

Artificial Distinction and Real Discrimination*

By ELENA INARRA[†] AND ANNICK LARUELLE^{‡§}

April 11, 2011

Abstract

In this paper we consider the hawk-dove game played by a population formed by two types of individual who fail to recognize their own type but do observe the type of their opponent. In this game we find two evolutionarily stable strategies and show that in each of them, and for any distribution of types, one type of individuals suffers more aggression than the other. Our theoretical results are consistent with the conclusions drawn from an experimental study into the behavior of a group of domestic fowls when a subgroup has been marked. (JEL C72)

*This research is supported by the Spanish Ministerio de Ciencia e Innovación under project ECO2009-11213, co-funded by the ERDF, and by Basque Government funding for Grupo Consolidado GIC07/146-IT-377-07. We thank Inma Estevez for drawing our attention to this case study. We thank Andre Barreira da Silva Rocha, Ken Binmore, Francis Bloch, Federico Grafe, Alan Grafen, Jean-François Laslier and Federico Valenciano for their comments. We also thank Marta San Martin for her help in the drawings. Elena Iñarra also gratefully acknowledges the hospitality of Oxford University and financial support from the Basque Government.

[†]*BRiDGE*, Departamento de Fundamentos del Análisis Económico I, Universidad del País Vasco, Avenida Lehendakari Aguirre, 83, E-48015 Bilbao, Spain; elena.inarra@ehu.es.

[‡]*BRiDGE*, Departamento de Fundamentos del Análisis Económico I, Universidad del País Vasco, Avenida Lehendakari Aguirre, 83, E-48015 Bilbao, Spain; a.laruelle@ikerbasque.org.

[§]IKERBASQUE, Basque Foundation of Science, 48011, Bilbao, Spain.

1 Introduction

Scientific experiments using animals frequently involve the marking of animals in a way that artificially changes their phenotype. For instance penguins with flipper bands have been used to study climate change. However, as Saraux *et al.* (2011) show, this banding reduces penguins' survival rate. Consequently these studies may mix up the effects of banding with other changes in penguin life and therefore be inaccurate. To analyze the effect of marking within a group of birds, Dennis *et al.* (2008) conduct an experiment in which a subgroup of domestic fowls is marked. The two most salient results of their research are that marked birds, for any of the proportions of the population under study, suffer more aggression and have less body mass than their unmarked pen mates.

In this experiment each bird can observe whether others are marked or unmarked, but does not know whether it is marked itself or not. This stands in sharp contrast with the usual assumption made in game theory that a player knows her type, but may have only partial information concerning the type of the other players. To the best of our knowledge no situations where individuals lack self-perception but are able to observe others' type have ever been modeled¹. Here we consider this feature, which underlies the experimental work of Denis *et al.* (2008).

We propose a variation of the classic hawk-dove game where individuals are unaware of their own type but see their opponent's type. We use evolutionarily stable strategy (Maynard Smith and Price, 1973) for the analysis of this game.

We refer to the hawk-dove game in a finite population as the "homogenous game", and modify it by artificially dividing the population into two types. Each individual can watch all other individuals and observe their type, but does not have information concerning her own type. Hence individuals can condition their actions on their information about the other individuals' types but not on their own. We refer to this game as the "heterogeneous game".

Though the division of a population into two types of individuals is not linked to any disparity in capacity, we find that it affects the behavior of the players. The evolutionarily stable strategy of the homogeneous game played against any type of individual is not evolutionarily stable in the heterogeneous game. Interestingly enough, the heterogeneous game has exactly two evolutionarily stable strategies, in each of them an individual's action depends on her opponent's type. In each equilibrium, one type of individual always suffers more aggression than the other. We refer to the type which is attacked more as *alien* and

¹Two illustrations of such type of situations are a card game called the "Indian poker game" (See http://en.wikipedia.org/wiki/Blind_man's_bluff_poker) and a sequence in Tarantino's movie "Inglourious Basterds" (2009).

to the other type as *local*. Not surprisingly, we find that an alien is always worse off than a local, and that an alien's expected payoff increases as the proportion of aliens within the population increases. For a local, however, the expected payoff increases as the proportion of aliens increases up to a certain point and decreases thereafter. Furthermore for an individual taken at random the existence of a small proportion of aliens is beneficial whereas a larger proportion is detrimental.

Our theoretical results support the conclusions obtained by Dennis *et al.* (2008). Considering hawkish behavior as a proxy of the pecking and threatening between birds observed in their experiment, we find that a strategy of more aggressive behavior toward marked birds than toward unmarked ones can be evolutionarily stable whatever the proportion of the population that has been marked. Moreover, the fact that in this experiment marked birds have less body mass than their unmarked pen mates is also supported under the assumption that the body mass of a bird can be evaluated through the expected payoff of an individual in a heterogeneous game.

Now let us compare our results with the relevant literature. In the seminal hawk-dove game only the mixed strategy in which the probability of each individual playing hawk is equal to the ratio between the value of the resource and the cost of fighting is evolutionarily stable. Maynard Smith and Parker (1976) propose a variation in which individuals fight for a territory, each player being either the "owner" or the "intruder". They show that the "bourgeois" strategy, *i.e.* playing hawk when one is the owner and dove when the intruder, is evolutionarily stable. Going further, Selten (1980) proves that only pure strategies are evolutionarily stable. However, Binmore and Samuelson (2001a, 2001b) consider the two roles that an individual may play jointly with payoff perturbations, and show that under certain conditions mixed strategies can also be evolutionarily stable in this game. This last result goes along with the findings obtained in our variation. When individuals lack self-perception but observe the type of the others pure as well as mixed strategies may be evolutionarily stable, depending on the proportion of aliens in a population.

This study can also be linked to the literature on social dynamics, in particular to the work by Axtell, Epstein and Young (1991). These authors consider the divide-one-dollar game, in which each individual may choose among the following three actions: high (ask for 70 cents), medium (ask for 50 cents) and low (ask for 30 cents) claims. The dollar is divided according to claims whenever they are feasible, otherwise players receive nothing. The dynamics of random bilateral encounters in large populations show that in the long run any two players in the population tend to demand a medium claim. However, if the population is artificially divided into two groups then a discriminatory norm emerges in society. An equilibrium

where the members of one group make a high claim and the members of the other group make a low claim when pairs belonging to different groups meet may persist for substantial periods of time, in which there is intra-group dissension. We reach a similar conclusion in a completely different setting: artificial division of the individuals into two types gives rise to true discrimination against one of the types, the aliens, who moreover behave aggressively among to one another.

The rest of the paper is organized as follows. Section 2 presents the homogeneous game and its unique evolutionarily stable strategy. Section 3 introduces the heterogeneous game and derives the evolutionarily stable strategies and expected payoffs. Section 4 analyzes the experimental work in the light of our game theoretical results. Section 5 concludes.

2 The homogenous game

Consider a population of n identical individuals in which any pair faces a contested resource of value v and may fight at a cost c . The size of the population n , the value v and the cost c are considered to be fixed with $v < c$. Each individual can be either aggressive and behave as a hawk or passive and behave as a dove. If an individual behaves as a hawk and her opponent as a dove, the aggressive individual gets the resource v while the passive individual gets nothing. If both individuals act as hawks, there is a fight. The winner gets the resource while the loser faces the cost c . Assuming that the two individuals have the same probability of winning the fight, the expected payoff for each one is half the resource minus the cost. If the two individuals behave as doves one withdraws and the other gets the resource. Assuming that the two individuals have the same probability of withdrawing the expected payoff for each one is half the resource. This description corresponds to the classic hawk-dove game played by a population of identical individuals, which we refer to as the *homogeneous game* Γ whose payoff matrix may be represented as follows²:

	hawk	dove	
hawk	$(\frac{v-c}{2}, \frac{v-c}{2})$	$(v, 0)$	(1)
dove	$(0, v)$	$(\frac{v}{2}, \frac{v}{2})$	

Let α denote the probability of playing hawk so that an individual can choose either a pure hawk ($\alpha = 1$) or dove ($\alpha = 0$) strategy or a mixed strategy ($0 < \alpha < 1$). Let $u(\alpha, \beta)$ be the expected payoff of an individual that plays α when her opponent plays β . That is,

$$u(\alpha, \beta) = \frac{v}{2}(1 - \beta) + \frac{c}{2}(\frac{v}{c} - \beta)\alpha. \quad (2)$$

²If we have $v > c$ then the structure of the game is equivalent to a prisoner's dilemma, while if we have $v = c$ then it is equivalent to a coordination game.

Since game Γ is symmetric the opponent's expected payoff is given by $u(\beta, \alpha)$.

The concept of "evolutionarily stable strategy" introduced by Maynard Smith and Price (1973) is applied to solve the hawk-dove game. This notion captures the resilience of a given strategy against any other strategy in the following sense: Consider a population where most members play an evolutionarily stable strategy while a small proportion of mutants choose a different strategy. In this situation each mutant's expected payoff is smaller than the expected payoff of a "normal" individual, so that the mutants are driven out from the population³.

An evolutionarily stable strategy may be formally determined as follows. Recall that a best response is a strategy that yields the highest payoff given the opponent's strategy. Let $\mathcal{B}(\beta)$ denote the set of an individual's best responses to an opponent playing strategy β . The two conditions for strategy α^* to be evolutionarily stable are: (i) $\alpha^* \in \mathcal{B}(\alpha^*)$ and (ii) for any $\beta \neq \alpha^*$ such that $\beta \in \mathcal{B}(\alpha^*)$ we have $u(\alpha^*, \beta) > u(\beta, \beta)$. Condition (i) states that α^* has to be a best response to itself. That is, the pair of strategies (α^*, α^*) is a symmetric Nash equilibrium (Nash, 1951). Condition (ii) states that if the opponent plays a best response to α^* (other than α^*) then the payoff of playing α^* is strictly greater than the payoff of playing that best response.

For game Γ the set of an individual's best responses to an opponent playing β is

$$\mathcal{B}(\beta) = \begin{cases} \{1\} & \text{if } \beta < \frac{v}{c} \\ \{\alpha \mid \alpha \in [0, 1]\} & \text{if } \beta = \frac{v}{c} \\ \{0\} & \text{if } \beta > \frac{v}{c}. \end{cases}$$

That is, if the probability of the opponent playing hawk is smaller than the ratio between the resource and the cost, the unique best response is to play hawk, while if it is greater than that ratio the unique best response is to play dove. If the probability of the opponent playing hawk is equal to that ratio then any strategy is a best response. For this game strategy $\frac{v}{c}$ is the only evolutionarily stable strategy. It is the only strategy that is a best response to itself, $\frac{v}{c} \in \mathcal{B}(\frac{v}{c})$, and it satisfies Condition (ii): $u(\frac{v}{c}, \beta) - u(\beta, \beta) = \frac{c}{2}(\frac{v}{c} - \beta)^2 > 0$ for $\beta \neq \frac{v}{c}$.

3 The heterogenous game

3.1 The model

Consider a division of the n individuals into two types, A and B , where x ($0 < x < 1$) is the proportion of individuals of type B . Assume that a pair of individuals is selected at random. This is equivalent to assuming that an individual is randomly chosen from a group of n

³See Maynard Smith (1982), Chapter 2, and Weibull (1995), Chapter 2, for a detailed explanation of this notion. A good introduction can also be found in Osborne (2004), Chapter 13.

individuals, and then an opponent is randomly chosen from the remaining $n - 1$ individuals. Thus, in any bilateral encounter between two individuals there are four possible cases: both individuals are of type A ; the first individual is of type A and the second of type B ; the first is of type B and the second is of type A ; or both are of type B . The probabilities of these four possible encounters, which we denote respectively by $p(A, A)$, $p(A, B)$, $p(B, A)$ and $p(B, B)$ are given by

$$p(A, A) = \frac{(1-x)(n-nx-1)}{n-1}, p(A, B) = p(B, A) = \frac{x(1-x)n}{n-1} \text{ and } p(B, B) = \frac{x(nx-1)}{n-1}. \quad (3)$$

Since the division of individuals into types is done random, it does not reflect genetic differences between individuals. Consequently the probability of winning or losing a fight is not determined by the distribution of types. Hence the hawk-dove game, whose payoff matrix given by (1) is played in each state of nature.

The essential feature of this model is that individuals fail to recognize their own type but do observe their opponent's type. This implies that the first individual does not distinguish between states (A, A) and (B, A) nor between (A, B) and (B, B) while the second individual does not distinguish between states (A, A) and (A, B) nor between (B, A) and (B, B) .

What is a strategy in this context? Obviously individuals find themselves in a position of choosing a probability of playing hawk for each type of opponent. A *strategy* can thus be represented by $\alpha = (\alpha_A, \alpha_B)$ where α_A gives the probability of behaving as a hawk when facing an individual of type A and α_B gives the probability of behaving as a hawk when facing an individual of type B . Of course any strategy α played by an individual in a homogeneous game can be played against either type of opponent, *i.e.* $\alpha_A = \alpha_B = \alpha$. Such a strategy is referred to as *homogeneous* by contrast to a *heterogeneous* strategy where $\alpha_A \neq \alpha_B$. Thus there are two pure homogeneous strategies: hawk against either type of opponent, $(1, 1)$, and dove against either type of opponent, $(0, 0)$; and two pure heterogeneous strategies: dove against individuals of type A and hawk against individuals of type B , $(0, 1)$, and hawk against individuals of type A and dove against individuals of type B , $(1, 0)$.

The *expected payoff* of an individual who lacks selfperception, who plays $\alpha = (\alpha_A, \alpha_B)$ while the opponent plays $\beta = (\beta_A, \beta_B)$ is the sum of the expected payoffs that she would obtain in each distinct encounter weighted by its probability of occurrence. For instance, in the encounter (A, B) the first individual recognizes her opponent as being of type B and plays hawk with probability α_B while the latter recognizes the former as an individual of type A and plays hawk with probability β_A , being $u(\alpha_B, \beta_A)$ the individual's expected payoff derived from such bilateral encounter. This expected payoff is multiplied by the probability of encounter $p(A, B)$. The expected payoffs in each of the remaining encounters are defined

analogously. Therefore the expected payoff of an individual playing α against an opponent playing β is given by $U(\alpha, \beta)$. That is,

$$U(\alpha, \beta) = p(A, A)u(\alpha_A, \beta_A) + p(A, B)u(\alpha_B, \beta_A) + p(B, A)u(\alpha_A, \beta_B) + p(B, B)u(\alpha_B, \beta_B).$$

Using (2) and (3) the expression above can be written as

$$\begin{aligned} U(\alpha, \beta) &= \frac{v}{2} [1 - (1-x)\beta_A - x\beta_B] \\ &+ \frac{c}{2} \left[(1-x)\frac{v}{c} - \frac{(1-x)(n-nx-1)}{n-1}\beta_A - \frac{x(1-x)n}{n-1}\beta_B \right] \alpha_A \\ &+ \frac{c}{2} \left[x\frac{v}{c} - \frac{x(1-x)n}{n-1}\beta_A - \frac{x(nx-1)}{n-1}\beta_B \right] \alpha_B. \end{aligned} \quad (4)$$

In addition, the individual's expected payoff can be decomposed into the expected payoff of an individual of type A , ($U_A(\alpha, \beta)$) multiplied by $1-x$, the probability of being of type A and the expected payoffs of an individual of type B ($U_B(\alpha, \beta)$) multiplied by x , the probability of being of type B . That is, $U(\alpha, \beta)$ can be written as

$$U(\alpha, \beta) = (1-x)U_A(\alpha, \beta) + xU_B(\alpha, \beta)$$

where

$$\begin{aligned} U_A(\alpha, \beta) &= \frac{v}{2}(1-\beta_A) + \frac{c}{2}\left(\frac{v}{c}-\beta_A\right)\frac{n(1-x)-1}{n-1}\alpha_A + \frac{c}{2}\left(\frac{v}{c}-\beta_A\right)\frac{nx}{n-1}\alpha_B, \\ U_B(\alpha, \beta) &= \frac{v}{2}(1-\beta_B) + \frac{c}{2}\left(\frac{v}{c}-\beta_B\right)\frac{n(1-x)}{n-1}\alpha_A + \frac{c}{2}\left(\frac{v}{c}-\beta_B\right)\frac{nx-1}{n-1}\alpha_B. \end{aligned} \quad (5)$$

Note the similarity between (2) and (5).

Summarizing, we have modeled a population formed by two types of individuals who play a hawk-dove game. The main characteristic is that individuals fail to perceive their own type but recognize the type of their opponents. The probabilities of the different types of encounter, the strategies and expected payoffs are defined. Thus, we have all the ingredients of a game, hereafter referred to as a *heterogeneous game* and denoted by Γ_x , where $0 < x < 1$ is the proportion of individuals of type B , which is the key parameter in this paper.

We proceed to solve the heterogeneous game Γ_x by applying the concept of evolutionarily stable strategy as for game Γ . Let $\mathcal{B}_x(\beta)$ be the set of an individual's best responses to an opponent playing β . Strategy α^* is *evolutionarily stable* if and only if (i) $\alpha^* \in \mathcal{B}_x(\alpha^*)$, and (ii) for any $\beta \in \mathcal{B}_x(\alpha^*)$ such that $\beta \neq \alpha^*$ we have $U(\alpha^*, \beta) > U(\beta, \beta)$.

3.2 Best responses

First, we determine the set of an individual's best responses given her opponent's strategy. If her opponent plays strategy $\beta = (\beta_A, \beta_B)$ the individual's best response is to choose

$\alpha = (\alpha_A, \alpha_B)$ such that $U(\alpha, \beta)$ is maximized. The set of best responses become clearer if (4) is rewritten as follows

$$U(\alpha, \beta) = f_0(\beta) + f_A(\beta) \alpha_A + f_B(\beta) \alpha_B$$

where

$$\begin{aligned} f_0(\beta) &= \frac{v}{2} [1 - (1-x)\beta_A - x\beta_B] \\ f_A(\beta) &= \frac{c}{2} \left[(1-x)\frac{v}{c} - \frac{(1-x)(n-nx-1)}{n-1}\beta_A - \frac{x(1-x)n}{n-1}\beta_B \right] \\ f_B(\beta) &= \frac{c}{2} \left[x\frac{v}{c} - \frac{x(1-x)n}{n-1}\beta_A - \frac{x(nx-1)}{n-1}\beta_B \right]. \end{aligned} \quad (6)$$

The optimal choice of an individual is $\alpha_A = 1$ whenever $f_A(\beta) > 0$, $\alpha_A = 0$ whenever $f_A(\beta) < 0$ and any α_A whenever $f_A(\beta) = 0$. Similarly the choice of α_B depends on the sign of $f_B(\beta)$. Thus $\mathcal{B}_x(\beta)$, the set of an individual's best responses to an opponent playing β , is given by

$$\mathcal{B}_x(\beta) = \begin{cases} \{(\zeta_A, \zeta_B) \mid \zeta_A, \zeta_B \in [0, 1]\} & \text{if } [f_A(\beta) = 0 \text{ and } f_B(\beta) = 0] \\ \{(\alpha_A, 1) \mid \alpha_A \in [0, 1]\} & \text{if } [f_A(\beta) = 0 \text{ and } f_B(\beta) > 0] \\ \{(0, 1)\} & \text{if } [f_A(\beta) < 0 \text{ and } f_B(\beta) > 0] \\ \{(0, \alpha_B) \mid \alpha_B \in [0, 1]\} & \text{if } [f_A(\beta) < 0 \text{ and } f_B(\beta) = 0] \\ \{(1, \gamma_B) \mid \gamma_B \in [0, 1]\} & \text{if } [f_A(\beta) > 0 \text{ and } f_B(\beta) = 0] \\ \{(1, 0)\} & \text{if } [f_A(\beta) > 0 \text{ and } f_B(\beta) < 0] \\ \{(\gamma_A, 0) \mid \gamma_A \in [0, 1]\} & \text{if } [f_A(\beta) = 0 \text{ and } f_B(\beta) < 0] \\ \{(1, 1)\} & \text{if } [f_A(\beta) > 0 \text{ and } f_B(\beta) > 0] \\ \{(0, 0)\} & \text{if } [f_A(\beta) < 0 \text{ and } f_B(\beta) < 0]. \end{cases} \quad (7)$$

Second, we determine the strategies that are best responses to themselves. As will be proven in the next theorem only three strategies satisfy this property. The first is independent of x , whereas for the other two the probability of hawkish behavior depends on x . These strategies denoted respectively as $\frac{v}{c}$, α_x^* and γ_x^* are given by

$$\begin{aligned} \frac{v}{c} &= \left(\frac{v}{c}, \frac{v}{c} \right) \\ \alpha_x^* &= \begin{cases} (\alpha_A^*(x), 1) & \text{if } x \leq \underline{w} \\ (0, 1) & \text{if } \underline{w} < x < \bar{w} \\ (0, \alpha_B^*(x)) & \text{if } x \geq \bar{w} \end{cases} \end{aligned} \quad (8)$$

$$\gamma_x^* = \begin{cases} (\gamma_A^*(x), 0) & \text{if } x \leq 1 - \bar{w} \\ (1, 0) & \text{if } 1 - \bar{w} < x < 1 - \underline{w} \\ (1, \gamma_B^*(x)) & \text{if } x \geq 1 - \underline{w}, \end{cases} \quad (9)$$

where

$$\begin{aligned}\alpha_A^*(x) &= \frac{v}{c} - \left(1 - \frac{v}{c}\right) \frac{nx}{n-nx-1}, & \alpha_B^*(x) &= \frac{n-1}{nx-1} \frac{v}{c}, \\ \gamma_A^*(x) &= \frac{n-1}{n(1-x)-1} \frac{v}{c}, & \gamma_B^*(x) &= \frac{v}{c} - \left(1 - \frac{v}{c}\right) \frac{n(1-x)}{n-n(1-x)-1}, \\ \underline{w} &= \frac{v}{c} \left(1 - \frac{1}{n}\right), & \text{and } \bar{w} &= \frac{v}{c} \left(1 - \frac{1}{n}\right) + \frac{1}{n}.\end{aligned}$$

We can now state and prove the following result:

Theorem 1 *In any heterogeneous game Γ_x , only strategies $\frac{v}{c}$, α_x^* and γ_x^* are best responses to themselves.*

Proof. Using (6) and (7) it can be checked that these three strategies satisfy $\beta \in \mathcal{B}_x(\beta)$ and that no other strategy does.

1. Strategy $\frac{v}{c} \in \mathcal{B}_x(\frac{v}{c})$ since $f_A((\zeta_A, \zeta_B)) = f_B((\zeta_A, \zeta_B)) = 0$ iff $\zeta_A = \zeta_B = \frac{v}{c}$.
2. For strategy α_x^* , we first check that the following holds: $f_A((\alpha_A, 1)) = 0$ iff $\alpha_A = \alpha_A^*(x)$ with $\alpha_A^*(x) \geq 0$ if $x \leq \underline{w}$, and $f_B((\alpha_A^*(x), 1)) > 0$. Second, we have $f_A((0, 1)) < 0$ if $x > \underline{w}$ and $f_B((0, 1)) > 0$ if $x < \bar{w}$. Third, we have $f_B((0, \alpha_B^*(x))) = 0$ iff $\alpha_B = \alpha_B^*(x)$, with $\alpha_B^*(x) \leq 1$ if $x \geq \bar{w}$ and $f_A((0, \alpha_B^*(x))) < 0$. Hence $\alpha_x^* \in \mathcal{B}_x(\alpha_x^*)$.
3. For strategy γ_x^* , we first check that $f_A((\gamma_A, 0)) = 0$ iff $\gamma_A = \gamma_A^*(x)$ with $\gamma_A^*(x) \leq 1$ if $x \leq 1 - \bar{w}$, and $f_B((\gamma_A^*(x), 0)) < 0$. Second, we have $f_A((1, 0)) > 0$ if $x > 1 - \bar{w}$ while $f_B((1, 0)) < 0$ if $x < 1 - \underline{w}$. Third, $f_B((1, \gamma_B)) = 0$ iff $\gamma_B = \gamma_B^*(x)$, with $\gamma_B^*(x) \geq 0$ if $x \geq 1 - \underline{w}$ and $f_A((0, \gamma_B^*(x))) > 0$. Hence $\gamma_x^* \in \mathcal{B}_x(\gamma_x^*)$.
4. It remains to show that no other strategy can be a best response to itself. This is done by checking that $f_A((1, 1)) < 0$, and $f_A((0, 0)) > 0$.

■

Thus, we have one homogeneous and two heterogeneous strategies that are best responses to themselves. The homogeneous strategy $\frac{v}{c}$ consists of playing the evolutionarily strategy $\frac{v}{c}$ of the homogeneous game toward an individual of any type. Note that the strategy is adopted by the individuals of two types. Obviously it could not be otherwise since in our game individuals do not know their own type.

3.3 Evolutionarily stable strategies

The next question is whether the strategies that are best responses to themselves are evolutionarily stable. As the following result shows, this is not the case for strategy $\frac{v}{c}$ (where individuals play hawk with probability $\frac{v}{c}$ against any type of opponent).

Theorem 2 *In any heterogeneous game Γ_x , strategy $\frac{v}{c}$ is not evolutionarily stable.*

Proof. It is immediately apparent that $f_A(\frac{v}{c}) = f_B(\frac{v}{c}) = 0$, and by (7) we know that any strategy is a best response to strategy $\frac{v}{c}$. To show that $\frac{v}{c}$ is not evolutionarily stable, we choose a strategy β such that the difference $U(\frac{v}{c}, \beta) - U(\beta, \beta)$ is negative. Using (4) this difference can be written as

$$U(\frac{v}{c}, \beta) - U(\beta, \beta) = \frac{c}{2(n-1)} \left[(\frac{v}{c} - \beta_A)^2 (n - nx - 1)(1 - x) + 2(\frac{v}{c} - \beta_A)(\frac{v}{c} - \beta_B)n(1 - x)x + (\frac{v}{c} - \beta_B)^2 (nx - 1)x \right].$$

(i) If $x = \frac{n-1}{n}$ the above difference is reduced to

$$U(\frac{v}{c}, \beta) - U(\beta, \beta) = \frac{c}{2n} (\frac{v}{c} - \beta_B) \left[2(\frac{v}{c} - \beta_A) + (\frac{v}{c} - \beta_B)(n - 2) \right]$$

which turns out to be negative if strategy β is chosen such that

$$0 < \beta_B < \frac{v}{c} \text{ and } \frac{v}{c} < \beta_A < 1 \text{ so that } \beta_A - \frac{v}{c} > \frac{n-2}{2} (\frac{v}{c} - \beta_B). \quad (10)$$

(ii) If $x \neq \frac{n-1}{n}$, denote $Z = \frac{\frac{v}{c} - \beta_A}{\beta_B - \frac{v}{c}}$. In this case the difference is a quadratic equation in Z :

$$U(\frac{v}{c}, \beta) - U(\beta, \beta) = \frac{c}{2(n-1)} (\beta_B - \frac{v}{c})^2 \left[(n - nx - 1)(1 - x)Z^2 - 2n(1 - x)xZ + (nx - 1)x \right]$$

whose discriminant is $\Delta = 4(1 - x)x(n - 1) > 0$. Thus, the difference under study is negative for any Z such that $\frac{2n(1-x)x - \sqrt{\Delta}}{2(n-nx-1)(1-x)} < Z < \frac{2n(1-x)x + \sqrt{\Delta}}{2(n-nx-1)(1-x)}$ and, in particular, for a strategy $\beta = (\beta_A, \beta_B)$ such that the following equality is satisfied:

$$\frac{\frac{v}{c} - \beta_A}{\beta_B - \frac{v}{c}} = \frac{nx}{n - nx - 1}. \quad (11)$$

■

Thus, while playing hawk with probability $\frac{v}{c}$ is evolutionarily stable in the homogeneous game, doing so against any type of opponent is not evolutionarily stable in the heterogeneous game. At first sight, this result may appear surprising as the differentiation introduced within the members of the population is merely artificial. But from the proof of Theorem 2 it can be easily understood why strategy $\frac{v}{c}$ is not evolutionarily stable. The strategies β that satisfy

(10) or (11) perform better against $\frac{v}{c}$ than strategy $\frac{v}{c}$ does against itself. These strategies are such that the probability of aggressive behavior toward one type is higher than $\frac{v}{c}$ while the probability of aggressive behavior toward the other type is smaller than $\frac{v}{c}$. The observation of the opponent's type gives a signal that players use.

By contrast the result for the other two strategies that are best responses is the following.

Theorem 3 *In any heterogeneous game Γ_x , only strategies α_x^* and γ_x^* are evolutionarily stable.*

Proof. By Theorem 1 we know that the only strategies that are best responses to themselves are α_x^* , γ_x^* and $\frac{v}{c}$ and by Theorem 2 we know that $\frac{v}{c}$ is not evolutionarily stable. Hence, it remains only to analyze strategies α_x^* and γ_x^* .

Using (7) the set of an individual's best responses to an opponent playing α_x^* is

$$\mathcal{B}_x(\alpha_x^*) = \begin{cases} \{(\alpha_A, 1) \mid \alpha_A \in [0, 1]\} & \text{if } x \leq \underline{w} \\ \{(0, 1)\} & \text{if } \underline{w} < x < \bar{w} \\ \{(0, \alpha_B) \mid \alpha_B \in [0, 1]\} & \text{if } \bar{w} \leq x. \end{cases}$$

For $\underline{w} < x < \bar{w}$ we have that $\alpha_x^* = (0, 1)$ is the *only* best response to itself, hence Condition (ii) of the evolutionarily stable strategy definition becomes empty. For the other values of x , however, it must be checked whether the difference $U(\alpha_x^*, \beta) - U(\beta, \beta)$ is strictly positive for any $\beta \in \mathcal{B}_x(\alpha_x^*)$ where $\beta \neq \alpha_x^*$. Using (4) we obtain that

$$U(\alpha_x^*, \beta) - U(\beta, \beta) = \begin{cases} \frac{c}{2} \frac{(1-x)(n-nx-1)}{n-1} (\alpha_A^*(x) - \alpha_A)^2 & \text{if } x \leq \underline{w} \\ \frac{c}{2} \frac{x(nx-1)}{n-1} (\alpha_B^*(x) - \alpha_B)^2 & \text{if } \bar{w} \leq x \end{cases}$$

Since this difference is strictly positive for any $\beta \neq \alpha_x^*$ then strategy α_x^* is evolutionarily stable.

The proof that strategy γ_x^* is evolutionarily stable is omitted because it is similar to the previous one. ■

So in general evolutionarily stable strategies are mixed strategies, although they may be pure strategies for some specific proportions of individuals of type B . This is the case when x is equal to $\frac{v}{c}$, that is for the game $\Gamma_{v/c}$. For this game an interesting comparison between $\frac{v}{c}$, the evolutionarily stable strategy in game Γ , and $(0, 1)$, the evolutionarily stable strategy in game $\Gamma_{v/c}$ can be made. In game $\Gamma_{v/c}$ an individual who plays strategy $(0, 1)$ behaves as a hawk whenever her opponent is of type B , which occurs with frequency $\frac{v}{c}$. Therefore strategies $\frac{v}{c}$ in Γ and $(0, 1)$ in $\Gamma_{v/c}$ are "similar" in the sense that a probability in game Γ is substituted by a frequency in game $\Gamma_{v/c}$. Furthermore strategy $(0, 1)$ is also evolutionarily stable in game Γ_x for values of x close to $\frac{v}{c}$, *i.e.* $\underline{w} < x < \bar{w}$. Note, however, that for values

of x smaller than \underline{w} , an individual that plays strategy $(0, 1)$ plays dove so often that her opponent is better off playing hawk at probabilities greater than $\frac{v}{c}$ while the opposite occurs for values of x higher than \bar{w} . Analogously an individual who plays strategy $(1, 0)$ in game $\Gamma_{1-v/c}$ plays hawk with a frequency of $1 - x = v/c$. The same comparison can be made between $\frac{v}{c}$ in game Γ , and strategy $(1, 0)$ in game $\Gamma_{1-v/c}$.

Now let us move on to the expected payoffs for the evolutionarily stable strategies starting with α_x^* . Given game Γ_x , plugging (8) into (4) we obtain

$$U(\alpha_x^*, \alpha_x^*) = \begin{cases} \frac{v}{2}(1 - \frac{v}{c}) + \frac{c^2 - v^2}{2c} \frac{x}{n - nx - 1} & \text{if } x \leq \underline{w} \\ \frac{v}{2} - \frac{nx - 1}{n - 1} \frac{cx}{2} & \text{if } \underline{w} < x < \bar{w} \\ \frac{v}{2}(1 - \frac{v}{c}) - \frac{v^2}{2c} \frac{1 - x}{nx - 1} & \text{if } x \geq \bar{w}. \end{cases}$$

Observe that an individual's payoff is larger for smaller proportions than for large proportions of individuals of type B . In the next section we come back to this result and explain why this is so.

If the payoff is decomposed according to the type of individuals we have that (5) yields to

$$U_A(\alpha_x^*, \alpha_x^*) = \begin{cases} \frac{v}{2}(1 - \frac{v}{c}) + \frac{c - v}{c} \frac{vnx}{n - nx - 1} & \text{if } x \leq \underline{w} \\ \frac{v}{2}(1 - \frac{v}{c}) + \frac{v}{2c} \frac{v(n - 1) + cnx}{n - 1} & \text{if } \underline{w} < x < \bar{w} \\ \frac{v}{2}(1 - \frac{v}{c}) + \frac{v^2}{2c} \frac{2nx - 1}{nx - 1} & \text{if } x \geq \bar{w}, \end{cases}$$

and

$$U_B(\alpha_x^*, \alpha_x^*) = \begin{cases} \frac{v}{2}(1 - \frac{v}{c}) - \frac{c - v}{2c} \frac{2nv(1 - x) - c - v}{n - nx - 1} & \text{if } x \leq \underline{w} \\ \frac{v}{2}(1 - \frac{v}{c}) - \frac{c - v}{2c} \frac{(n - 1)v + (nx - 1)c}{n - 1} & \text{if } \underline{w} < x < \bar{w} \\ \frac{v}{2}(1 - \frac{v}{c}) - \frac{v^2}{c} \frac{n - nx}{nx - 1} & \text{if } x \geq \bar{w}. \end{cases}$$

Clearly an individual of type A always obtains a larger payoff than an individual of type B . The reverse holds in the second evolutionarily stable strategy. Moreover, by plugging (9) into (4) and (5) it can be checked that the following relations hold:

$$U(\gamma_x^*, \gamma_x^*) = U(\alpha_{1-x}^*, \alpha_{1-x}^*) \tag{12}$$

$$U_A(\gamma_x^*, \gamma_x^*) = U_B(\alpha_{1-x}^*, \alpha_{1-x}^*) \tag{13}$$

$$U_B(\gamma_x^*, \gamma_x^*) = U_A(\alpha_{1-x}^*, \alpha_{1-x}^*). \tag{14}$$

3.4 Interpretation of the results in terms of aliens and locals

Let us start by examining strategy α_x^* , given by (8). Note that, whatever x may be, the probability of aggression toward individuals of type B is always greater than toward individuals of type A . The reverse holds for strategy γ_x^* given by (9). So the individuals of type B

are treated worse in strategy α_x^* while the individuals of type A are treated worse in strategy γ_x^* . We refer to the better-treated individuals as *locals* and to the worse-treated individuals as *aliens*. A closer look at (8) and (9) reveals the relationship between the two evolutionarily stable strategies. To clarify this, define

$$\begin{aligned} g_{\text{I}}(y) &= \begin{cases} \frac{v}{c} - (1 - \frac{v}{c})\frac{ny}{n-ny-1} & \text{if } y \leq \underline{w} \\ 0 & \text{otherwise} \end{cases} \\ g_{\text{II}}(y) &= \begin{cases} 1 & \text{if } y < \bar{w} \\ \frac{n-1}{ny-1} \frac{v}{c} & \text{otherwise,} \end{cases} \end{aligned} \quad (15)$$

which allows (8) and (9) to be rewritten as $\alpha_x^* = (g_{\text{I}}(x), g_{\text{II}}(x))$ and $\gamma_x^* = (g_{\text{II}}(1-x), g_{\text{I}}(1-x))$ clearly showing that the two strategies are the two faces of a single coin. Moreover, if these probabilities are expressed as a function of the proportion of aliens (denoted by y), $g_{\text{I}}(y)$ can be interpreted as the probability of hawkish behavior toward a local and $g_{\text{II}}(y)$ as the probability of hawkish behavior toward an alien. To illustrate the evolution of these probabilities with an example, in *Figure 1* we plot $g_{\text{II}}(y)$ and $g_{\text{I}}(y)$ for $n = 10$ and $v/c = 1/3$.

Figure 1 about here

The following conclusions can be derived on the trend in the probability of aggression. The probability of aggression toward aliens is 1 for y smaller than \bar{w} , and then it decreases toward $\frac{v}{c}$. The probability of aggression toward locals, however, decreases from $\frac{v}{c}$ to 0 for y smaller than \underline{w} and remains at 0 thereafter. Furthermore, aggression toward aliens is always greater than the aggression suffered by individuals in a homogeneous game ($g_{\text{II}}(y) > \frac{v}{c}$) while the reverse holds for locals ($g_{\text{I}}(y) < \frac{v}{c}$).

Similarly we may wonder what happens to the trend in payoffs assuming that individuals play an evolutionarily stable strategy. The payoffs in equality (12) can be written as a function of y , which we denote by $U^*(y)$, so that $U^*(y) = U(\alpha_y^*, \alpha_y^*)$.

To facilitate the interpretation of the trend in $U^*(y)$ we take as a reference the equilibrium payoff obtained in the homogeneous game Γ which we denote by \bar{U}^* . From (2) we determine $\bar{U}^* = u(\frac{v}{c}, \frac{v}{c}) = \frac{v}{2}(1 - \frac{v}{c})$. In *Figure 2* we graphically represent $U^*(y)$ and \bar{U}^* for $n = 10$, $v/c = 1/3$.

Figure 2 about here.

This figure shows that the introduction of aliens proves beneficial for individuals as long as y is smaller than \underline{w} . For y greater than \bar{w} the contrary effect arises. The maximal payoff is obtained for y equal to \underline{w} .

We can also study the trend in the payoff depending on the type of the individuals. Once more, equality (13) allows us to write *the payoff of a local*, $U_I^*(y)$, as a function of y so that $U_I^*(y) = U_A(\alpha_y^*, \alpha_y^*)$. Similarly equality (14) allows us to write *the payoff of an alien* $U_{II}^*(y) = U_B(\alpha_y^*, \alpha_y^*)$. In *Figure 3* we plot $U_I^*(y)$ and \bar{U}^* while in *Figure 4* we plot $U_{II}^*(y)$ and \bar{U}^* for $n = 10$ and $v/c = 1/3$.

Figure 3: about here.

Figure 4: about here.

At first glance these figures reveal that the situation is as expected: aliens are worse off than individuals in a homogeneous game, while locals are better off. Furthermore the maximal payoff for a local is obtained for a proportion of aliens of \bar{w} . A closer examination enables the trend in these payoffs according to the proportion of aliens to be explained. Two effects may be identified.

First, the less aggression suffered, the greater the well-being. This result partially explains the trend in the payoffs. *Figure 1* shows that $g_I(y)$ decreases for values of y smaller than \underline{w} and $g_{II}(y)$ decreases for values of y larger than \bar{w} . Consequently for values of y smaller than \underline{w} a local's payoff increases, and for values of y larger than \bar{w} an alien's payoff also increases (see *Figures 3* and *4*). This increase in payoffs for the same intervals is also observed for an individual (see *Figure 2*).

Second, when the aggression is constant, the trend in the payoff may be explained by the response to the level of aggression. *Figure 1* shows that function $g_I(y)$ is 0 for values of y greater than \underline{w} , meaning that locals suffer no aggression. In that case their best response is to play hawk. But for y greater than \bar{w} they play hawk less often ($g_{II}(y)$ decreases and $g_I(y) = 0$). Therefore in this interval the payoff of locals decreases. Analogously function $g_{II}(y)$ is 1 for values of y smaller than \bar{w} , meaning that aliens suffer maximum aggression. In that case their best response is to play dove. For y smaller than \underline{w} they play hawk less often ($g_I(y)$ decreases and $g_{II}(y) = 1$). Therefore in this interval the payoff of aliens increases.

It remains to explain the trend in the payoffs for values of y lying between \underline{w} and \bar{w} where locals suffer no aggression and aliens suffer maximum aggression. Here the best response for a local is to play hawk and for an alien it is to play dove. As y grows, the hawkish behavior

increases, which turns out to be beneficial for locals and harmful for aliens. For an "average" individual, however, the overall effect is negative.

The following proposition summarizes the main features that are observed in the foregoing figures. The proof is omitted because of its simplicity.

Proposition 1 *For any proportion of aliens $0 < y < 1$ we have that:*

$$\begin{aligned}
 (i) \quad & U_{II}^*(y) < \bar{U}^* < U_I^*(y), \\
 (ii) \quad & U^*(y) > \bar{U}^* \text{ if } y < \underline{w} \quad \text{and} \quad U^*(y) < \bar{U}^* \text{ if } y > \bar{w}, \\
 (iii) \quad & U_I^*(y) \text{ is maximal for } y = \bar{w} \quad \text{and} \quad U^*(y) \text{ is maximal for } y = \underline{w}.
 \end{aligned} \tag{16}$$

4 The experiment

Dennis *et al.* (2008) have conducted several experiments with groups of domestic fowls. They consider group sizes of 10 and 50 birds in which different proportions (20%, 50% and 100% respectively) are marked. They study the birds' aggressive behavior measured by the number of pecks and threats in the encounters between them. The most significant results of this experiment are: (i) Marked domestic fowls receive more pecks than their unmarked pen mates. (ii) Marked domestic fowls in the 20% group receive significantly more threats than domestic fowls in the 100% marked group. (iii) There is no significant difference in the aggression suffered by marked fowls in the 20 and 50% marked groups. (iv) Aggressiveness toward marked fowls in populations with 100% of marked birds is lower than in any mixed population. (v) Marked fowls have a lower body mass than their unmarked pen mates.

To evaluate these experimental results in the light of our model we assume that behaving as a hawk is a good proxy for the pecking and threatening between birds observed in the experiment. We also assume that the strategy played by the population is the evolutionarily stable strategy where the aliens are the marked fowls. In addition we consider that the expected payoff of an individual can be used as a proxy for a bird's body mass. With these assumptions we find that some of our theoretical results are consistent with the experimental ones:

(i) *Marked domestic fowls receive more pecks than their unmarked pen mates.* By (15) we have $g_I(y) < g_{II}(y)$.

(ii) *Marked domestic fowls in the 20% group receive significantly more threats than domestic fowls in the 100% marked group.* By (15) we have $g_{II}(0.2) = 5 \frac{(n-1)}{n-5} \frac{v}{c} > \frac{v}{c}$.

(iii) *There is no significant difference in the aggression suffered by marked birds in the 20 and 50% marked groups.* By (15) we have $g_{II}(0.5) = g_{II}(0.2) = 1$ if $0.5 \leq \frac{v}{c} (1 - \frac{1}{n})$. Therefore

this empirical result is supported by our theoretical findings under the assumption that the value of the resource v is basically greater than half the cost c .

(iv) *Aggressiveness toward marked birds in populations with 100% of marked birds is lower than any other mixed population.* By (15) we have that for any $0 < y < 1$, $g_{\text{II}}(y) > \frac{v}{c}$.

(v) *Marked birds have a lower body mass than their unmarked pen mates.* By (16) we have $U_{\text{II}}^*(y) < U_{\text{I}}^*(y)$: the payoff of a marked bird is smaller than the payoff of an unmarked one.

5 Concluding comments

The contribution of this paper can be summarized as follows: We introduce a variation of the hawk-dove game in which there is a population formed by two types of individual who do not perceive their own type but do recognize the type of their opponent. Although the difference between the two types is "artificial" it is not innocuous. Our game has two evolutionarily stable strategies in which the probability of being aggressive toward one type of individual is always higher than the probability of being aggressive toward the other type. It is worth stressing that the type of individual treated worse may not be the minority group. This contradicts the intuition according to which the type which constitutes the minority of the population seems likely to be discriminated against. The probability of aggression toward aliens does however decrease with the proportion of aliens. Increasing the proportion of aliens also decreases the probability of aggressive behavior toward locals. For a random individual the effect is positive for small proportion of aliens, and negative for large proportions.

In conclusion we would like to point out that although our research was inspired by a biological experiment, the approach followed in this paper might also serve to explain other social situations. First, the behavior of the individuals in our setting could be interpreted as if they were acting behind the veil of ignorance. It is known that individuals under this device conscientiously ignore their position in society with the objective of reaching a more "just" society. This idea transferred to game theory could be interpreted as if every agent when playing a strategic game ignores who she is while she is aware of the type of opponent is playing with. Second, our work can also be related to the classic work developed by Cass and Schell (1983) who show that extrinsic uncertainty may play a role in rational expectations equilibrium models. For these authors a variable is intrinsic whenever it has an affect on the fundamentals otherwise it is extrinsic. In our paper the random division of the population can be considered as an extrinsic variable that matters since it changes the results of the hawk-dove in homogenous populations. Finally, it should be emphasized that our results are

similar to those obtained in the paper by Axtell et al. (1991) in which an artificial division of a group of individuals into two subgroups generates real discrimination.

References

- [1] Axtell, R. L., J. M. Epstein and H. P. Young (2001) The Emergence of Classes in a Multi-Agent Bargaining Model. *Social Dynamics* edited by S. N. Durlauf and H. P. Young.
- [2] Binmore, K., and L. Samuelson (2001a), Evolution and Mixed Strategies. *Games and Economic Behavior* **34**, 200–226.
- [3] Binmore, K., and L. Samuelson (2001b), Can Mixed Strategies be Stable in Asymmetric Games. *Journal of Theoretical Biology* **210**, 1–14.
- [4] Cass, D., and K. Shell, (1983), Do Sunspots Matter? *The Journal of Political Economy* **91(2)**, 193–227.
- [5] Dennis, R. L., R. C. Newberry, H. W. Cheng, and I. Estevez, (2008), Appearance Matters: Artificial Marking Alters Aggression and Stress, *Poultry Science* **87**, 1939–1946.
- [6] Maynard Smith, J., (1982), *Evolution and the theory of games*, Cambridge University Press; Cambridge, UK.
- [7] Maynard Smith, J., and G. R. Price, (1973), The Logic of Animal Conflict, *Nature* **246**, 15–18.
- [8] Maynard Smith, J. and G. A. Parker (1976), The Logic of Asymmetric Contests. *Animal Behavior* **24**, 159–75.
- [9] Nash, J. F., (1951), Non-Cooperative Games, *Annals of Mathematics* **54**, 286–295.
- [10] Osborne, M. J., (2004), *An Introduction to Game Theory*, Oxford University Press; New York, USA.
- [11] Saraux, C., C. Le Bohec, J. M. Durant, V. A. Viblanc, M. Gauthier-Clerc, D. Beaune, Y-H. Park, N. G. Yoccoz, N. C. Stenseth and Y. Le Maho, (2011), Reliability of flipper-banded penguins as indicators of climate change. *Nature* **469**, 203–206.
- [12] Selten, R., (1980), A note on Evolutionarily Stable Strategies in Asymmetric Animal Contests, *Journal of Theoretical Biology* **84**, 93–101.

- [13] Weibull, J. W., (1995), *Evolutionary Game Theory*, MIT Press; Cambridge, Massachusetts, USA.

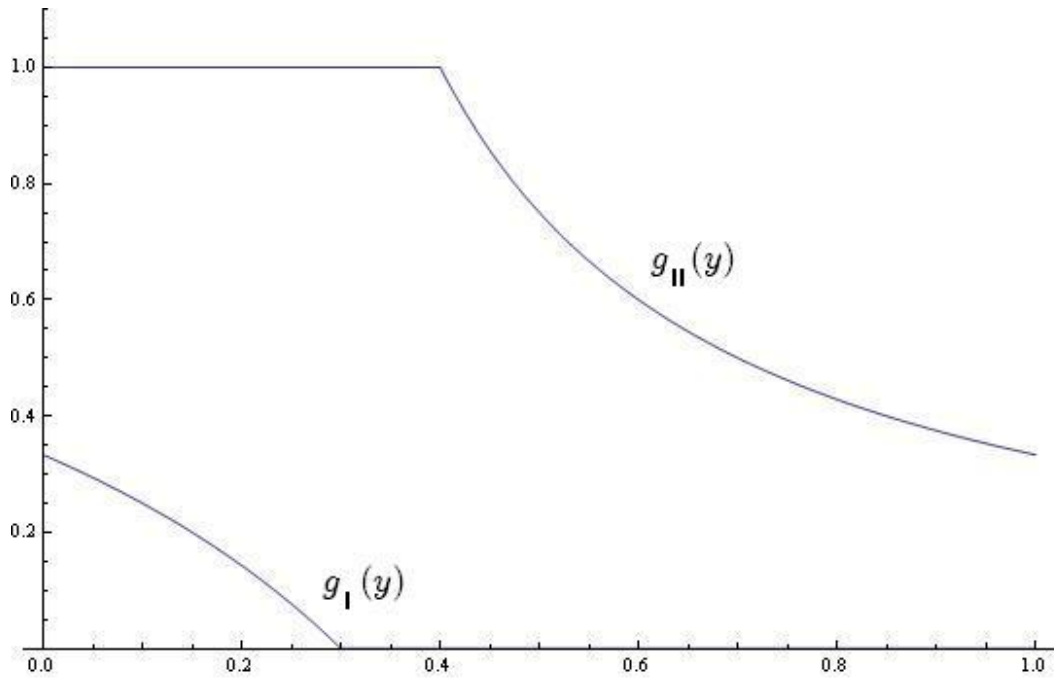


Figure 1: Probabilities of a hawk behavior toward local (g_I) and toward aliens (g_{II}) as a function of the proportion of aliens (y)

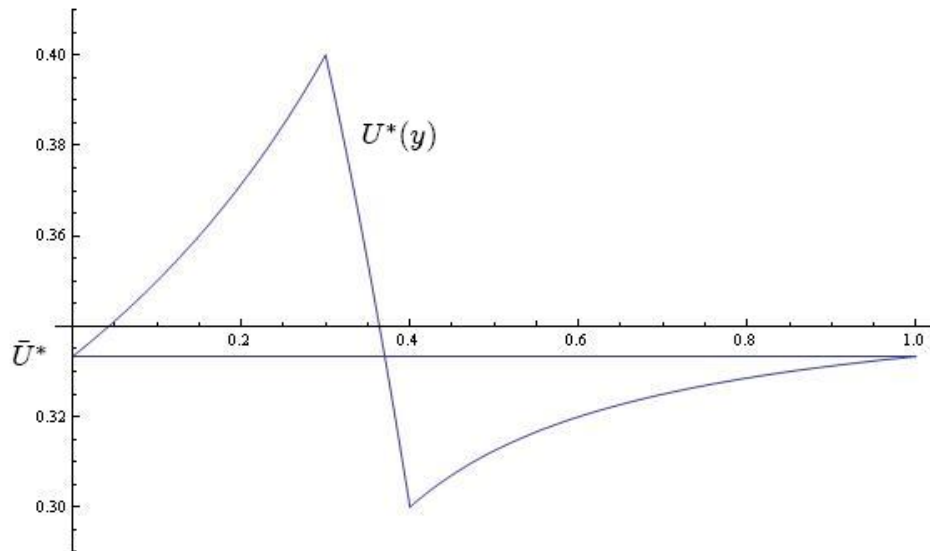


Figure 2: Payoff in the homogeneous game (\bar{U}^*) and payoff of a random individual in the heterogeneous game ($U^*(y)$) as a function of the proportion of aliens (y)

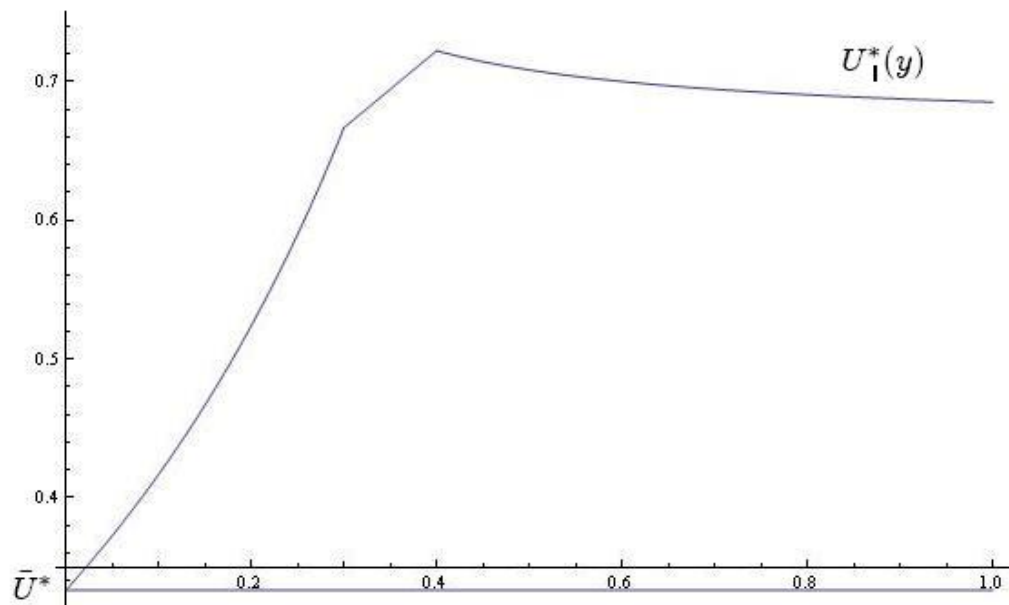


Figure 3: Payoff in the homogeneous game (\bar{U}^*) and payoff of a local in the heterogeneous game ($U_I^*(y)$) as a function of the proportion of aliens (y)

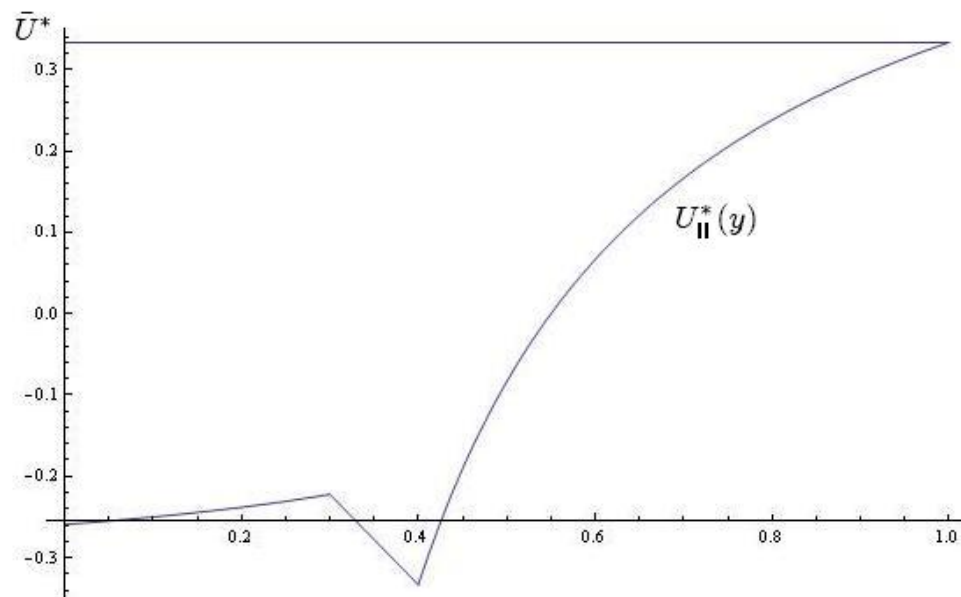


Figure 4: Payoff in the homogeneous game (\bar{U}^*) and payoff of an alien in the heterogeneous game ($U_{II}^*(y)$) as a function of the proportion of aliens (y)