The evolution of taking roles[*]

Florian Herold[†] and Christoph Kuzmics[‡]

August 17, 2017

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. Evolutionary stable social structures are different 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
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. The research leading to these results has received 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. Florian Herold gratefully acknowledges this support. This work was also supported by the COLIBRI initiative of the Karl-Franzens University of Graz, Austria.

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

[‡] Universität Graz, christoph.kuzmics@uni-graz.at

1 Introduction

Social interactions among humans (and arguably also among animals) are influenced by diverse social structures that evolved over time. The influence of a social structure is particularly striking when small payoff-irrelevant distinguishing features of players who started out as equals with identical capabilities result in very different behavior. We investigate under which circumstances people endogenously evolve into assuming different roles despite symmetric interaction, starting as equals, and despite completely random matching. We further study the stability of the emerging social structures.

The canonical example we consider is an interaction which has the structure of a symmetric 2×2 game in which both pure strategy equilibria are asymmetric. The best-responses to the opponents' pure strategies are as in the hawk-dove game: The best response to one pure strategy is the other pure strategy. We are interested in games where the two players earn different payoffs in the asymmetric equilibria and the two players therefore prefer different pure strategy equilibrium play. Games with such a structure are often employed by economists and other social scientists as archetypes to model either situations of conflict between two parties or situations in which players need to anti-coordinate (e.g. division of labor, specialization in different tasks). Consider the base game

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

W.l.o.g. we assume $a > b$ (otherwise simply switch the labels H and D). The key restrictions we impose on our game are $b > c$ and $a > d$.

This class of games is also the canonical example in the evolutionary game theory literature for a game in which the evolutionary stable strategy in a one population model is very different from the evolutionary stable strategies in two-population models. In the one-population model both players playing the game are drawn from the same population and the unique evolutionary stable strategy is the symmetric mixed Nash equilibrium. In the two population model the mixed Nash equilibrium is not even a neutrally stable strategy (and thus neither evolutionary stable) and the only evolutionary 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 their player-position and thus it is essentially impossible to play an asymmetric mixed strategy profile.

When we move from the one population model to the two population model we do however not only allow players to play contingent on their position, but we also assume that evolutionary competition occurs only within the population of each player-positions separately. We here want to consider a framework in which players have different labels which do not affect payoffs directly yet allow matched players with different labels to play asymmetrically. Evolutionary competition however is still evaluated across the entire population. Thus, not only actions conditional on labels evolve, but also the distribution of labels itself. More precisely we consider neutrally stable strategies (NSS) and evolutionary stable strategies (ESS) of the meta-game in which players can choose their label and an action that can condition on the opponent's label or type.[1]

An important observation (Lemma 2, points (a) and (b)) for all further results is that in any NSS matched players with different labels must anti-coordinate, while players must play the mixed Nash-equilibrium when matched with their own label.[2] To develop a better intuition for the coming results one can imagine that for the existing labels a social-structure develops that specifies for each combination of different labels which of the two asymmetric Nash equilibria is the convention. Given such a social structure, labels that do earn lower payoffs become less frequent over time. Hence in an NSS of the meta-game all labels in the supports of the NSS must earn the same payoff given this distribution of labels and given a social structure consistent with Lemma 2.

For a situation with two potential labels only, there is - up to a permutation in labels - only one social structure that is consistent with Lemma 2. A 'top' label will play H against the other 'bottom' label. A key distinction emerges from the evolutionary analysis between two (sub)classes. While all symmetric base-games we consider have the same best-response structure, it is interesting to note that these two classes of games are economically quite different games. We call the first sub-class the class of conflict games. In a game of conflict you always prefer the opponent to play the dove strategy D, independently of your own choice of action.[3] We call the second sub-class anti-coordination games. In anti-coordination games a player who did commit to playing action D would actually prefer the opponent to play action H. Thus, this class of games corresponds rather to a game of specialization in which, for instance, the two players act in a team and success requires two different skills. One skill is more attractive to acquire than the other, but given you acquired one particular skill, you prefer that your partner has acquired the complementary skill necessary for joint success. In conflict games with two potential labels only the top label can be in the

[1] Note that types play a somewhat different role here compared to the literature on the evolution of preferences under observability in which a type implies some subjective preferences and therefore certain strategic behavior (compare e.g. Dekel, Ely, and Yilankaya (2007) and Herold and Kuzmics (2009)). Here choosing a type does not constrain the strategies you may choose.

[2] The result that players with different roles must anti-coordinate is already present in Selten (1980), but he keeps these roles exogenous.

[3] The interaction in a conflict game corresponds e.g. to the Hawk-Dove game described in Maynard Smith (1982) in which the reward is equally shared if both animals retreat.

support of an NSS and despite the existence of the two labels we are essentially back to the result of the one-population model without labels. For anti-coordination games, in contrast, both labels are present in the support of the unique NSS, but bottom types in a lower proportion than top types. Thus the inefficiency is reduced, but some inefficiency remains in the NSS.

For more then two labels Lemma 2 is consistent with several social structures: For instance one might have a structure in which labels have a transitive order and higher labels play H against lower labels. We call this a hierarchical structure. Yet, also some circular structures are possible, which for an odd number of labels can lead to egalitarian outcomes (and for an even number of labels to approximately egalitarian outcomes). For a specific anti-coordination game Hurkens and Schlag (2002) showed this already and they conjecture that their result extends to other anti-coordination games. We confirm this conjecture, after we provide general characterizations of ESS and NSS in our general setup.

Then we focus our analysis on games of conflict. First, in games of conflict no hierarchical structure can form an NSS with more than one label in its support. Second another interesting distinction emerges. In conflict games in which the payoff d (if both players play dove) is below the average payoff that the two players obtain in the asymmetric equilibrium, an egalitarian types structure forms an ESS.

For a large number of labels complex social structures composed of stable substructures can be evolutionary stable. In the online supplement to this paper we provide a characterization of when such group substructures are stable. Finally, we summarize some key results contingent on the parameter d and discuss the welfare implications. It turns out that in our setting a stable egalitarian social structure is good for a society - even from an efficiency point of view. Intuitively, an egalitarian social structure makes all roles in society equally attractive and thus helps to avoid situations where too many players choose the more attractive roles in society.[4]

1.1 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 evolutionary stable. While in Selten (1980) these different roles are given exogenously, we are interested here in their endogenous evolution.

[4] Results in this spirit have been established for repeated symmetric games by Bhaskar (2000) and Kuzmics, Palfrey, and Rogers (2014), where the promise of an egalitarian continuation play induces efficient randomization in early rounds of play.

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

Most of this related cheap talk literature, such as Robson (1990), Wärneryd (1993), Sobel (1993), Blume, Kim, and Sobel (1993), Schlag (1993), Schlag (1995), Kim and Sobel (1995), Bhaskar (1998), Banerjee and Weibull (2000), Hurkens and Schlag (2002), focusses on coordination games and in how far cheap talk will - or will not - help to select against inefficient equilibria. The most closely related formal setup to our work is Hurkens and Schlag (2002) and Banerjee and Weibull (2000). While both papers focus on coordination games, the work by Hurkens and Schlag (2002) also has a section on a task allocation game which falls in our sub-class of anti-coordination games.

For this task-allocation game they find necessary conditions for ESS corresponding 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 type structure and respectively egalitarian or approximate egalitarian type structures. They also conjecture that that these results extend to a larger class of anti-coordination games. Our Propositions 1, 2, and 3, which 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 not discussed in the literature before.

We come back to the relation to Farrell (1987), Banerjee and Weibull (2000), and Hurkens and Schlag (2002) and discuss it in more detail in Section 4 after we 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.[5] W.l.o.g. we assume that $a, b, c \geq 0$ (one could always add a constant to all payoffs without affecting the incentives in the game). W.l.o.g. we assume $a > b$ (if not we would simply switch labels H and D).

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.[6] 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 in this paper will 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 her own choice of action. We shall call such games *conflict games.* In contrast, when $b > d$ then 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 later 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 $u^* = \frac{ab-cd}{a-d+b-c}$.*

[5] 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.

[6] 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 lead to everyone playing the dominant action. Players may send messages but they will not impact play. Coordination games are of interest in our context, but have already been subjected to a thorough analysis in Banerjee and Weibull (2000) and Hurkens and Schlag (2002), among others.

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).*

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

6. *Keeping the parameters a, b, c fixed the mixed equilibrium payoff u^* is strictly increasing 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 A.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) *types.* One could also call these roles or labels or (commonly distinguishable) messages. To avoid confusion we exclusively use the term *types* throughout the main Sections 2 and 3 of this paper. The finite set of types is given by *type space* Θ. Types are, therefore, payoff-irrelevant, but perfectly observable and players can condition their play on the opponents type $\theta' \in \Theta$.

The formal setup is, thus, almost identical to that of Banerjee and Weibull (2000) and Hurkens and Schlag (2002), except that we study the entire class of H-D base-games and focus on the resulting social-structures, while their focus is to study equilibrium selection in coordination games and in an anti-coordination game.[7]

[7] 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 type. We prefer to reduce the number of strategies, without losing anything, by allowing players to condition only on their opponent's type. We thus, follow Schlag (1993), Schlag (1995) and Hurkens and Schlag (2002), 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 clarifying when a failure of evolutionary but not neutral stability is simply due to

Let $F = \{f : \Theta \to A\}$ the (finite) set of action-functions. Then $f(\theta')$ provides the action that a player chooses against an opponent of type θ'.

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 types as well as, for each adopted type, 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 type $\theta \in \Theta$. Furthermore we denote the (conditional) probability that a player of type θ, given strategy $\sigma \in \Delta(S)$ plays H against an opponent of type θ' by $x_\theta(\theta') = \frac{\sum_{f \in F, f(\theta')=H} \sigma(\theta,f)}{\sigma(\theta)}$.[8]

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.[9] 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 $\theta, \theta' \in \Theta$:[10]

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

2.3 The Solution Concept

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

a large number of equivalent strategies or due to a more fundamental problem intrinsic to the game under analysis.

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

[9] 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 $\theta, \theta' \in \Theta$.

[10] 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.[11] 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 evolutionary 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 type 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 type 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 type 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 $\theta, \theta' \in \Theta$:
$x_\theta(\theta') = 1 - x_{\theta'}(\theta) \in \{0, 1\}$.

(c) For all $\theta \in \Theta$: $\sigma(\theta) > 0$ (all types 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') $\frac{a+b}{2} \geq d$.

[11] 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 A.2. 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 types 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 type 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 in the single population case (i.e. here, whenever two individuals of the same type meet) the only evolutionary stable outcome is the symmetric mixed equilibrium. We known from Selten (1980) that in the multiple population model (i.e. here, whenever two individuals of different types meet) the only evolutionary stable outcome must be a strict, and, hence, pure and possibly asymmetric equilibrium.

For a specific anti-coordination game (task allocation game) Hurkens and Schlag (2002, Lemma 2(ii)) provided corresponding necessary conditions for evolutionary stability and showed that in any ESS all types must be played with positive probability.

Note that Lemma 2, in its characterization of NSS, is mute about strategies σ without full type 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 types Θ. For any strategy σ of this game let Θ_S denote the set of types $\theta \in \Theta$ with $\sigma(\theta) > 0$. Let furthermore $\sigma|\Theta_S$ denote the strategy σ restricted to the set of types Θ_S. Then the following two statements are true.*

(a) If σ is an NSS of the meta game with set of types Θ then $\sigma|\Theta_S$ is a (full type support) NSS of the meta game with the same H-D base game with the set of types Θ_S.

(b) Let $\Theta_S \subset \Theta$, with $|\Theta_S| \geq 2$. Let σ' be a strategy of the meta game with set of types Θ with support only on Θ_S. 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 $\theta \notin \Theta_S$, $\sigma(\theta) = \sigma'|\Theta_S(\theta)$ for all $\theta \in \Theta_S$ and identical $x_\theta(\theta')$ for all $\theta, \theta' \in \Theta_S$.

Note that Lemmas 2 and 3 are silent about the distribution over types in Θ in an NSS. Understanding this is where the 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 types Θ. Given a meta-game strategy $\sigma \in \Delta(S)$, we call the induced type behavior x, with $x_\theta(\theta') \in \Delta(A)$ the behavior of type θ when meeting type θ' as defined in Section 2.2, the induced* type structure*.*

(i) A pre-stable type structure *is a type structure satisfying the following conditions:*

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

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

(ii) The induced type game *of a pre-stable type 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 type structure we can now investigate how the composition of types evolves in the corresponding reduced form "type game". First, we investigate which distribution of types leads to an Nash-equilibrium in this type game. Then it is straightforward to check whether the corresponding strategies are 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 types Θ with $|\Theta| \geq 2$.*

(a) There exists an ESS $\sigma \in \Delta(S)$ of the meta game with a certain type structure, if and only if the type structure is pre-stable, the corresponding type game has a full support Nash-equilibrium, and $\frac{a+b}{2} > d$.[12]

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

The proof follows immediately from Lemma 2. Furthermore, if σ is an ESS of the meta game, then it must be the unique ESS with this type 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 type game. In the type game it is a full support ESS and must therefore be unique.[13] But then no other strategy of the meta game with the same type game can form an ESS.

[12] Within the type game the ESS condition is only $\frac{a+b}{2} > u^*$. Yet, for stability in the meta game, we need the more restrictive condition $\frac{a+b}{2} > d$: there are H-D base-games 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 3 optimal pure strategies playing H against its own type, and then the three optimal pure strategies playing D against its own type). Then, e.g., the mutant strategy $\mu = \frac{1}{27}(4, 4, 4, 8, 2, 5)$ violates the SOC of NSS (and thus also for ESS).

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

To check whether there exist NSS of the meta game with a certain pre-stable type structure without full type support, it is still useful to look at Nash equilibria of the type game (without full support), yet we need to check that all best responses in the meta game, do perform weakly worse against themselves than the NSS strategy against this mutant strategy.

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 type structure with $x_\theta(\theta) = x^*$ for all $\theta \in \Theta$.*

(a) If there is an order of types $\succ$ 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 type structure*.*

(b) Suppose $|\Theta|$ is odd. If $x_\theta(\theta') = 1$ (plays H) for exactly half of all types $\theta' \neq \theta$ and $x_\theta(\theta') = 0$ (plays D) for the other half of all types $\theta' \neq \theta$, then x is called an egalitarian type structure*.*[14]

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

Note that these definitions are not empty, meaning that we can indeed construct a strategy $\sigma \in \Delta(S)$ with a hierarchical type structure and we can also construct one with an egalitarian type-structure, provided the number of types in Θ is odd.[15] If the number of types in Θ is even we can also construct a strategy $\sigma \in \Delta(S)$ with an approximate egalitarian type structure.[16]

[14] If $|\Theta|$ is even, then an exactly egalitarian type structure is obviously impossible. Yet, one can define an close to egalitarian structure in which half the types play exactly one more time H than D and the other half one more time D then H. The results would be very similar to the results we obtain for odd numbers of types, but would complicate the arguments and notation.

[15] There exist several ways to visualize an egalitarian type structure. For instance, one could arrange types in Θ on a circle such that each type θ plays H against the $(n-1)/2$ types located clockwise from θ and plays D against all other types.

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

As an example consider the case $\Theta = \{T, M, B\}$, i.e. $|\Theta| = 3$. The hierarchical and egalitarian, respectively, type structures can be represented in terms of the induced "type game" given by the following two matrices.

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

hierarchical egalitarian

For the hierarchical type structure type T is the "top type" and plays H against all other types. Type M is the middle type, who plays D against type T and H against type B. Type B is the bottom type, who plays D against all other types.

In the egalitarian type structure all types play H against one other type and D against the remaining other type. Thus, they are all in equal or "egalitarian" positions.

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

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

(a) There exists an ESS of the meta game with an hierarchical type structure if and only if the base game is an anti-coordination game, i.e. $b > d$. *This ESS is the unique symmetric equilibrium with hierarchical type structure.*

(b) There exists an NSS of the meta game with a hierarchical type structure for all H-D base-games. For anti-coordination games this NSS is the unique ESS with hierarchical type structure with full type support from above. For conflict games this hierarchical NSS has only the top-type in its support.

For conflict games the unique symmetric Nash-equilibrium of the hierarchical type game puts all weight on the top-type 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. There exists a strategy of the meta-game of any H-D base-game that induces an egalitarian type structure and has full type 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 $d < \frac{a+b}{2}$ then such a strategy inducing an egalitarian type structure forms an ESS (and thus also an NSS) of the meta game.

(b) If $d > \frac{a+b}{2}$ then such a strategy inducing an egalitarian type structure is not an NSS (and thus also not an ESS) of the meta game.

The proofs of Proposition 1 and 2 are relegated to Appendices A.5 and A.6. They both follow the same steps. First, we compute the unique symmetric full type 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 approximates $\frac{a+b}{2}$ as n gets large - 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 type 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 there exists a strategy of the meta-game of any H-D base-game that induces an approximate egalitarian type structure and has full type support.*

In the supplementary material, not intended for publication, we show that for conflict games and for $|\Theta| = 4$ and for $|\Theta| = 6$ the meta game has no ESS.

3.3 Welfare

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

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

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

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

(c) *If $\frac{a+b}{2} > d \geq b$: For $|\Theta| \geq 3$ several NSS with distinct payoffs exist. The NSS with only one type in its support give the minimum payoff among pre-stable equilibria. For odd $|\Theta|$ 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) *If $b > d$: In these anti-coordination games many ESS and even more NSS exist. Egalitarian ESS (which exist for odd $|\Theta|$) give the maximum expected payoff. For even $|\Theta|$ approximate egalitarian ESS exist. Hierarchical ESS also exist and are Pareto dominated by the egaliatrian or approximate 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 types 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 types 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 type NSS in this case.

4.1.2 Connection with Farrell, 1987

Farrell (1987) was, as far as we know, the first who studied 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 our language this is a set of types. For each pair of types $\theta, \theta' \in \Theta$ 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 of our paper, Farrell (1987) investigates equilibria of the meta-game that satisfy

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

It is straightforward to see that this corresponds to what we here call the "hierarchical" type structure. We can reproduce Farrell's (1987) result by noting that every meta-game with a finite number of types has a (unique - up to relabelling of types) 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 game, we know that there is no hierarchical NSS (or even Nash equilibrium). Imposing the hierarchical structure in these games would yield the result that every hierarchical NSS places probability 1 on a single type. We also know now, however, that there are other NSS, based for instance 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 "types" 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 type game has an NSS (provided $u^* < \frac{a+b}{2}$), in which the first three types are used with probability $\frac{1}{3}$ each, while type (D, D) is not used.

4.1.3 Connection with Hurkens and Schlag, 2002

Hurkens and Schlag (2002) consider the effect 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 of the base game without conditioning on the opponent's message. For coordination games Hurkens and Schlag (2002) find that without the option to stay away from cheap talk there exists an inefficient ESS, but with the option to stay away from cheap talk, the set of strategies resulting in the efficient outcome is the unique evolutionary stable set. Most related to our paper 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 for their task allocation game evolutionary stable strategies that correspond to our hierarchical type structure and to our egalitarian or approximate egalitarian type structure, respectively. They continue their analysis of the task allocation game by adding again the option to stay away from communication (Section 4.2) and show that while (in our terminology) the hierarchical type structure equilibrium remains evolutionary stable, all evolutionary stable strategies are bounded away from the efficient outcome (and hence the egalitarian type structure is not evolutionary 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 (types) of the opponent it seems difficult to visibly commit not to do so.

4.2 Adding one more type

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

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

One way to think about this is to assume nothing other than evolution will now lead to some new NSS, in which the old types interact with each other as before, but any stable behavior between the new and old types can emerge. If this is our view, then there is an even sharper distinction 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 types in its support. So the multiplicity of types is in this sense stable or robust to the introduction of a radically new type.

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 type can evolve to be such that it plays H against all other types and they play D against it. In this case, however, this new type dominates all other types, and the only NSS with this type structure is the one in which the new type receives probability weight one. In this sense, the possible multiplicity of types in conflict games is not stable or robust to the introduction of a radically new type.

Going back to anti-coordination games, it is interesting to note, that, while the multiplicity of types is robust, the NSS can nevertheless change dramatically from before to after the introduction of the new type. To see this consider the three-type hierarchical structure with one added type 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 type game without type X has a unique NSS, and that NSS has full type support for anti-coordination base games. The meta game with type X has an egalitarian NSS with equal support on T,M, and X, 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

We investigated the evolution of taking roles in symmetric 2×2 games with asymmetric pure strategy equilibria. We provided a characterization of evolutionary (and neutrally) stable strategies in the meta game. Depending on the parameters and the number of payoff-irrelevant types (or labels or roles) we discussed social structures that can emerge as evolutionary (and neutrally) stable strategies and their welfare implications. Two structures of particular interest are the egalitarian and hierarchical type structure. The results are very different for two sub-classes of these base games: conflict games and anti-coordination games. In situations in which different payoffs can be sustained by neutrally stable strategies, the payoffs in the egalitarian type structure Pareto dominate the payoffs of the hierarchical types structure. In this sense an egalitarian organized social structure can promote efficiency in our setting.

It remains an interesting question which neutrally (or evolutionary) stable strategies are likely to emerge if several exist. One way to think about it is to consider further equilibrium refinement. Hierarchical social structures seem fairly common and typically a structure that is cognitively easier to process. If we would assume in addition that players have small cognitive costs for perceiving different types as separate, this may influence which equilibria remain stable and seems an interesting road for further inquiry.

Another way to deal with multiple stable equilibria is to embrace them: It is in the case of multiple stable equilibria that our theoretical analysis and recommendations might have an impact by changing from one focal point equilibrium to a new (and hopefully better) one.[17] Thus, economic theorist might see multiple stable equilibria as a blessing rather than a curse.

[17] For instance, Roger Myerson pointed out in his lecture at GAMES 2016 that much of economic activity (such as exchanging goods for a piece of paper called money) might be best interpreted as a shift from one focal equilibrium to another.

A Proofs

A.1 Proof of Lemma 1

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

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

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 $u^* = d + y^*(b-d) = \frac{ab-cd}{a-d+b-c}$, independently of his 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.

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

Now we prove point 5: For conflict games $(b \leq d)$ we now 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 $(b > d)$ we now 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^* is strictly increasing in d. Furthermore,

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

and, by continuity of u^*

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

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, thus, have

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

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

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

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.

A.2 Proof of Lemma 2

First, we prove that having a full type 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 types are played with positive probability (Condition (c)). Suppose to the contrary that there exists a type $\hat{\theta} \in \Theta$ with $\sigma(\hat{\theta}) = 0$. Then consider the strategy σ' which is identical to σ except (potentially) when playing against the type $\hat{\theta}$ (which does not happen in equilibrium). If playing against an opponent with type $\hat{\theta}$, strategy σ' would plays D (dove). Clearly, $u(\sigma', \sigma) = u(\sigma, \sigma) = u(\sigma', \sigma') = u(\sigma, \sigma')$ and hence σ can only be ESS if it happens to be identical to σ'. But then the strategy s playing $\hat{\theta}$ with probability one and playing H (hawk) against all types in the support of σ would obtain a payoff $u(s, \sigma) = a > u(\sigma, \sigma)$ which contradicts that σ was ESS.

Second, we prove that conditions (a) to (d) are necessary conditions for having a full type 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 there is a $\theta \in \Theta$ 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 words, strategy σ' mimics strategy σ in all respects except when adopting type θ and meeting type θ it plays D.[18] Strategy σ' thus generates different payoff than strategy σ against σ only when both strategies adopt type 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 type 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 $\sigma(\theta') > 0$. A similar argument to the one above that proves part (a) implies that each of the two types must play a best response to the other type, otherwise the FOC of neutral stability is violated. It remains to be shown that the two types playing the symmetric equilibrium of the base game against each other can not 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 types θ and θ' are employed and then, as σ prescribes the

[18] It is easy to see that such a strategy exists. It may not be unique. See footnote 9.

mixed strategy equilibrium strategy x^* both pure actions of the base game H and D give equally 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 follow inequalities are equivalent

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

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 type support NSS and for ESS.

Condition (c) is obviously necessary for full type support NSS and hence 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 A.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, there is no further equivalent strategy since for any match of different types a pure (contingent) strategy is played. 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 type $\theta \in \Theta$ there are two corresponding pure best replies to σ: Select type θ, play against other types $\theta' \neq \theta$ whatever $x_\theta(\theta')$ the strategy σ prescribes, and play against your own type 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).[19] This corresponds to part (a) of the following lemma. In part(b)

[19] Van Damme attributes Theorem 9.2.7 to Haigh (1975) and Abakus (1980). Similar arguments are used in the proofs of Hurkens and Schlag (2002) and inspired our strategy of proof.

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, $supp(\sigma)$. 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 $supp(\sigma)$.*

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

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

(b) Then σ is an NSS if and only if

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

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

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 type, and strategies from $|\Theta| + 1$ to $2|\Theta|$ those in which "D" is played against an opponent with the same type. 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

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

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 of $a > b > c$. 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, $\frac{a+b}{2} \leq d$ then 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 $\mathbf{y}^T M\mathbf{y} \geq 0$ 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 $\mu \in \Delta(supp(\sigma))$:

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

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 $y \equiv \mu - \sigma$

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

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 a 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 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.

A.3 Proof of Lemma 3

We first prove Point 1. Suppose not. Then there exists a mutant strategy $\mu|\Theta_S$ in the restricted meta game such that violates either the FOC or SOC of NSS in the restricted meta game. The same strategy extended to the full meta game with full set of types Θ must violate the same NSS condition in the meta game, since all extra types are played with probability 0 and do not change expected payoffs. Hence any NSS of the meta game, must also form an NSS in the meta game restricted to types in the support of its strategy, which proves Point 1 of Lemma 3.

We not turn to proving Point 2. A strategy that supports the NSS in the meta game with the larger set of types Θ is a straightforward extension of the full type support NSS strategy from the meta game with smaller set of types Θ_S by specifying that $x_\theta(\theta')$ corresponds to the strategy of the base game that gives the opponent only his 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 type in the support of an NSS strategy can play always D against all other types in the support (otherwise it is dominated by any of the other strategies in the support, as these obtain sometimes $a > b$ and never below b). Hence there is a type θ' with $\sigma(\theta') > 0$ such that $v \geq \sigma(\theta')a + (1 - \sigma(\theta'))b > b = \min\{u^*, b\}$. This proves Point 2 of Lemma 3.

A.4 Proof of Lemma 4

The proof follows immediately from Lemma 2.

A.5 Proof of Proposition 1

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

Lemma 7. *Let $n \equiv |\Theta| \geq 2$. There exists a full support Nash equilibrium (σ^*, σ^*) of the type-game of the (pre-stable) hierarchical type-structure if and only if the base game is an anti-coordination game ($u^* < b$). If the types $\theta_1, \theta_2, \ldots, \theta_n$ are ordered according to the hierarchical structure (with θ_1 top* *type), 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 type earning the average payoff

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

Equivalently,

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

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 type 1 top-type) 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 type k (fixing the hierarchical type structure) must yield the same expected payoff. The payoff to type 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. If the game in hand is one of anti-coordination, 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.

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

A.6 Proof of Proposition 2

Under an egalitarian type structure each type plays H against half of all other types 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 $|\Theta| = 2l + 1$) there are such egalitarian pre-stable structures, i.e. it is a well defined structure. We can, for instance, locate the $2l+1$ types on a circle and each type plays H against the next l types located clockwise and D against the next l types located anti-clockwise.

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

The expected expected payoff of each strategy in the type 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 $d < \frac{a+b}{2}$ it cannot form an NSS, q.e.d.

A.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 approximate egalitarian pre-stable type structure with a full support Nash equilibrium in the type game. (For their task allocation game Hurkens and Schlag (2002) have a similar construction in the proof of their Prop. 3). Imagine the $2l$ types placed on a circle. Types $i \in \{1, \ldots, l\}$ play H against the l next types located clockwise and D against the $l-1$ types located anti-clockwise. Types $i \in \{l+1, \ldots, 2l\}$ play H against the $l-1$ next types located clockwise and D against the l types located anti-clockwise. This forms an pre-stable structure if all types play also x^* against their own type.

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

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

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

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

Note first, that for conflict games ($u^* \geq b$) the equation $\alpha_l\left(u^* - b\right) + \alpha_{l+1}\left(a - u^*\right) = 0$ has no solution (with $\alpha_l, \alpha_{l+1} \geq 0$). For $|\Theta| = 4$ it is straightforward to show that all approximate egalitarian structures have the structure above and thus no approximate egalitarian structure can be part of an ESS of the meta-game in this case.

Now we prove that there is an ESS of the meta-game under the approximate egalitarian type structure 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 G-induced payoff difference matrix 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 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:[20]

(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 $\bar{D}x = h$.

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

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

Consider the type game with an approximate egalitarian structure as described above with an even number of types $n = 2l$, 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 $u^* - b$, the second 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 recent 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

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

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 $v \in \mathrm{I\!R}^{n-1}$ such that $v^T d_i > 0$ for all i. The proof is by contradiction. Thus, suppose there is such a $v \in \mathrm{I\!R}^{n-1}$ 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 that $v_i > 0$ for all i. But then $v^T d_1 < 0$ 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 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 $r^* \equiv ((a - u^*, 0, \dots, 0, (b - a), (b - a), 0, \dots, 0, (u^* - b)$. We must have $r^* x = 0$ (which corresponds to the payoff difference between the last and first type). In particular, $x_{2l} = 0$ (otherwise $r^* x$ would be negative 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 $x_1 > 0$. Yet, 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.[21]

A.8 Proof of Lemma 5

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

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

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 type $\theta_T \in \Theta$ and with $\sigma(\theta) = 0$ for all remaining types $\theta \neq \theta_T$. 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, q.e.d.

A.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 type support can exist for $|\Theta| \geq 2$. Now if any NSS with more than two types in its support would exist, then, by Lemma 3 it would also be an NSS in the game restricted to the set of types in the support Θ_S. But in this restricted meta game it would be a full support equilibrium, a contradiction.

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

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

[21] Note that there do exist equilibria 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 $2l-1$ types. But these equilibria without full label support do not satisfy the Equal Payoff Condition (I) and thus there must also exist one with full label support satisfying the Equal Payoff Condition (I).)

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

References

Abakus, A. (1980): "Conditions for evolutionary stable strategies," *J. Appl. Prob.*, 17, 559–562.

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

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

Bhaskar, V. (2000): "Egalitarianism and efficiency in repeated symmetric games," *Games and Economic Behavior*, 32(2), 247–262.

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

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," *The American Economic Review*, 88.1, 157–179.

Farrell, J. (1987): "Cheap Talk, Coordination, and Entry," *The RAND Journal of Economics*, 18.1, 34–39.

Haigh, J. (1975): "Game Theory and Evolution," *Adv. Applied Prob.*, 7, 8–11.

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 (2002): "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.

Kuzmics, C., T. Palfrey, and B. W. Rogers (2014): "Symmetric play in repeated allocation games," *Journal of Economic Theory*, 154, 25–67.

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.

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

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

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.