, by Proposition 2, and a cluster point at u^*, since we always have a single label NSS in this case.

4.1.2 Connection with Farrell, 1987

Farrell (1987) was, as far as we know, the first to study a model in which there is cheap talk before a game of anti-coordination is played. In his model, players engage in $T \geq 1$ rounds of communication. At each stage $t \leq T$, both players simultaneously and independently of each other send one of two messages, labelled H and D. Farrell (1987) investigates equilibria of this game, in which play after communication is given by the following rule. The player who sent message H at the first point in time at which both players sent different messages (if there is such a time) then plays action H in the anti-coordination game. The other player then plays action D. If both players send identical messages in every round, then they play the symmetric equilibrium x^* in the anti-coordination game.

More formally, let $\theta = (\theta_t)_{t=1}^T$ be a vector of messages, one message for each point in time. Let Θ be the set of all such vectors. In the language used in this paper, this is a set of labels. For each pair of labels $\theta, \theta' \in \Theta$ let $t^*(\theta, \theta') = \min_t\{\theta_t \neq \theta'_t\}$. If $\theta_t = \theta'_t$ for all t let $t^*(\theta, \theta') = \infty$.

In the language used in this paper, Farrell (1987) investigates equilibria of the meta-game that satisfy

$$x_\theta(\theta') = \begin{cases} x^* & \text{if} \qquad t^*(\theta,\theta') = \infty \\ 1 & \text{if} \quad t^*(\theta,\theta') < \infty \text{ and } \theta_{t^*(\theta,\theta')} = H \\ 0 & \text{if} \quad t^*(\theta,\theta') < \infty \text{ and } \theta_{t^*(\theta,\theta')} = D \end{cases} \tag{4}$$

It is straightforward to see that this corresponds to what we call the "hierarchical" label structure. We can reproduce Farrell's (1987) result by noting that every meta-game with a finite number of labels has a (unique - up to relabelling) hierarchical NSS. The ex-ante expected payoff in this NSS is bounded from above by b (the lower payoff in the asymmetric pure strategy equilibrium of the anti-coordination game). As T tends to infinity, the payoff in this NSS tends to b and is, thus, even in this limit, far away from the efficient payoff of $\frac{a+b}{2}$. Note that all this requires that the game is one of anti-coordination.

For conflict games, we know that no hierarchical NSS (or even Nash equilibrium) exists. Imposing the hierarchical structure in these games would yield the result that every hierarchical NSS places probability 1 on a single label. We also know now, however, that other NSS exist, e.g., one based on the egalitarian structure.

For a final example, to see how the egalitarian structure could be implemented in Farrell's (1987) model, consider the case $T = 2$. We then have four "labels" given by $(H,H),(H,D),(D,H),(D,D)$. The egalitarian structure could then be imposed as follows.

	(H,H)	(H,D)	(D,H)	(D,D)
(H,H)	u^*	a	b	a
(H,D)	b	u^*	a	a
(D,H)	a	b	u^*	a
(D,D)	b	b	b	u^*

This label game has an NSS (provided $u^* < \frac{a+b}{2}$), in which the first three labels are used with a probability of $\frac{1}{3}$ each, while label (D,D) is not used.

4.1.3 Connection with Hurkens and Schlag (2003)

Hurkens and Schlag (2003) consider the effects of cheap talk messages on equilibrium in coordination games and - more closely related to our paper - a task allocation game which is a specific anti-coordination game in our terminology. They also conjecture that their results hold for a class of games that corresponds to anti-coordination games, a conjecture that our analysis confirms. Specifically, they are interested in the effect of an option not to take part in cheap talk communication, which they model as a special cheap talk message "stay away from cheap talk" which commits the sender to play one action in the base game without conditioning on the opponent's message. Regarding coordination games, Hurkens and Schlag (2003) find that an inefficient ESS exists if the option to stay away from cheap talk is not available, but, if this option is available, the set of strategies resulting in the efficient outcome is the unique evolutionarily stable set. Most closely related to our work is their analysis of the task allocation game without the option to stay away (Section 4.1): We originally worked only with NSS until we became aware of the connection to their results. The necessary conditions for ESS that they established in their Lemma 2 inspired our characterization of ESS (Lemma 2 in our paper). In their proof of Proposition 3, they construct evolutionarily stable strategies for their task allocation game that correspond to our hierarchical label structure and to our egalitarian or approximately egalitarian label structure, respectively. They continue their analysis of the task allocation game by adding the option again to stay away from communication (Section 4.2) and show that while (in our terminology) the hierarchical label structure equilibrium remains evolutionarily stable, all evolutionarily stable strategies are bounded away from the efficient outcome (and, hence, the egalitarian label structure is not evolutionarily stable anymore). While the option to stay away from pre-play communication seems plausible in the cheap talk context, in our context where players meet automatically and can condition their play on visible features (labels) of the opponent, it seems difficult to visibly commit not to do so.

4.2 Adding one more label

Throughout this paper so far, we always focused on a finite set of possible labels. This means, for full-label-support NSS in particular, that there are no unused labels. But this restriction to a fixed set of labels seems somewhat arbitrary. In this section, we discuss the possible ways one can think about what could happen if one additional, "radically new" label suddenly appeared. Suppose we have a finite set of labels Θ and evolution has progressed to the point that an NSS of the meta-game has established itself. Now suppose a previously unheard of label $\theta^* \notin \Theta$ appears.

One could think about what could happen next in many ways, but we feel it reasonable to assume that the presence of this new label, now available to be adopted by individuals, will not upset the label structure of the incumbent labels. Suppose, for instance, the NSS of the original game (without label θ^*) has a full egalitarian structure. Let us assume that the introduction of the new label does not change that. Having assumed that, we now have to think about what behavior the old labels will display when they meet the new label and, conversely, what behavior the new label will display when meeting other labels.

One way to think about this is to assume simply that evolution will now lead to some new NSS, in which the old labels interact with each other as they did before, but any evolutionarily stable behavior between the new and old labels can emerge. If this is our view, an even sharper distinction can be drawn between anti-coordination games and conflict games.

In anti-coordination games, the new NSS (with the given restriction) may possibly look quite different from the old NSS, but we know that any new NSS must still have at least two labels in its support. So the multiplicity of labels is, in this sense, stable or robust to the introduction of a radically new label.

This is not true for conflict games (and here it does not matter whether $u^* < \frac{a+b}{2}$ or not). For conflict games, the new label can evolve such that it plays H against all other labels and they play D against it. In this case, however, this new label dominates all other labels, and the only NSS with this label structure is the one in which the new label receives probability weight one. In this sense, the possible multiplicity of labels in conflict games is not stable or robust to the introduction of a radically new label.

Returning to anti-coordination games, it is interesting to note that, while the multiplicity of labels is robust, the NSS can, nevertheless, change dramatically from before to after the introduction of the new label. To see this, consider the three-label hierarchical structure with one added label X as given below

$$\begin{array}{c|ccc|c} & \text{T} & \text{M} & \text{B} & \text{X} \\ \hline \text{T} & u^* & a & a & b \\ \text{M} & b & u^* & a & a \\ \text{B} & b & b & u^* & b \\ \hline \text{X} & a & b & a & u^* \end{array}$$

The label game without label X has a unique NSS, and that NSS has full label-support for anti-coordination base games. The meta-game with label X has a NSS with equal support on T,M, and X and an egalitarian structure among these three labels, while B is not in its support. This is true for conflict games (provided $u^* < \frac{a+b}{2}$) but more importantly also for anti-coordination games as long as $\frac{1}{3}(u^* + a + b) > b$. This is, for instance, true when $c = d = 0$ and $b = 1$ and $a = 3$.

5 Conclusion

This may explain why animals, when they compete over food, typically do not adopt a social structure based on costless signals, but instead use costly signalling (such as showing strength). Humans may sometimes sustain more complex egalitarian social systems, which are then highly prosperous for some time. But in conflict games, the stability of egalitarian social structures is vulnerable to the emergence of new dominating label-types, such as a new conquering elite.

On the other hand, in anti-coordination games, we always expect a non-trivial social structure to emerge, even if it is based only on costless signals. Moreover, societies lucky enough to evolve an egalitarian structure achieve higher overall welfare than those that evolve a hierarchical structure.

Finally, the insights provided by our model may even be helpful if we, acting as social scientists or historians, observe a social structure that seemingly contradicts the findings in this paper, for instance, a hierarchical social structures with conflict games. Arguably, in the medieval Western-European social hierarchy, the social interaction between the warrior nobility and their peasant vassals, for instance, corresponded more typically to a conflict game situation than to a pure anti-coordination game. Then the logic of our analysis suggests that there must be some mechanism that prevents the free choice of labels. Indeed, hierarchical societies usually restrict the choice of adapting a higher-ranked role: Members of the nobility in Europe or the caste in India were determined by birth and not by free choice. Marriages across estates or castes were typically not accepted. Catholic clergy members could not have legitimate offspring and usually vowed poverty (especially if they were of lower status). In fact, the logic underlying our analysis suggests that, even for anti-coordination games, prosperity may increase under a hierarchical social structure if the adoption of high labels becomes more costly, at least for some, although inequality may also increase. Such a hierarchical structure, which is based on the lack of movement between social groups, may be more efficient than a hierarchical structure with free movement across social groups. Still, the most efficient social structure, at least in our simplified model that abstracts away from many things, would be an egalitarian structure with free movement across social groups. Our analysis also suggests an explanation for why we do not always observe these in the real world, based on evolutionary stability and the distinction between anti-coordination and conflict games.

Appendix

A Further notions of evolutionary stability

A.1 Evolutionarily stable sets

An evolutionarily stable set (ES set) of strategies $X \subset \Delta(S)$, as defined by Thomas (1985) - see also Weibull (1995, p. 51, Definition 2.6 and Proposition 2.10), is a non-empty and closed set such that, for each $\sigma \in X$, there is a neighborhood U such that $u(\sigma, \sigma') \geq u(\sigma', \sigma')$ for all $\sigma' \in U$ with strict inequality if $\sigma' \notin X$.

If we apply this concept in our setting, we obtain the following proposition. We omit the discussion of the knife-edge case of $\frac{a+b}{2} = d$.

Proposition 5. *Let $\frac{a+b}{2} \neq d$. A set of strategy profiles $X \subset \Delta(S)$ is an ES set of the meta-game if and only if it consists exclusively of ESS of that meta-game.*

Proof: By the result in Thomas (1985), see also Weibull (1995, Proposition 2.11), any union of ESS is an ES set. To see that an ES set cannot include any strategy other than ESS in our setting, let X be an ES set and let $\sigma \in X$ with induced action matrix x. By Weibull (1995, Proposition 2.7) σ must be an NSS. Suppose it is not an ESS. Then, by Lemma 2, it cannot have full label-support. Thus, there is a label $\theta \in \Theta$ such that $\sigma(\theta) = 0$. Define strategy $\sigma' \in \Delta(S)$ such that $\sigma'(\theta') = \sigma(\theta')$ and such that its induced action matrix x' coincides with x for all labels other than θ. I.e. $x'_{\theta'}(\theta'') = x_{\theta'}(\theta'')$ for all $\theta', \theta'' \neq \theta$. Finally, let $x'_{\theta}(\theta') = 1$ (plays H) and $x'_{\theta'}(\theta) = 0$ (plays D). Then let $\sigma^{\epsilon} = \epsilon\sigma' + (1-\epsilon)\sigma$. For $\epsilon > 0$ small enough, $\sigma^{\epsilon} \in U$ and by construction $u(\sigma^{\epsilon}, \sigma^{\epsilon}) = u(\sigma, \sigma^{\epsilon})$. Thus, by definition of an ES set, $\sigma^{\epsilon} \in X$ as well. By the closedness of X, there is a maximum ϵ such that $\sigma^{\epsilon} \in X$. Denote this maximum ϵ by ϵ^*. Suppose now that $\epsilon^* < 1$. The argument above can be repeated in that a (proper) convex combination of σ^{ϵ^*} and σ' is again in a neighborhood of X. This finally implies that σ' must be in X as well. But σ' is not even a Nash equilibrium of the meta-game (the best response is a strategy that places all probability on label θ), cannot be an NSS and, thus, cannot be in X. QED

Another setwise notion that has evolutionary appeal is that of CURB (closed under rational behavior) sets proposed by Basu and Weibull (1991). See e.g., Ritzberger and Weibull (1996) and Balkenborg, Hofbauer, and Kuzmics (2013). A (symmetric) CURB set is a subset B of the set of pure strategies S such that if $s \in S$ is a best response to some (mixed) strategy $\sigma \in \Delta(B)$ then $s \in B$.

Proposition 6. *The only CURB set of the meta-game is the set of all pure strategies S.*

Proof: Consider a CURB set $B \subset S$ and consider any pure strategy $s \in B$. As it is a pure strategy, there is a label $\theta \in \Theta$ such that $s(\theta) = 1$. Consider first the case that $x_\theta(\theta) = 1$ (plays H). The case of $x_\theta(\theta) = 0$ (plays D) is analogous. Then every strategy $s' \in S$ with $s'(\theta) = 1$ and $x'_\theta(\theta) = 0$ (plays D) is a best response to s and, therefore, all such $s' \in B$. By an analogous argument, any pure strategy $s'' \in S$ with $s''(\theta) = 1$ and $x''_\theta(\theta) = 1$ (plays H) is a best reply against such a strategy s' and, therefore, all such $s'' \in B$ as well. Now consider an arbitrary $\theta' \in \Theta$ with $\theta' \neq \theta$. Then by the above argument there is a strategy $s' \in B$ with $s'(\theta) = 1$ and $x'_\theta(\theta') = 0$ (plays D). Then any strategy $s''' \in S$ with $s'''(\theta') = 1$ and $x'''_{\theta'}(\theta) = 1$ (plays H) is a best response to s' and thus $s''' \in B$ as well. A repetition of the first argument completes the proof. QED

A.2 Uniform limit ESS

The games we study here can be viewed as extensive form games in which players first choose a label and then after observing all chosen labels choose an action. Because of the possibility of unreached subgames, extensive form games typically have many strategies that are equivalent for a range of strategies of the opponents. An ESS may, therefore, not exist simply for the trivial reason that there is a strategy that is equivalent on the path of play. On the other hand being an NSS is often considered a very weak evolutionary stability property. It does, for instance, not generally rule out weakly dominated strategies. This concern led Selten (1983) to consider limit ESS as a refinement of NSS that is intended to capture something of the stronger evolutionary stability property of an ESS. As in Selten's (1975) notion of trembling hand perfection, Selten (1983) considers a sequence of slightly perturbed games, in which each strategy has to be used by a strategy-specific minimal probability. A strategy is then a limit ESS if there is a sequence of such perturbed games, there is an ESS in each of the perturbed games in this sequence, and the given strategy is the limit of this sequence of ESSs. Recently, Heller (2014) showed that not every limit ESS is actually also an

NSS. This led Heller (2014) to define the concept of a uniform limit ESS that is indeed a proper refinement of NSS. Thus, every ESS is a uniform limit ESS, and every limit ESS is an NSS. One can show that the key NSS in our analysis that is not simultaneously an ESS, i.e. the NSS for all conflict games ($d > b$) with the hierarchical label structure and single label support only on the top label, is a uniform limit ESS. To do so we need to consider a sequence of perturbed games in which the strategy-specific minimal probabilities are sufficiently larger for those non-top label strategies that play D against the top label than for those that play H against the top label. This argument can, in fact, be extended to show that every non-full-label-support NSS of every meta-game (at least for the most relevant case of $d < \frac{a+b}{2}$) is a uniform limit ESS. Heller (2015) also introduces another, much stronger, refinement of NSS in his concept of a strict limit ESS. This is defined, analogously to a strict trembling hand perfect equilibrium as in Okada (1981), by requiring that the strategy in question is a limit point of such a sequence of ESS's for every sequence of perturbed games. One can show that any non-full-label-support NSS in our analysis is not a strict limit ESS.

B Proofs of the main results

B.1 Proof of Lemma 1

The payoff u of Player 1 in the base game, if he or she plays H with probability $x \in [0,1]$, and Player 2 plays H with probability $y \in [0,1]$ is given by

$$\begin{aligned} u(x,y) &= xyc + x(1-y)a + (1-x)yb + (1-x)(1-y)d \\ &= d + y(b-d) + x\left[a - d - y\left(a - d + b - c\right)\right]. \end{aligned} \tag{5}$$

If the opponent plays $y^* = \frac{a-d}{a-d+b-c}$ then the term in square brackets is zero, Player 1's expected payoff is $u^* = d + y^*(b-d) = \frac{ab-cd}{a-d+b-c}$, independent of his or her own action. Player 1 is thus willing to mix, and (x^*, x^*) is the mixed equilibrium. If $y > \frac{a-d}{a-d+b-c}$, then the term in square brackets is negative, and $x = 0$ is the unique best response. If $y < \frac{a-d}{a-d+b-c}$, then the term in square brackets is positive, and $x = 1$ is the unique best response. Hence, in addition to the mixed equilibrium there are exactly two more Nash equilibria: $(x = 0, y = 1)$ and $(x = 1, y = 0)$. This proves the first three points.

To prove point 4 consider the equivalence of the following inequalities.

$$\begin{aligned} & \frac{ab-cd}{a-d+b-c} < b, \\ \Leftrightarrow \quad & -cd < -bd + b^2 - cb, \\ \Leftrightarrow \quad & d(b-c) < b(b-c), \\ \Leftrightarrow \quad & d < b. \end{aligned}$$

Now we prove point 5: For conflict games ($b \leq d$), we know from the previous point that $b \leq u^*$ and, hence, $b = \min\{b, u^*\}$. The opponent can limit the payoff of a player to b by playing H. (Each player can also guarantee himself a payoff of at least b by playing strategy D. Hence, a player's minmax value is indeed b for conflict games.)

For anti-coordination games ($d < b$), we know from the previous point that $b > u^*$ and, hence, $u^* = \min\{b, u^*\}$. The opponent can limit the expected payoff of a player to u^* by playing x^*.

To prove point 6, consider the function $u^* : (-\infty, a) \to \mathbb{R}$ defined by $u^*(d) = \frac{ab-cd}{a-c+b-d}$ for fixed parameter values a, b, c. Then the first derivative of this function is strictly positive:

$$(u^*(d))' = \frac{(-c)(a-d+b-c) - (ab-cd)(-1)}{(a-d+b-c)^2} = \frac{(a-c)(b-c)}{(a-d+b-c)^2} > 0.$$

Hence, u^* strictly increases in d. Furthermore,

$$(6) \quad \lim_{d\to-\infty} u^*(d) = \lim_{d\to-\infty} \frac{ab}{a+b-c-d} - c\frac{1}{\frac{a+b-c}{d} - 1} = 0 + (-c)\frac{1}{0-1} = c,$$

and, by continuity of u^*,

$$(7) \quad \lim_{d\to a} u^*(d) = u^*(d=a) = \frac{ab-ca}{a-a+b-c} = \frac{a(b-c)}{b-c} = a.$$

To prove point 7, note that

$$\begin{aligned} & u^* = \frac{ab-cd}{a-d+b-c} > \frac{a+b}{2} \\ \Leftrightarrow \quad & 2(ab-cd) > (a+b)(a-d+b-c) \\ \Leftrightarrow \quad & da + db - 2dc > a^2 + b^2 - ac - bc \\ \Leftrightarrow \quad & d > \frac{a^2+b^2-c(a+b)}{a+b-2c}. \end{aligned}$$

We arrive at

$$(8) \quad \bar{d} = \frac{a^2+b^2-c(a+b)}{a+b-2c}.$$

Next, we show $\bar{d} \in (\frac{a+b}{2}, a)$:

$$\begin{aligned}
(9) \qquad & & \bar{d} &> \frac{a+b}{2} \\
&\Leftrightarrow & \tfrac{a^2+b^2-c(a+b)}{a+b-2c} &> \frac{a+b}{2} \\
&\Leftrightarrow & 2\left(a^2+b^2-ac-bc\right) &> (a+b)\,(a+b-2c) \\
&\Leftrightarrow & 2a^2+2b^2-2ac-2bc &> a^2+b^2+2ab-2ac-2bc \\
&\Leftrightarrow & a^2+b^2 &> 2ab \\
&\Leftrightarrow & (a-b)^2 &> 0.
\end{aligned}$$

The last inequality is obviously true, which implies the first inequality.

Furthermore, by assumption, we have $a > b$ and $b > c$, which implies

$$\begin{aligned}
& a(b-c) > b(b-c) \\
\Leftrightarrow \quad & 0 > b^2 + ac - cb - ab \\
\Leftrightarrow \quad & a^2 + ab - 2ac > a^2 + b^2 - ca - cb \\
\Leftrightarrow \quad & a > \frac{a^2+b^2-c(a+b)}{a+b-2c} \\
\Leftrightarrow \quad & a > \bar{d}.
\end{aligned}$$

This completes the proof of point 7. QED

B.2 Proof of Lemma 2

First, we prove that having a full-label-support NSS is a necessary condition for having an ESS of the meta-game. Since ESS always implies NSS (see, e.g. Weibull (1995)), it is sufficient to show that σ can only be an ESS if all labels are played with positive probability (Condition (c)). Suppose to the contrary that a label $\hat{\theta} \in \Theta$ exists with $\sigma(\hat{\theta}) = 0$. Then consider a strategy σ', which is identical to σ except when playing against the label $\hat{\theta}$. That is, there is a label θ with $\sigma(\theta) > 0$ such that strategy σ' describes a different action distribution for this label θ against label $\hat{\theta}$ than σ does. Clearly, such a strategy exists. But then $u(\sigma', \sigma) = u(\sigma, \sigma) = u(\sigma', \sigma') = u(\sigma, \sigma')$ and, hence, σ is not an ESS.

Second, we prove that conditions (a) to (d) are necessary conditions for having a full-label-support NSS (and, therefore, also for ESS) of the meta-game.

To prove that Condition (a) is necessary for NSS, consider a $\sigma \in \Delta(S)$ such that a $\theta \in \Theta$ exists with $\sigma(\theta) > 0$ and $x_\theta(\theta) > x^*$ (the case $x_\theta(\theta) < x^*$ can be proven analogously), where x^* is the symmetric equilibrium probability of H in the base game (see Lemma 1.2). Now consider a strategy $\sigma' \in \Delta(S)$ with the property that $\sigma'(\theta') = \sigma(\theta')$ for all $\theta' \in \Theta$ and

$x'_{\theta'}(\theta'') = x_{\theta'}(\theta'')$ for all $\theta', \theta'' \in \Theta$ such that at least one of θ', θ'' is not equal to θ, and finally $x_\theta(\theta) = 0$. In other words, strategy σ' mimics strategy σ in all respects except that, when adopting label θ and meeting label θ, it plays D.[21] Strategy σ' thus generates different payoff than strategy σ against σ only when both strategies adopt label θ (which happens with positive probability). Then, however, strategy σ' describes the unique best response and, thus, generates a higher payoff than strategy σ does (as σ does not prescribe this best response in this case). This violates the FOC of neutral stability and proves that Condition (a) is necessary for NSS and, therefore, also for full-label-support NSS and ESS.

To prove that Condition (b) is necessary for NSS, consider a strategy $\sigma \in \Delta(S)$ such that there are $\theta, \theta' \in \Theta$ with $\sigma(\theta) > 0$ and $\sigma(\theta') > 0$. An argument similar to the one above that proves part (a) implies that each of the two labels must play a best response to the other label or the FOC of neutral stability will be violated. It remains to be shown that the two labels playing the symmetric equilibrium of the base game against each other cannot be part of an NSS either. Thus, suppose $x_\theta(\theta') = x_{\theta'}(\theta) = x^*$. Then consider strategy $\sigma' \in \Delta(S)$ such that σ' mimics σ in all respects except in its prescription for $x'_\theta(\theta')$ and $x'_{\theta'}(\theta)$. In fact, let $x'_\theta(\theta') = 1$ and $x'_{\theta'}(\theta) = 0$. It is easy to see that $u(\sigma', \sigma) = u(\sigma, \sigma)$, as the only difference that could occur is when labels θ and θ' are employed and then, as σ prescribes the mixed strategy equilibrium strategy x^*, both pure actions of the base game H and D prove an equal payoff against x^*. Thus, the FOC for neutral stability is satisfied with equality. We then need to check the SOC and compare $u(\sigma', \sigma')$ with $u(\sigma, \sigma')$. We find that the following inequalities are equivalent

$$\begin{aligned}
u(\sigma', \sigma') &> u(\sigma, \sigma') \\
\sigma(\theta)\sigma(\theta')a + \sigma(\theta')\sigma(\theta)b &> \sigma(\theta)\sigma(\theta')\left(ax^* + d(1-x^*)\right) + \\
&\quad + \sigma(\theta')\sigma(\theta)\left(cx^* + b(1-x^*)\right) \\
\sigma(\theta)\sigma(\theta')\left[a+b\right] &> \sigma(\theta)\sigma(\theta')\left[(c+a)x^* + (b+d)(1-x^*)\right] \\
a + b &> (c+a)x^* + (b+d)(1-x^*) \\
a + b &> \frac{(c+a)(a-d)}{a-d+b-c} + \frac{(b+d)(b-c)}{a-d+b-c} \\
a(b-c) + b(a-d) &> c(a-d) + d(b-c),
\end{aligned}$$

where the final inequality is true for all hawk-dove games as, by assumption, $a > d$ and $b > c$. Thus, the SOC for neutral stability is not satisfied, and we arrive at a contradiction, proving that Condition (b) is necessary for NSS and, therefore, also for full-label-support NSS and for ESS.

Condition (c) is obviously necessary for full-label-support NSS and, thus,

[21] It is easy to see that such a strategy exists. It may not be unique. See footnote 13.

also for ESS. The necessity of Condition (d) follows directly from the first-order condition of the definition of an ESS. Otherwise σ could not even form a Nash equilibrium with itself.

Third, in order to prove that the conditions in Lemma 2 are sufficient, we will show that conditions (a), (b), and (d) jointly imply a quasi-strict symmetric Nash equilibrium. This fact then allows us to use Lemma 6 below that characterizes NSS, or respectively ESS, for quasi-strict symmetric Nash equilibria. A symmetric Nash equilibrium (σ, σ) is called quasi-strict if σ has all pure best responses to σ in its support. First note that, under conditions (a) and (b), if $\sigma(\theta)$ is specified for all $\theta \in \Theta$ and $x_\theta(\theta') \in \{0, 1\}$ for all $\theta' \neq \theta \in \Theta$, then σ is uniquely determined. In particular, no further equivalent strategy exists, since a pure (contingent) strategy is played for any match of different labels. Note that there are $2|\Theta|$ pure best responses to any σ that satisfies conditions (a)-(d), and that these are all in the support of such a σ: For each label $\theta \in \Theta$, there are two corresponding pure best replies to σ: Select label θ, play against other labels $\theta' \neq \theta$ whatever $x_\theta(\theta')$ the strategy σ prescribes, and play against your own label θ either “H” or “D”.

For a quasi-strict Nash equilibrium (σ, σ), strategy σ is an ESS if and only if the payoff matrix is negative definite with respect to the support of σ (see van Damme (1991), Theorem 9.2.7, and the preceding text on pages 220/221).[22] This corresponds to part (a) of the following lemma. In part(b) of the lemma we adapt this characterization for NSS.

Lemma 6. *Let (σ, σ) be a quasi-strict Nash equilibrium, i.e. the set of pure best responses $B(\sigma)$ corresponds to the support of σ, supp(σ). Define $K \equiv |supp(\sigma)| = |B(\sigma)|$ and let M denote the $K \times K$ matrix corresponding to the restriction of the full payoff matrix to pure strategies in supp(σ).*

(a) Then σ is an ESS if and only if M is negative definite (with respect to the support of σ), i.e.

$$(10) \qquad \mathbf{y}^T M \mathbf{y} < 0 \textit{ for all } \mathbf{y} \in \mathbb{R}^K \textit{ with } \mathbf{y} \neq 0, \sum_{i=1}^{K} y_i = 0.$$

(b) Then σ is an NSS if and only if

$$(11) \qquad \mathbf{y}^T M \mathbf{y} \leq 0 \textit{ for all } \mathbf{y} \in \mathbb{R}^K \textit{ with } \sum_{i=1}^{K} y_i = 0.$$

We relegate the proof of Lemma 6 to the end of this subsection.

[22] Van Damme attributes Theorem 9.2.7 to Haigh (1975) and Abakus (1980). Similar arguments are used in the proofs of Hurkens and Schlag (2003) and inspired our strategy of proof.

Let M be the $2|\Theta| \times 2|\Theta|$ payoff matrix when we restrict the set of pure strategies to the pure best responses to σ. Let the first $|\Theta|$ pure strategies be those in which "H" is played against an opponent with the same label, and strategies from $|\Theta|+1$ to $2|\Theta|$ be those in which "D" is played against an opponent with the same label. On the diagonal, this matrix M has then first $|\Theta|$ times the entry c and then $|\Theta|$ times the entry d. Off the diagonal, it has half of the entries a and half of the entries b, such that $M_{ij}+M_{ji}=a+b$ for all $i \neq j$. Hence, for any $\mathbf{y} \in \mathbb{R}^{2|\Theta|}$ with $\mathbf{y} \neq 0$, $\sum_{i=1}^{2|\Theta|} y_i = 0$ we have

$$\begin{aligned}\mathbf{y}^T M \mathbf{y} &= c\sum_{i=1}^{|\Theta|} y_i^2 + d \sum_{i=|\Theta|+1}^{2|\Theta|} y_i^2 + \frac{a+b}{2}\sum_{i=1}^{2|\Theta|}\sum_{j=1, j\neq i}^{2|\Theta|} y_i y_j \\ &= \left(c - \frac{a+b}{2}\right)\sum_{i=1}^{|\Theta|} y_i^2 + \left(d - \frac{a+b}{2}\right)\sum_{i=|\Theta|+1}^{2|\Theta|} y_i^2, \qquad (12)\end{aligned}$$

where we used $\sum_i \sum_{j\neq i} y_i y_j = \sum_i y_i \left(\sum_{j\neq i} y_j\right) = \sum_i y_i(-y_i) = -\sum_i y_i^2$ to obtain the last line. The first term is always negative because $a > b > c$. To show that conditions (a) to (e) of Lemma 2 imply ESS, note that the second term is negative for $\frac{a+b}{2} > d$, which then implies $\mathbf{y}^T M \mathbf{y} < 0$ for any $\mathbf{y} \in \mathbb{R}^{2|\Theta|}$ with $\mathbf{y} \neq 0$, $\sum_{i=1}^{2|\Theta|} y_i = 0$. This implies that the payoff matrix is negative definite with respect to its carrier and, together with the fact that (σ, σ) is quasi-strict, it implies that σ is an ESS.

To show that conditions (a) to (d) and (e') of Lemma 2 imply NSS note that the second term is non-positive for $\frac{a+b}{2} \geq d$, which then implies $\mathbf{y}^T M \mathbf{y} \leq 0$ for any $\mathbf{y} \in \mathbb{R}^{2|\Theta|}$ with $\sum_{i=1}^{2|\Theta|} y_i = 0$. Together with the fact that (σ, σ) is quasi-strict, Lemma 6 implies that σ is an NSS.

Condition (e') of Lemma 2 is also necessary to have an NSS: If, in contrast, $\frac{a+b}{2} < d$ then, for $|\Theta| \geq 2$, we can choose a vector $\mathbf{y} \in \mathbb{R}^{2|\Theta|}$ that has zeros in the first $|\Theta|$ entries and some non-zero entries in the remaining entries. Then $\mathbf{y}^T M \mathbf{y} > 0$ and the corresponding σ cannot be an NSS.

Finally, we prove that $\frac{a+b}{2} > d$ is also a necessary condition for ESS. If, in contrast, $\frac{a+b}{2} \leq d$, we can then choose a vector $\mathbf{y} \in \mathbb{R}^{2|\Theta|}$ that has zeros in the first $|\Theta|$ entries and some non-zero entries in the remaining entries. Then $\mathbf{y}^T M \mathbf{y} \geq 0$ and the corresponding σ cannot be an ESS.

The following proof of Lemma 6 then finalizes the proof of Lemma 2.

Proof of Lemma 6: Part (b): Now, we show for quasi-strict (σ, σ) that σ NSS implies Condition 11. Let σ be the K-dimensional restriction of σ to the pure strategies in its support. Quasi-strictness of (σ, σ) implies for all $\mu \in \Delta(supp(\sigma))$: $\mu^T M \sigma = \sigma^T M \sigma$. Furthermore, since the FOC holds with equality, the SOC for NSS implies for all $\mu \in \Delta(supp(\sigma))$:

$$\begin{aligned} & \mu^T M \mu && \leq \sigma^T M \mu \\ \Leftrightarrow\ & \mu^T M \mu - \mu^T M \sigma + \sigma^T M \sigma && \leq \sigma^T M \mu \\ \Leftrightarrow\ & (\mu - \sigma)^T M (\mu - \sigma) && \leq 0. \end{aligned} \tag{13}$$

Now suppose, with the aim to construct a contradiction, that $\exists \mathbf{y} \in \mathbb{R}^K$ with $\sum_i y_i = 0$ such that

$$\mathbf{y}^T M \mathbf{y} > 0. \tag{14}$$

Then we can construct a $\mu \in \Delta(supp(\sigma))$ that violates Inequality 13 in the following way: First, define $\epsilon \equiv min\{(\min_i \sigma_i), (\min_i(1 - \sigma_i))\}$, and $y_{max} \equiv \max_i |y_i|$ and then set $\tilde{y}_i \equiv \frac{\epsilon}{y_{max}} y_i$. Then $\sum_i \tilde{y}_i = 0$ and $\tilde{\mathbf{y}}^T M \tilde{\mathbf{y}} > 0$. If we set $\mu_i \equiv \tilde{y}_i + \sigma_i$ then $\mu_i \in [0, 1]$ and $\sum_i \mu_i = 1$, hence, $\mu \in \Delta(supp(\sigma))$, and, furthermore, $(\mu - \sigma)^T M (\mu - \sigma) > 0$, which contradicts Inequality 13.

Now we show that Condition 13 implies for any σ that forms a quasi-strict Nash equilibrium against itself, that σ is NSS. First, consider the case of a mutant strategy $\mu \in \Delta(supp(\sigma))$. For any such μ Condition 13 implies with $y \equiv \mu - \sigma$

$$\begin{aligned} & (\mu - \sigma)^T M (\mu - \sigma) && \leq 0, \\ \Rightarrow\ & \mu^T M \mu - \mu^T M \sigma + \sigma^T M \sigma - \sigma^T M \sigma && \leq 0, \\ \Rightarrow\ & \mu^T M \mu && \leq \sigma^T M \sigma, \end{aligned} \tag{15}$$

where we used that $\mu^T M \sigma = \sigma^T M \sigma$ for $\mu \in \Delta(supp(\sigma))$ if (σ, σ) is a quasi-strict Nash equilibrium. Hence, for $\mu \in \Delta(supp(\sigma))$ the FOC for NSS is satisfied with equality and the SOC is satisfied by Inequality 15. In order to complete the proof that σ is an NSS, note that for any mutant strategy $\mu \notin \Delta(supp(\sigma))$ it follows from the assumption that (σ, σ) is a quasi-strict Nash equilibrium, that $\mu^T M \sigma < \sigma^T M \sigma$. Hence, the FOC for NSS is strictly satisfied and the SOC, therefore, is irrelevant. This completes the proof of part (b) of Lemma 6. The proof of part (a) is very similar and can be found in van Damme (1991), Theorem 9.2.7, and the preceding text on pages 220/221. QED

B.3 Proof of Lemma 3

We first prove part (a). Suppose the statement is not true. Then a mutant strategy $\mu|\Theta_S$ exists in the restricted meta-game such either the FOC or the SOC of NSS are violated in the restricted meta-game. The same strategy extended to the full meta-game with full set of labels Θ must violate the same NSS condition in the meta-game, since all extra labels are played with probability 0 and do not change the expected payoffs. Hence, any NSS of the meta-game, must also form an NSS in the meta-game restricted to labels in the support of its strategy, which proves part (a) of Lemma 3. By Lemma 2, unless $\frac{a+b}{2} = d$, it is also an ESS of that restricted game.

We now turn to proving part (b). A strategy that supports the NSS in the meta-game with the larger set of labels Θ is a straightforward extension of the full-label-support NSS strategy from the meta-game with a smaller set of labels Θ_S by specifying that $x_\theta(\theta')$ corresponds to the strategy of the base game that gives the opponent only his or her minmax value: $\min\{u^*, b\}$. Note that the expected payoff (call it v) of any strategy in the support of an NSS with $|\Theta_S| \geq 2$ is strictly above this minmax value: For $u^* \neq b$ and any $\theta \in \Theta_S$: $v \geq \sigma(\theta)u^* + (1-\sigma(\theta))b > \min\{u^*, b\}$. For $u^* = b$, no label in the support of an NSS strategy can always play D against all other labels in the support (otherwise it is dominated by any of the other strategies in the support, as these sometimes obtain $a > b$ and never fall below b). Hence, a label θ' with $\sigma(\theta') > 0$ exists such that $v \geq \sigma(\theta')a + (1-\sigma(\theta'))b > b = \min\{u^*, b\}$. This proves part (b) of Lemma 3. QED

B.4 Proof of Lemma 4

The proof follows immediately from Lemma 2.

B.5 Proof of Proposition 1

The following lemma, in conjunction with Lemma 4, immediately proves part (a) of Proposition 1.

Lemma 7. *Let $n \equiv |\Theta| \geq 2$. A full support Nash equilibrium (σ^*, σ^*) of the label game of the (pre-stable) hierarchical label structure exists if and only if the base game is an anti-coordination game ($u^* < b$). If the labels $\theta_1, \theta_2, \ldots, \theta_n$ are ordered according to the hierarchical structure (with θ_1 top label), then for $i \in \{2, \ldots, n\}$:*

$$\begin{aligned}
\sigma^*(\theta_i) &= \sigma^*(\theta_{i-1})\left(\frac{b-u^*}{a-u^*}\right) = \sigma^*(\theta_1)\left(\frac{b-u^*}{a-u^*}\right)^{i-1} \\
&= \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)\left(\frac{b-u^*}{a-u^*}\right)^{i-1}, \qquad (16)
\end{aligned}$$

and each label earning the average payoff

$$\begin{aligned}
h_n &\equiv \sigma^*(\theta_1)u^* + \left(1-\sigma^*(\theta_1)\right)a \\
&= \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)u^* + \left(\frac{\left(\frac{b-u^*}{a-u^*}\right)-\left(\frac{b-u^*}{a-u^*}\right)^n}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)a \qquad (17)
\end{aligned}$$

Equivalently,

$$\begin{aligned}
h_n &= \sigma^*(\theta_n)u^* + (1-\sigma^*(\theta_n))\,b \\
&= \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)\left(\frac{b-u^*}{a-u^*}\right)^{n-1} u^* + \\
&\quad + \left(1-\left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)\left(\frac{b-u^*}{a-u^*}\right)^{n-1}\right) b \\
&= \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right)\left(\frac{b-u^*}{a-u^*}\right)^{n-1} u^* + \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)^{n-1}}{1-\left(\frac{b-u^*}{a-u^*}\right)^n}\right) b
\end{aligned}$$

Note that (for anti-coordination games) $h_n < b$ and $\lim_{n\to\infty} h_n = b$.

Proof of Lemma 7: For convenience, let $\Theta = \{1, 2, ..., n\}$ (with label 1 top label) and, for any mixed strategy $\sigma \in \Delta(S)$, let $\alpha_k \equiv \sigma(k)$. Let $\alpha \in \Delta(S)$ denote the full support symmetric Nash equilibrium. Given α, every label k (fixing the hierarchical label structure) must yield the same expected payoff. The payoff to label k is given by $A_k = \sum_{l=1}^{k-1} \alpha_l b + \alpha_k u^* + \sum_{l=k+1}^{n} \alpha_l a$. Equating A_k and A_{k+1} yields $\alpha_k u^* + \alpha_{k+1} a = \alpha_k b + \alpha_{k+1} u^*$. This, in turn yields $\frac{\alpha_{k+1}}{\alpha_k} = \frac{b-u^*}{a-u^*}$, which must be true for all $k \in \{1, ..., n-1\}$. This corresponds to the first equality. This is only possible with full support if $b > u^*$ and, hence,the base game must be an anti-coordination game. For anti-coordination games, this ratio is a number strictly between 0 and 1. The second equality follows by induction and the third equality from the requirement $1 = \sum_{i=1}^{n} \alpha_i = \alpha_1 \sum_{i=1}^{n} \left(\frac{b-u^*}{a-u^*}\right)^{i-1} = \alpha_1 \left(\frac{1-\left(\frac{b-u^*}{a-u^*}\right)^n}{1-\frac{b-u^*}{a-u^*}}\right)$, where the last step follows from the well-known equality $\sum_{i=0}^{N} \delta^i = \frac{1-\delta^{N+1}}{1-\delta}$, which is easily proved by induction over N. This proves Lemma 7. QED

To prove part (b) of Proposition 1, note first that part (b) follows directly from part (a) for anti-coordination games, since every ESS is also NSS. For conflict games, the argument in the proof of Lemma 7 shows that any NSS with a hierarchical label structure must have all weight on the top label. It remains only to be shown that this is indeed an NSS of the meta-game: Any strategy playing any other label with positive probability earns b or less against the incumbent top label population, while incumbents earn $u^* \geq b$. In games with $u^* > b$, the mutant earns strictly less in the FOC. In the knife-edge case of a base game with $u^* = b$ the FOC is satisfied with equality if D is played against the top label with certainty, but then the incumbents earn a against the mutants, while mutants earn strictly less than a against themselves. QED

B.6 Proof of Proposition 2

In an egalitarian label structure, each label plays H against half of all other labels and D against the other half. It is easy to see that, for odd $|\Theta|$ (then we can find a natural number l such that $|\Theta| = 2l + 1$), such egalitarian pre-stable structures exist, i.e. it is a well-defined structure. We can, for instance, locate the $2l + 1$ labels on a circle, and each label plays H against the next l labels located clockwise and D against the next l labels located anti-clockwise.

For convenience, let $\Theta = \{1, 2, ..., n\}$. Let $\alpha \in \Delta(\Theta)$ denote the full support symmetric Nash equilibrium of the label game given by $\alpha_k = \frac{1}{n}$ for all $k \in \{1, ..., n\}$.

The expected expected payoff of each strategy in the label game against α is given by:

$$v_n \equiv \frac{u^*}{n} + \frac{n-1}{n}\frac{a+b}{2}. \tag{18}$$

It follows immediately from Lemma 4 that for $d < \frac{a+b}{2}$ the corresponding strategy of the meta-game forms an ESS, and that for $d > \frac{a+b}{2}$ it cannot form an NSS. QED

B.7 Proof of Proposition 3

If $|\Theta| \geq 4$ is an even number, we can find a natural number $l \geq 2$ such that $|\Theta| = 2l$. For anti-coordination games, we now construct an approximately egalitarian pre-stable label structure with a full support Nash equilibrium in the label game. (For their task allocation game, Hurkens and Schlag (2003) have a similar construction in the proof of their Prop. 3). Imagine that the $2l$ labels are placed on a circle. Labels $i \in \{1, \ldots, l\}$ play H against the l next labels located clockwise and D against the $l - 1$ labels located anti-clockwise. Label-types $i \in \{l + 1, \ldots, 2l\}$ play H against the $l - 1$ next labels located clockwise and D against the l labels located anti-clockwise. This forms an pre-stable structure if all labels play also x^* against their own label.

Consider now the corresponding label game. This has a full support Nash equilibrium if and only if a full support mixed strategy $\alpha = (\alpha_1, \ldots \alpha_{2l}) \in \Delta(\Theta)$ exists in the label game such that all labels earn the same expected payoff. Hence, the difference between the payoff of any label θ_i and the payoff of the clockwise next label $\theta_{i+1(mod\ 2l)}$ must be zero:
For $1 \leq i < l$:

$$\alpha_i \left(u^* - b\right) + \alpha_{i+1} \left(a - u^*\right) + \alpha_{i+l+1} \left(b - a\right) = 0, \tag{19}$$

for $i = l$:

$$\alpha_l (u^* - b) + \alpha_{l+1} (a - u^*) = 0, \tag{20}$$

for $l + 1 \leq i \leq 2l - 1$:

$$\alpha_i (u^* - b) + \alpha_{i+1} (a - u^*) + \alpha_{i-l} (b - a) = 0. \tag{21}$$

Note first that for conflict games ($u^* \geq b$) the equation

$$\alpha_l (u^* - b) + \alpha_{l+1} (a - u^*) = 0$$

has no solution (with $\alpha_l, \alpha_{l+1} \geq 0$). For $|\Theta| = 4$, it is straightforward to show that all approximately egalitarian structures have the structure above and, thus, no approximately egalitarian structure can be part of an ESS of the meta-game in this case.

Now we prove that an ESS of the meta-game under the approximately egalitarian label structure exists when the base game is an anti-coordination game. We proceed by first establishing a lemma that provides a necessary condition for an arbitrary finite, symmetric, two-player game to have a symmetric completely mixed Nash equilibrium.

A few definitions are necessary. For a symmetric finite two player game with $n \times n$ payoff matrix G, let $D = D(G)$ denote the G-induced payoff difference matrix that is given by the $n \times n - 1$ matrix obtained from G as follows. The l-th row of D is the difference between rows l and $l + 1$, for $l = 1, 2, ..., n - 1$. Finally, denote by $\bar{D} = \bar{D}(G)$ the $n \times n$ matrix that coincides with D for the first $n - 1$ rows and has the unit vector (vector of all ones) in row n. Let $h \in \mathbb{R}^n$ denote the vector that is equal to the zero vector except that $h_n = 1$.

A vector $x \in \mathbb{R}^n$ represents a completely mixed Nash equilibrium of the finite, symmetric, two-player game with $n \times n$ payoff matrix G if and only if the following two conditions hold:[23]

(I) **Equal Payoff Condition**
$x \geq 0$ (that is $x_i \geq 0 \; \forall \; 0 \leq i \leq n$ and $\exists \; i$ such that $x_i > 0$) and $\bar{D}x = h$.

(II): **Full Support Condition**
$x_i > 0$ for $1 \leq i \leq n$.

Consider the label game with an approximately egalitarian structure as described above with an even number of labels $n = 2l$, for any $l = 1, 2, ...$ The payoff difference matrix D induced by this game is as follows. Column 1 has two non-zero entries, the first in row 1 given by $u^* - b$, and the second

[23] This characterization and the characterization of the Equal Payoff Condition (I) below are taken from our note Herold and Kuzmics (2017).

in row $l+1$ given by $b-a$. Column i with $2 \leq i \leq l-1$ has three non-zero entries at row $i-1$ given by $a-u^*$, at row i given by u^*-b, and at row $l+i$ given by $b-a$. Column l has two non-zero entries at row $l-1$ given by $a-u^*$ and at row l given by u^*-b. Column $l+1$ has two non-zero entries at row l given by $a-u^*$ and at row $l+1$ given by u^*-b. Column i with $l+2 \leq i \leq 2l-1$ has three non-zero entries at row $i-(l+1)$ given by $b-a$, at row $i-1$ given by $a-u^*$, and at row i given by u^*-b. Finally, column $2l$ has two-non-zero entries, one at row $l-1$ given by $b-a$ and one at row $2l-1$ given by $a-u^*$.

We will now use a result from our note Herold and Kuzmics (2017) (compare with Lemma 2 and the sentence directly after Lemma 2): The Equal Payoff Condition (I), has a solution if and only if

$$\nexists\ w \in \mathbb{R}^{n-1} \textit{ such that } w^T D > 0. \tag{22}$$

Let d_i denote the i-th column of this matrix D. To establish the Equal Payoff Condition (I) we need to show that there is no vector $v \in \mathbb{R}^{n-1}$ such that $v^T d_i > 0$ for all i. The proof is by contradiction. Thus, suppose there is such a $v \in \mathbb{R}^{n-1}$ with $v^T d_i > 0$ for all i. Let d^* denote the sum of all columns 1 to n. Then d^* has only one non-zero coordinate, which is at row l and is given by $a-b$. As $v^T d_i > 0$ for all i we have that $v^T d^* > 0$ and, as $a-b>0$, we have $v_l > 0$.

By $v^T d_l > 0$, we then obtain that $(a-u^*)v_{l-1} + (u^*-b)v_l > 0$. Given that $v_l > 0$ and $a-u^* > 0$ and $u^*-b < 0$, we have $v_{l-1} > 0$. By $v^T d_{2l} > 0$, we obtain that $(b-a)v_{l-1} + (a-u^*)v_{2l-1} > 0$, which, given the results so far, implies that $v_{2l-1} > 0$. Next, consider $v^T d_{l-1} > 0$. This implies that $(a-u^*)v_{l-2} + (u^*-b)v_{l-1} + (b-a)v_{2l-1} > 0$. Given the results so far, this implies that $v_{l-2} > 0$. Going through all columns of D in this way, except column 1, we obtain the result that $v_i > 0$ for all i. But then we obtain $v^T d_1 < 0$ which provides a contradiction to our supposition. We thus have established the Equal Payoff Condition (I).

Next, we need to show that this mixed strategy which satisfies the Equal Payoff Condition (I) must be completely mixed, i.e. satisfy the Full Support Condition (II). Suppose it does not. Suppose $x \geq 0$ and there is a coordinate i such that $x_i = 0$ and, nevertheless, $Dx = 0$. Note that each row i of D has exactly one strictly positive entry $d_{i(i+1)} = (a-u^*)$ at column position $i+1$. Note that the other non-zero entries are negative: $(b-a) < 0$ and $d_{ii} = (u^*-b) < 0$ for anti-coordination games.

Suppose $x_1 = 0$. If we add together all rows of D we obtain $r^* \equiv ((u^*-a), 0, \ldots, 0, (a-b), (a-b), 0, \ldots, 0, (b-u^*))$. We must have $r^* x = 0$ (which corresponds to the payoff difference between the last and first label). In particular, $x_{2l} = 0$ (otherwise $r^* x$ would be positive if $x_1 = 0$). But if any $x_{i+1} = 0$ then from row i we see that $x_i = 0$ and thus all $x_i = 0$, which contradicts $\sum_i x_i = 1$.

Thus, we must have $x_1 > 0$. Still, if $x_i > 0$ for any $1 \leq i \leq 2l - 1$. Then also $x_{i+1} > 0$ (otherwise row i would stay strictly negative. Thus, $x_i > 0$ for all $1 \leq i \leq 2l$ (by induction), and we are done.[24] By Lemma 2 this is an ESS. QED

B.8 Proof of Lemma 5

The average payoff in any label game induced by a pre-stable structure is given by

$$\begin{aligned}\sum_{\theta,\theta'\in\Theta} \sigma(\theta)T_{\theta,\theta'}\sigma(\theta') &= u^* \sum_{\theta\in\Theta} (\sigma(\theta))^2 + \frac{a+b}{2} \sum_{\theta\neq\theta'} \sigma(\theta)\sigma(\theta') \\ &= u^* \left(\sum_{\theta\in\Theta} (\sigma(\theta))^2 \right) + \frac{a+b}{2} \left(1 - \left(\sum_{\theta\in\Theta} (\sigma(\theta))^2 \right) \right). \qquad (23)\end{aligned}$$

Note that $\left(\sum_{\theta\in\Theta} (\sigma(\theta))^2\right) \in [\frac{1}{|\Theta|^2}, 1]$, under the constraint $\sum_{\theta\in\Theta} \sigma(\theta) = 1$, is minimized by σ with $\sigma(\theta) = \frac{1}{|\Theta|}$ for all $\theta \in \Theta$ and is maximized by a σ with $\sigma(\theta_T) = 1$ for one label $\theta_T \in \Theta$ and with $\sigma(\theta) = 0$ for all remaining labels $\theta \neq \theta_T$. Thus, the average payoff is a weighted average of u^* and $\frac{a+b}{2}$ and is maximized by putting as much weight as possible on the higher number of the two. QED

B.9 Proof of Proposition 4

We know from Lemma 2 that for $d > \frac{a+b}{2}$ (i.e. for cases (a) and (b)) no ESS and no NSS with full label-support can exist for $|\Theta| \geq 2$. Now if any NSS with more than two labels in its support would exist, then, by Lemma 3, it would also be an NSS in the game restricted to the set of labels in the support Θ_S. But in this restricted meta-game, it would be a full support equilibrium, a contradiction.

Parts (a) and (b) follow directly from this argument.

(c) Follows directly from Proposition 1, Proposition 2, and Lemma 5.

(d) Follows directly from Proposition 1, Proposition 2, Proposition 3, and Lemma 5.

QED

[24] Note that equilibria do exist with no full label-support, e.g. $x_{l+1} = 0$ and equal weight on all other x_i, $i \neq l+1$ corresponds to the egalitarian equilibrium with $2l-1$ labels. But these equilibria without full label-support do not satisfy the Equal Payoff Condition (I) and, thus, one with full label-support must satisfy the Equal Payoff Condition (I).

C The meta-game with four labels

Next consider the case $|\Theta| = 4$. With four labels, the hierarchical structure emerges again. Since the number of labels is even, there is no egalitarian structure, but there is a circular structure that is approximately egalitarian.

$$\begin{array}{c|cccc} & L_1 & L_2 & L_3 & L_4 \\ \hline L_1 & u^* & a & a & a \\ L_2 & b & u^* & a & a \\ L_3 & b & b & u^* & a \\ L_4 & b & b & b & u^* \end{array} \qquad \begin{array}{c|cccc} & L_1 & L_2 & L_3 & L_4 \\ \hline L_1 & u^* & a & a & b \\ L_2 & b & u^* & a & a \\ L_3 & b & b & u^* & a \\ L_4 & a & b & b & u^* \end{array}$$

hierarchical approximately egalitarian

It will turn out to be useful to consider a further reduced form label game in which some labels are summarized in sub-groups, which are treated equally by all other labels. For instance:

$$\begin{array}{c|ccc} & G_{12} & L_3 & L_4 \\ \hline G_{12} & h_2 & a & a \\ L_3 & b & u^* & a \\ L_4 & b & b & u^* \end{array} \qquad \begin{array}{c|ccc} & L_1 & G_{23} & L_4 \\ \hline L_1 & u^* & a & b \\ G_{23} & b & h_2 & a \\ L_4 & a & b & u^* \end{array}$$

hierarchical approximately egalitarian

For both label structures, full-label-support equilibria exist only for anti-coordination games, but not for conflict games.

Furthermore, for $|\Theta| \geq 4$, structures with a partial hierarchy among some intra-egalitarian groups exist. Consider the case $|\Theta| = 4$:

$$\begin{array}{c|cccc} & L_1 & L_2 & L_3 & L_4 \\ \hline L_1 & u^* & a & a & a \\ L_2 & b & u^* & a & b \\ L_3 & b & b & u^* & a \\ L_4 & b & a & b & u^* \end{array} \qquad \begin{array}{c|cccc} & L_1 & L_2 & L_3 & L_4 \\ \hline L_1 & u^* & a & b & a \\ L_2 & b & u^* & a & a \\ L_3 & a & b & u^* & a \\ L_4 & b & b & b & u^* \end{array}$$

Top label and egalitarian group Egalitarian group - bottom label

Considering the labels L_2-L_4 in the first label game as one egalitarian sub-group G_B, and the labels L_1-L_3 in the second label game as one sub-group G_T, a further reduction of the label structures is given by

$$\begin{array}{c|cc} & T & G_B \\ \hline T & u^* & a \\ G_B & b & v_3 \end{array} \qquad \begin{array}{c|cc} & G_T & B \\ \hline G_T & v_3 & a \\ B & b & u^* \end{array}$$

Top label and egalitarian group Egalitarian group - bottom label

Analogous pre-stable label structures with a hierarchy between a single label and an egalitarian group of $k \equiv |\Theta| - 1$ labels exist, of course, for any even number of labels $|\Theta|$ and lead to the correspondingly further reduced label structure:

$$\begin{array}{c|cc} & T & G_B \\ \hline T & u^* & a \\ G_B & b & v_k \end{array} \qquad\qquad \begin{array}{c|cc} & G_T & B \\ \hline G_T & v_k & a \\ B & b & u^* \end{array}$$

Top label and egalitarian-group   Egalitarian-group - bottom label

Remember that $v_k \in [u^*, \frac{a+b}{2}]$ or $[\frac{a+b}{2}, u^*]$. In conflict games with $u^* \geq b$, it follows for all $k \geq 3$ that $v_k > b$. Hence, the top-label (or the labels of the top egalitarian group in the second reduced label-game) dominates the labels of the bottom egalitarian group (or the bottom label, respectively). In equilibrium all probability weight must, therefore, be on the top-label or, respectively, on the labels of the top egalitarian group. In anti-coordination games with $u^* < b$, in the first game there is a full label-support equilibrium with a top label and an egalitarian group at the bottom of the hierarchy. In the second game it depends: For sufficiently small k, the expected payoff within the egalitarian group v_k is still smaller than b, and there is a full-label-support equilibrium, yet there must be a $\bar{k}$ such that, for all $k \geq \bar{k}$ the payoff $v_k \geq b$, the payoff of the top-group dominates the payoff of the bottom label payoff and the bottom label cannot be played in equilibrium.

D Group sub-structures

These arguments can be generalized for more hierarchies among groups with different sub-structures.

Definition 4. *Group sub-structures:*

(a) A pre-stable label structure has a **group sub-structure** *if the set of labels Θ can be partitioned into non-empty sets $\Theta_1, ..., \Theta_M$ with $M < |\Theta|$ such that for all $i, j \in \{1, \ldots, M\}$ holds $x_{\theta_i}(\theta_j) = x_{\theta'_i}(\theta'_j)$ for all $\theta_i, \theta'_i \in \Theta_i$ and $\theta_j, \theta'_j \in \Theta_j$.*

(b) A pre-stable structure has a **hierarchy among groups** *if Θ can be partitioned into two nonempty sets Θ_T and Θ_B such that $x_\theta(\theta') = 1$ for all $\theta \in \Theta_T$ and $\theta' \in \Theta_B$.*

(c) In a pre-stable label structure, a label θ_T is called a **top label** *if $x_{\theta_T}(\theta) = 1$ for all $\theta \in \Theta \setminus \{\theta_T\}$ and a* **top label within subgroup** *Θ_g if $x_{\theta_T}(\theta) = 1$ for all $\theta \in \Theta_g \setminus \{\theta_T\}$.*

(c) *In a pre-stable label structure, a label θ_B is called a* **bottom label** *if $x_{\theta_B}(\theta) = 0$ for all $\theta \in \Theta \setminus \{\theta_B\}$ and a* **bottom label within subgroup** Θ_g *if $x_{\theta_B}(\theta) = 0$ for all $\theta \in \Theta_g \setminus \{\theta_B\}$.*

Consider the label game of a pre-stable structure with a group sub-structure. A full-support strategy in this label game can only be an equilibrium if the payoffs are equilibrated within each group. More precisely, for any subset of labels $\Theta_j \subset \Theta$, let for all $\theta \in \Theta_j$

$$\sigma|_{\Theta_j}(\theta) = \frac{\sigma(\theta)}{\sum_{\theta' \in \Theta_j} \sigma(\theta')}. \tag{24}$$

Definition 5. *Consider the label game of a pre-stable structure with a set of labels Θ and with a group sub-structure $(\Theta_1, \ldots, \Theta_M)$.*

(a) *Let for each Θ_j the* **sub-group label game** $\mathfrak{G}_{\Theta_j}$ *denote the $|\Theta_j| \times |\Theta_j|$ game derived from the full label game by eliminating all rows and all columns for labels $\theta_k \notin \Theta_j$.*

(b) *A full support strategy σ of the label game with full set of labels Θ is called* **equilibrated within group** Θ_j *if under $\sigma|_{\Theta_j}$ every label $\theta \in \Theta_j$ obtains exactly the same expected payoff w_j in the sub-group label game $\mathfrak{G}_{\Theta_j}$.*

(c) *A full support strategy of the label game with full set of labels $\Theta = \biguplus_{j=1}^{M} \Theta_j$ is called* **within sub-group equilibrated** *if, for every Θ_j with $j \in \{1, \ldots, M\}$, it is equilibrated within group Θ_j.*

For a strategy of the meta-game σ that induces a pre-stable label structure that is within sub-group equilibrated, we can now introduce a further reduced **inter-group label game** that has one pure strategy ϑ_j, $j \in \{1, \ldots M\}$ for every sub-group and a $M \times M$ payoff matrix with payoff w_j on the diagonal and payoffs a and b as induced by the original label game. Also note that a strategy σ of the original meta-game induces a strategy $\hat{\sigma}$ in the inter-group label game via

$$\hat{\sigma}(\vartheta_j) = \sum_{\theta' \in \Theta_j} \sigma(\theta'). \tag{25}$$

Definition 6. *Consider a full label-support strategy of the meta-game with a set of labels Θ inducing a label game of a pre-stable structure with a group sub-structure $(\Theta_1, \ldots, \Theta_M)$ that is within sub-group equilibrated. The induced full support strategy σ of the label game with full set of labels Θ is called* **inter-group equilibrated** *if every induced strategy in the inter-group label game earns the same expected payoff.*

Definition 7. *A full support strategy σ of the label game induced by a pre-stable structure is called* **equilibrated** *if every pure strategy in the label game earns the same expected payoff under σ.*

Lemma 8. *Consider a pre-stable label structure with a group sub-structure $\Theta_1, ..., \Theta_M$. A full-label-support strategy σ of the label game with a set of labels $\Theta = \biguplus_{j=1}^{M} \Theta_j$ is equilibrated if and only if*

- *it is within sub-group equilibrated, and*
- *it is inter-group equilibrated.*

Proof: Note that the expected payoff of any label θ_i, $i \in \{1, \ldots, |\Theta|\}$, in some group of labels Θ_j, $j \in \{1, \ldots, M\}$, can be decomposed in the probability of playing against a label in its own group Θ_j times the conditional expected payoff w_j in that case, and the probability of playing against any label not in the group and the conditional expectation in that case.

Consider a full support strategy σ of the label game induced by a pre-stable structure.

Proof of "only if" statement: Suppose there is a group Θ_j which is not within subgroup equilibrated. Then there are at least two labels which earn a different expected payoff conditional on playing in that group. But since all labels outside the group play identically against both labels, this implies that they also earn different expected payoffs overall and, thus, the full support strategy of the full label game cannot be equilibrated. This means that being within sub-group equilibrated is a necessary condition for σ to be equilibrated. Next we show that σ can only be equilibrated if it is inter-group equilibrated. Suppose it is not. Then pick two labels from different groups Θ_i and Θ_j. Then both labels earn different payoffs, contradicting the supposition that σ is equilibrated.

Proof of the "if" statement: Suppose the full-label-support strategy σ is within sub-group equilibrated and inter-group equilibrated. Then, due to within-subgroup equilibration, every label $\theta_i in \Theta_j$ earns the same expected payoff as Θ_j in the inter-group label game. Furthermore, all Θ_j, $j \in \{1, \ldots, M\}$ earn the same expected payoff (since σ is inter-group equilibrated), all labels earn the same expected payoff and σ is equilibrated. QED

Note that a full support strategy σ of the label game is equilibrated if and only if it is a Nash equilibrium of the label game. Lemma 8, in conjunction with Lemma 4, therefore, gives us a clear picture when a full-label-support NSS or ESS exists for a pre-stable structure with sub-group structure.

The following Proposition 7 helps us to exclude some pre-stable label structures if we are searching for full-label-support NSS in situations with conflict base-games.

Proposition 7. *Consider any conflict base-game. For the corresponding meta-game with $|\Theta| \geq 2$ in any full-label-support NSS (and thus in any ESS), there*

(a) can **not** *exist a top label,*

(b) can **not** *exist a bottom label,*

(c) can **not** *exist a hierarchy among groups,*

(d) can **not** *exist a group sub-structure in which one group with more than one label has has a top player (within that group),*

(e) can **not** *exist a group sub-structure in which one group with more than one label has has a bottom player(within that group).*

Proof:

(a) Since $u^* \geq b$ for conflict games, the top label (who plays hawk and earns the largest possible payoff a against all other labels) is a (at least weakly) dominant strategy in the label game and would earn strictly more than any other strategy of the label game under full label-support.

(b) Since $u^* \geq b$ for conflict games, a bottom label (who plays dove against all other labels) is weakly dominated by all other strategies and cannot be part of any full support equilibrium of the label game.

(c) The same argument as (a) now applies to the top group in the inter-group label game.

(d) The same argument as in (a) now applies to the sub-group label game $\mathfrak{G}_{\Theta_j}$ of such a sub-group Θ_j with a top label.

(e) The same argument as in (b) now applies to the sub-group label game $\mathfrak{G}_{\Theta_j}$ of such a sub-group Θ_j with a bottom label.

QED

D.1 There is no ESS when $|\Theta| = 4$ in the conflict case

Suppose the base game is one of conflict (i.e. $d > b$) and $|\Theta| = 4$. We now show that this meta-game has no ESS. By Lemma 2, any ESS must have full support on all four labels. We then show that any candidate ESS that satisfies properties a and b of Lemma 2 necessarily has a dominated label, and, thus, cannot have full support, violating property (c) of Lemma 2.

We have to go through a series of cases. First, suppose that one label plays H against all other labels. This label dominates all other labels, and we arrive at a contradiction. Second, suppose that one label plays D against all other labels. This label is dominated by all other labels, and again we reach a contradiction. The only case remaining is such that all labels play H against at least one other label and at most two other labels. This leads to a unique label structure (subject to relabeling), the unique, approximately egalitarian structure (subject to relabeling) given by the following matrix.

$$\begin{array}{c|cccc} & L_1 & L_2 & L_3 & L_4 \\ \hline L_1 & u^* & a & a & b \\ L_2 & b & u^* & a & a \\ L_3 & b & b & u^* & a \\ L_4 & a & b & b & u^* \end{array}$$

Given $d > b$ and thus $u^* > b$, label L_3 is dominated by label L_2.

D.2 There are only two ESS when $|\Theta| = 5$ in the conflict case

Suppose the base game is one of conflict with $\frac{a+b}{2} > d > b$ and $|\Theta| = 5$. Then $u^* > b$. This game has exactly two ESS (modulo relabeling). One is the egalitarian one. Only one other pre-stable structure can exist that forms an ESS: In any pre-stable label structure, overall half of the off-diagonal entries are a and the other half are b. In the egalitarian pre-stable structure, each label has exactly two a entries and two b entries. In any non-egalitarian label structure, at least one label, denote it L_1, must have three a entries, but it cannot have four because it would dominate all other labels. Denote by L_2 the unique label against which L_1 has an entry of b. Now note that the remaining three labels $L_3 - L_5$ must all obtain a if matched against label L_2, otherwise they would be dominated by L_1. Thus, $L_3 - L_5$ form a sub-group of labels, and we know from the results for $|\Theta| = 3$ that it can only have the egalitarian structure. Thus, any non-egalitarian ESS must have the following pre-stable structure (modulo relabeling)

$$\begin{array}{c|ccccc} & L_1 & L_2 & L_3 & L_4 & L_5 \\ \hline L_1 & u^* & b & a & a & a \\ L_2 & a & u^* & b & b & b \\ L_3 & b & a & u^* & a & b \\ L_4 & b & a & b & u^* & a \\ L_5 & b & a & a & b & u^* \end{array}$$

with the labels L_3, L_4, and L_5 receiving equal probability weight and the other two also receiving positive probability weight. To verify that such an ESS exists, we have to show that a full-label-support distribution exists that satisfies the equal payoff condition. Note that labels $L_3 - L_5$ are treated equally by L_1 and L_2 and, thus, form an egalitarian sub-group, call it G_3. Each label within this subgroup does equally well, and the conditional expected payoff when two labels of G_3 meet is $v_3 = \frac{u^*+a+b}{3}$, where $a > v_3 > u^* > b$. The resulting symmetric two-player label-game must have a symmetric Nash equilibrium (see, e.g. (Weibull 1995), p.27). It is easy to see that this symmetric Nash equilibrium must have full (label) support: First, no label is best responding to itself. Second, if one label would have zero probability weight, one of the remaining labels always dominates the other. Thus, we have a full-label-support equilibrium which by Lemma 2 is an ESS.

D.3 On ESS when $|\Theta| = 6$ in the conflict case

Suppose the base game is one of conflict with $d > b$ (hence, $a > u^* > b$) and $|\Theta| = 6$. We now show that this meta-game has no ESS for some parameters and only an approximately egalitarian ESS for other parameters. By Lemma 2, any ESS must have full support on all six labels. We then show for some parameters that any candidate ESS that satisfies properties a and b of Lemma 2 either has a dominated label or has label-game equilibrium that does not have full support. In either case, it then follows that the label game cannot have a full support ESS. This is immediately obvious in the dominated label case and true in the other case because a full support ESS is necessarily the unique Nash equilibrium of a game (see, e.g. Weibull (1995, Proposition 2.2)).

There are a series of cases to consider. First, consider the case that one label plays H against all other labels. Then this label dominates all other labels, and we arrive at a contradiction. Second, suppose one label plays D against all other labels. Then this label is dominated by all other labels, and again we arrive at a contradiction.

Third, consider the case that one label plays H against all but one other label. Then, if we want to avoid having dominated labels, the label game must have the following sub-structure (otherwise L_1 would dominate at least one of the labels L_3 to L_6):

	L_1	L_2	L_3	L_4	L_5	L_6
L_1	u^*	b	a	a	a	a
L_2	a	u^*	b	b	b	b
L_3	b	a	u^*			
L_4	b	a		u^*		
L_5	b	a			u^*	
L_6	b	a				u^*

Note that the four labels L_3 to L_6 are all treated equally by labels L_1 and L_2. They can only differ in how they play against each other. The problem, thus, is reduced to considering only these four labels and, by the argument above (case $|\Theta| = 4$), no label structure with four labels exists in which there is no dominated label in conflict games.

Fourth, a similar argument can be made when we consider the case that one label plays D against all but one other label. This also leads to the existence of a dominated label in much the same way as in the previous case.

In all remaining cases, every label plays H against at least two and at most three opponents. Given the fact that the total number of H plays in the matrix must be 15, we need exactly three labels to play H against two opponent labels and exactly three labels to play H against three opponent labels. Let us call the first group the $2H$-group and the latter the $3H$-group. There are now, without loss of generality, four cases. Each group (of three labels each) can only be either egalitarian or hierarchical among themselves . Each case leads to a different label structure, all of which are approximately egalitarian.

Case1: If both groups are hierarchical among themselves, this leads to a structure in which the lowest label in the $3H$-group internal hierarchy (call it L_3), dominates the label (call it L_4) that is highest in the $2H$-group internal hierarchy.

$$\begin{array}{c|cccccc}
 & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 \\
\hline
L_1 & u^* & a & a & & & \\
L_2 & b & u^* & a & & & \\
L_3 & b & b & u^* & a & a & a \\
L_4 & b & b & b & u^* & a & a \\
L_5 & & & & b & u^* & a \\
L_6 & & & & b & b & u^*
\end{array}$$

Case 2: If the $3H$-group is egalitarian and the $2H$-group is hierarchical then let us denote the medium label of the internal $2H$-group hierarchy by L_5 and let L_1 denote the unique label from the $3H$-group against which L_5 plays H. Then the following label structure is implied:

$$\begin{array}{c|cccccc}
 & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 \\
\hline
L_1 & u^* & a & b & a & b & a \\
L_2 & b & u^* & a & a & a & b \\
L_3 & a & b & u^* & a & a & b \\
L_4 & b & b & b & u^* & a & a \\
L_5 & a & b & b & b & u^* & a \\
L_6 & b & a & a & b & b & u^*
\end{array}$$

If we set $x_4 = 0$ (no weight on label L_4), then the remaining five labels form an egalitarian sub-structure. If we put equal weight on the remaining five labels $x_1 = x_2 = x_3 = x_5 = x_6 = \frac{1}{5}$, they form a Nash equilibrium and obtain equilibrium payoff $\frac{2a+2b+u^*}{5}$. Label L_4 would earn only the smaller payoff $\frac{2a+2b+b}{5}$ in this equilibrium. Hence, this forms a partial support Nash equilibrium of the label game. But this contradicts the existence of a full-support ESS (by Weibull (1995, Proposition 2.2)).

Case 3: If both groups are egalitarian, then let L_1 be an arbitrary label from the $3H$-group and denote by L_2 the label from the $3H$-group against which L_1 plays H and denote by L_3 the label against which L_1 plays D. Then call L_6 the unique label from the $2H$-group against which L_1 plays D. Then the label structure must be one of the following two:

$$\begin{array}{c|cccccc}
 & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 \\ \hline
L_1 & u^* & a & b & a & a & b \\
L_2 & b & u^* & a & \mathbf{b} & \mathbf{a} & a \\
L_3 & a & b & u^* & \mathbf{a} & \mathbf{b} & a \\
L_4 & b & \mathbf{a} & \mathbf{b} & u^* & a & b \\
L_5 & b & \mathbf{b} & \mathbf{a} & b & u^* & a \\
L_6 & a & b & b & a & b & u^*
\end{array}
\qquad
\begin{array}{c|cccccc}
 & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 \\ \hline
L_1 & u^* & a & b & a & a & b \\
L_2 & b & u^* & a & \mathbf{a} & \mathbf{b} & a \\
L_3 & a & b & u^* & \mathbf{b} & \mathbf{a} & a \\
L_4 & b & \mathbf{b} & \mathbf{a} & u^* & a & b \\
L_5 & b & \mathbf{a} & \mathbf{b} & b & u^* & a \\
L_6 & a & b & b & a & b & u^*
\end{array}$$

label structure 1 label structure 2

In label structure 1 label L_1 dominates label L_4 and, thus, cannot form an ESS. In label structure 2 the distribution $x_1 = x_2 = x_3 = \frac{1}{3}$ and $x_4 = x_5 = x_6 = 0$ forms a partial-support Nash equilibrium with equilibrium payoff $\frac{a+b+u^*}{3}$ for L_1, L_2, and L_3. Labels L_4, L_5, and L_6 would earn only the lower payoff of $\frac{a+b+b}{3}$ in the equilibrium. Again this contradicts the existence of a full-support ESS (by Weibull (1995, Proposition 2.2)).

Case 4: If the $3H$-group is hierarchical and the $2H$-group is egalitarian, then let L_1 denote the top label of the internal $3H$-group hierarchy and let L_4 denote the unique label from the $2H$-group against which L_1 plays H. Then the following label structure is implied:

$$\begin{array}{c|cccccc}
 & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 \\ \hline
L_1 & u^* & a & a & a & b & b \\
L_2 & b & u^* & a & b & a & a \\
L_3 & b & b & u^* & a & a & a \\
L_4 & b & a & b & u^* & a & b \\
L_5 & a & b & b & b & u^* & a \\
L_6 & a & b & b & a & b & u^*
\end{array}$$

If we set $x_6 = 0$ (no weight on label L_6), the remaining labels form an induced label structure that is the non-egalitarian Nash equilibrium with five labels described above. If the expected payoff of label L_6 is below the equilibrium payoff of the other five labels, this constitutes a Nash equilibrium, which contradicts the existence of a full-support ESS (by Weibull (1995, Proposition 2.2)).

By using the LinearSolve command in Mathematica and by solving for the equilibrium using the corresponding payoff-difference matrix with five labels we could show that there exists a $v_0 \approx 0.3611$ such that for all $\frac{u^*-b}{a-b} > v_0$ this is indeed the case, and there cannot exist an ESS for $|\Theta| = 6$.

For sufficiently small $\frac{u^*-b}{a-b} > 0$ however, we could show, by using the LinearSolve command in Mathematica, and by solving for the equilibrium using the corresponding payoff-difference matrix with six labels, that a full-support equilibrium exists which is then also an ESS. Indeed, we found such an equilibrium with six labels for all numerical values of $\frac{u^*-b}{a-b} < v_0$ that we checked.

References

Abakus, A. (1980): "Conditions for evolutionary stable strategies," *Journal of Applied Probability*, 17, 559–562.

Balkenborg, D., J. Hofbauer, and C. Kuzmics (2013): "Refined best reply correspondence and dynamics," *Theoretical Economics*, 8(1), 165–192.

Banerjee, A., and J. W. Weibull (2000): "Neutrally stable outcomes in cheap-talk coordination games," *Games and Economic Behavior*, 32, 1–24.

Basu, K., and J. W. Weibull (1991): "Strategy subsets closed under rational behavior," *Economics Letters*, 36, 141–46.

Benndorf, V., I. Martinez-Martinez, and H.-T. Normann (2016): "Equilibrium selection with coupled populations in hawk–dove games: Theory and experiment in continuous time," *Journal of Economic Theory*, 165, 472–486.

Bhaskar, V. (1998): "Noisy communication and the evolution of cooperation," *Journal of Economic Theory*, 82, 110–31.

Blume, A., Y.-G. Kim, and J. Sobel (1993): "Evolutionary stability in games of communication," *Games and Economic Behavior*, 5, 547–575.

Brown, R. (1965): *Social Psychology.* The Free Press, New York.

Dekel, E., J. C. Ely, and O. Yilankaya (2007): "Evolution of preferences," *Review of Economic Studies*, 74, 685–704.

Eshel, I., L. Samuelson, and S. Shaked (1998): "Altruists, egoists, and hooligans in a local interaction model," *American Economic Review*, 88.1, 157–179.

Farrell, J. (1987): "Cheap talk, coordination, and entry," *The RAND Journal of Economics*, 18.1, 34–39.

Forsyth, D. R. (2016): *Group Dynamics, 5th edition.* Wadsworth.

Haigh, J. (1975): "Game theory and evolution," *Advances in Applied Probability*, 7, 8–11.

Heller, Y. (2014): "Stability and trembles in extensive-form games," *Games and Economic Behavior*, 84, 132–136.

——— (2015): "Three steps ahead," *Theoretical Economics*, 10(1), 203–241.

Herold, F., and C. Kuzmics (2009): "Evolutionary stability of discrimination under observability," *Games and Economic Behavior*, 67, 542–551.

——— (2017): "Note: A necessary condition for symmetric completely mixed Nash-equilibria," Mimeo.

Hurkens, S., and K. Schlag (2003): "Evolutionary insights on the willingness to communicate," *International Journal of Game Theory*, 31, 511–526.

Kim, Y.-G., and J. Sobel (1995): "An evolutionary approach to pre-play communication," *Econometrica*, 63, 1181–93.

Maynard Smith, J. (1982): *Evolution and the Theory of Games.* Cambridge University Press, Cambridge.

Maynard Smith, J., and G. R. Price (1973): "The logic of animal conflict," *Nature*, 246, 15–18.

Okada, A. (1981): "On stability of perfect equilibrium points," *International Journal of Game Theory*, 10, 67–73.

Oprea, R., K. Henwood, and D. Friedman (2011): "Separating the Hawks from the Doves: Evidence from continuous time laboratory games," *Journal of Economic Theory*, 146(6), 2206–2225.

Ritzberger, K., and J. W. Weibull (1996): "Evolutionary selection in normal form games," *Econometrica*, 63, 1371–1399.

Robson, A. J. (1990): "Efficiency in evolutionary games: Darwin, Nash and the secret handshake," *Journal of Theoretical Biology*, 144, 379–96.

Schlag, K. (1995): "When does Evolution Lead to Efficiency in Communication Games," Mimeo.

Schlag, K. H. (1993): "Cheap talk and evolutionary dynamics," Bonn University Economics Department Disc. Paper B-242.

Selten, R. (1975): "Re-examination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4, 25–55.

Selten, R. (1980): "A Note on Evolutionary Stable Strategies in Asymmetric Animal Conflicts," *Journal of Theoretical Biology*, 84, 93–101.

——— (1983): "Evolutionary stability in extensive two-person games," *Mathematical Social Sciences*, 5(3), 269–363.

Sobel, J. (1993): "Evolutionary Stability and Efficiency," *Economic Letters*, 42, 301–312.

Spence, M. (1973): "Job market signaling," *Quarterly Journal of Economics*, 87(3), 355–374.

Taylor, P., and L. Jonker (1978): "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, 40, 145–56.

Thomas, B. (1985): "On evolutionarily stable sets," *Journal of Mathematical Biology*, 22(1), 105–115.

van Damme, E. E. C. (1991): *Stability and Perfection of Nash Equilibria.* Springer-Verlag, Berlin, Heidelberg.

Wärneryd, K. (1993): "Cheap talk, coordination, and evolutionary stability," *Games and Economic Behavior*, 5, 532–46.

Weibull, J. W. (1995): *Evolutionary Game Theory.* MIT Press, Cambridge, Mass.