

Efficient social contracts and group selection

Simon M. Huttegger · Rory Smead

Received: 16 November 2010 / Accepted: 17 March 2011
© Springer Science+Business Media B.V. 2011

Abstract We consider the Stag Hunt in terms of Maynard Smith's famous Haystack model. In the Stag Hunt, contrary to the Prisoner's Dilemma, there is a cooperative equilibrium besides the equilibrium where every player defects. This implies that in the Haystack model, where a population is partitioned into groups, groups playing the cooperative equilibrium tend to grow faster than those at the non-cooperative equilibrium. We determine under what conditions this leads to the takeover of the population by cooperators. Moreover, we compare our results to the case of an unstructured population and to the case of the Prisoner's Dilemma. Finally, we point to some implications our findings have for three distinct ideas: Ken Binmore's group selection argument in favor of the evolution of efficient social contracts, Sewall Wright's Shifting Balance theory, and the equilibrium selection problem of game theory.

Keywords Cooperation · Equilibrium selection · Group selection · Haystack model · Prisoner's Dilemma · Shifting balance · Stag Hunt

Introduction

The problem of the evolution of cooperation has many facets. The traditional focus is on explaining how it could be possible that selfish behavior does not undermine altruistic behavior. Within game theory the corresponding basic problem can be stated in terms of the famous Prisoner's Dilemma game. In the Prisoner's Dilemma,

S. M. Huttegger
Department of Logic and Philosophy of Science, University of California, Irvine, Irvine, CA, USA
e-mail: shuttegg@uci.edu

R. Smead (✉)
Department of Philosophy and Religion, Northeastern University, Boston, MA, USA
e-mail: r.smead@neu.edu

defection (selfishness) strictly dominates cooperation (altruism). Cooperation, in this sense, is not a feasible evolutionary outcome for most standard models. One way of explaining the evolution of altruism is by postulating the effects of group-structure on evolution. This line of research is well known as the group selection controversy (see Maynard Smith 1976, and references therein; see Sober and Wilson 1998, and references therein; see Okasha 2009, and references therein).

One can, however, understand cooperation more generally as social collaboration. In this case, there are interactions other than the altruism-selfishness dichotomy that are relevant to investigating various aspects of cooperation (Skyrms 2004; Calcott 2008). One such aspect is the evolution of efficient rather than non-efficient population states.

Throughout this paper, we consider the Stag Hunt game. The basic idea of this game goes back to Rousseau. It is a simple representation of the conflict between risk and efficiency in social interactions. To illustrate this game, imagine two hunters who must decide independently whether or not to hunt stag (cooperate) or to hunt hare (defect). The stag provides a very good payoff, but can only be hunted successfully if both hunters cooperate. Hunting hare, on the other hand, ensures a moderate gain under any circumstance. Hunters that are expecting others to defect will want to defect, but hunters that expect others to cooperate will want to cooperate. This game of cooperation is different from the Prisoner's Dilemma: defection is stable, but not dominant. Cooperation is the efficient outcome, but defection is less risky.

From the standard evolutionary perspective, both cooperation and defection are possible in the Stag Hunt. If there are enough cooperators in the population, cooperation will evolve. Otherwise, defection will evolve. Under traditional assumptions, defection is often the more likely evolutionary outcome because it is less risky. This leads to questions about cooperation that go beyond stability concerns. Under what circumstances can we expect the efficient outcome to evolve? And, how could a population move from an inefficient state to a more efficient state?

Much of the group selection debate has focused on when it is possible for altruism (usually unstable) to evolve. However, these ideas may be relevant to addressing questions about the evolution of efficient (but stable) states as well. If different groups reach different social states in games like the Stag Hunt, some will be more efficient than others. What impact might this have on the evolution of efficient social states in general? The aim of this paper is to formulate a simple model that will allow us to examine the effect of group selection on the evolution of efficient social states. We find that even in very simple settings the role of group-structure on efficiency is not straightforward. Using Maynard Smith's famous Haystack model, we delineate when group selection helps populations reach efficient states and when it works against efficiency.¹

Questions about the evolution of efficient states are of particular importance in social philosophy for determining when we should expect social contracts to be

¹ It is important to note at the outset that we will be considering a model of group selection that is completely Hamiltonian in spirit. We are not considering the cases where groups are treated as cohesive units that are in direct competition, but rather the effect of a certain kind of group-structure in the population (with correlated interactions) on the selection of individual social strategies.

efficient (Binmore 2005). They are also relevant to issues in evolutionary biology and, in particular, to how populations may move from one fitness peak to another (e.g. Sewall Wright's shifting balance theory; see Wright 1932). Furthermore, understanding how group selection can affect the evolution of efficient outcomes will provide some insight to the equilibrium selection problem in game theory (see e.g., Samuelson 1998). We will explain these background issues in more detail in the section “Background”.

Background

The emergence of efficient outcomes is considered in a number of different fields. In this section we briefly describe how this problem is framed in game theory, social philosophy, and evolutionary biology. We return to these issues at the end of our paper, after having established our main results.

Equilibrium selection problem

Most games have more than one Nash equilibrium. Which of them will be chosen by the players? This is the so-called equilibrium selection problem in game theory (see e.g., Samuelson 1998). In evolutionary game theory, this question is answered by appealing to some evolutionary process. Such processes often do not choose the efficient Nash equilibrium. As an example, consider the Stag Hunt as specified in Fig. 1. There are two strict Nash equilibria, one where both players cooperate and one where both players defect. The first one is Pareto efficient, i.e. no change in the players' choices can be made that would result in a higher payoff to at least one of the players. The other one is risk dominant. In the simple case of the Stag Hunt, this means that defection yields a expected higher payoff when all opponent actions are considered equally likely.² Many evolutionary processes that have been considered in the literature choose the risk dominant over the Pareto efficient Nash equilibrium (e.g., Kandori et al. 1993; Young 1993). Considering the effect of group selection might show how the Pareto efficient equilibrium can be chosen over the risk dominant one.

Social philosophy

The emergence of efficient outcomes is also an important element in Ken Binmore's *Natural Justice* (2005). Binmore describes three “levels of priority” in understanding the evolution of social contracts: stability, efficiency and fairness. His primary aim is to understand fairness of contracts, but he is careful not to neglect the role of stability and efficiency: “A social contract must be internally stable, or it won't survive. It needs to be efficient, or it won't compete successfully with the social contracts of other societies” (5).

² In games with more strategies, risk dominance has to be defined differently.

Fig. 1 The Stag Hunt.
 $a > b \geq 0$, $b + 1 > a$

	c	d
c	$1 + a$	0
d	$1 + b$	1

The importance of stability leads Binmore to conceive of social contracts as the equilibria of some game. Given the equilibrium selection problem in game theory, there is a parallel problem of selecting social contracts. In response to this problem, Binmore argues that (in some circumstances) evolution will favor efficiency due to competition among groups and that efficient contracts will “eventually come to predominate”:

Why should we expect evolution to succeed in selecting one of the efficient equilibria rather than one of the many inefficient alternatives?... The answer postulates competition among groups... Societies operating [an efficient social contract] will therefore grow faster. Assuming societies cope with population growth by splitting off colonies which inherit the social contract of the parent society, we will then eventually observe large numbers of copies of societies operating [an efficient social contract] compared with those operating [an inefficient contract] (12).

Since social contracts are assumed to be stable here, the argument does not commit the group selection fallacy. If the societies were playing a game like the Prisoner’s Dilemma, the efficient cooperating outcome could be subverted by defecting free-riders. Hence, those outcomes would be unstable. However, societies that are playing stable equilibria are immune to subversion from within.

This is an interesting and suggestive idea about how efficient social contracts might develop. However, Binmore does not develop this argument in much more detail, leaving questions about its scope and applicability. For instance, it is not clear how often such conditions are met nor how frequently this setting will lead to the “predominance” of efficient contracts.³ For these reasons, it would be desirable to have a simple formal model resembling this setting which can be explored in more detail. Binmore’s argument was one inspiration behind the model we are going to introduce shortly. The game we shall consider is the Stag Hunt and the population dynamics will be determined by a Haystack model. These two elements capture efficiency as well as group structure.

Shifting balance theory

Fitness landscapes have been an enormously influential tool in evolutionary biology since their inception by Sewall Wright in his seminal paper “The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution” (Wright 1932; Gavrillets

³ It is important to also note that Binmore’s view on the equilibrium selection problem is much more nuanced than the quoted passage suggests. In particular Binmore argues that humans are especially good at solving equilibrium selection problems and that our fairness norms play an important role in how we solve these problems. However, for our purposes here, we will treat this as a stand-alone argument.

2004). In a fitness landscape, a genotype's fitness is associated with a number such that the numbers express (at least) ordinal relations between fitnesses. The genotypes can then be placed in a "geographical" relationship where degree of similarity corresponds to spatial distance. Higher fitness values can be represented with higher altitude. The resulting landscape can be pictured as having fitness peaks and fitness valleys. According to standard population genetic theory, natural selection will drive populations to occupy regions close to fitness peaks. However, fitness peaks do not have to be of the same height. This implies that populations can get stuck at suboptimal fitness peaks. It also raises the question of how a population can get from a lower fitness peak to a higher one. In order to do so, it has to cross fitness valleys; but, surely, this will be frustrated by natural selection. Thus we have a structurally similar problem to efficient social contracts.

There are also some similarities in the proposed solutions. Wright's shifting balance theory rests on a subdivision of the population:

[L]et us consider the case of a large species which is subdivided into many small local races, each breeding largely within itself but occasionally crossbreeding [...] With many local races, each spreading over a considerable field and moving relatively rapidly in the more general field about the controlling peak, the chances are good that one at least will come under the influence of another peak. If a higher peak, this race will expand in numbers and by crossbreeding with the others will pull the whole species toward the new position. The average adaptiveness of the species thus advances under intergroup selection, an enormously more effective process than intragroup selection (363).

The idea is that a population divided into small sub-populations leads to more genetic variation due to random genetic drift and hence for more "exploration" of the phenotype space (or the genotype space). This then would lead some sub-populations to the higher fitness peak.

Although Wright's shifting balance theory originally focused on fitness values that were non-strategic, there are some parallels between Wright's idea and the idea from Binmore. In particular, we could think of various social contracts as fitness peaks, which vary in height according to efficiency. Dividing a population into small groups allows some individuals to achieve higher fitness values than would be possible otherwise, which then has an impact on the evolution of the population as a whole. The group structure of the population allows a shift from lower to higher peaks or from inefficient social contracts to efficient ones.⁴

The Haystack model

We would like to construct a simple model to investigate the effect of a group-structure on efficient social contracts. One of the earliest models that attempted to

⁴ There are also some important differences between the two ideas. Wright's shifting balance theory rests heavily on small populations being subject to greater genetic drift, whereas Binmore's group selection argument simply depends on variation among which contracts are reached by different groups. The variation Binmore has in mind may have many possible sources.

make group selection explicit was Maynard-Smith's Haystack model (1964). As mentioned in the introduction, scholars have since investigated a wide variety of group selection models. However, the majority of these models have focused on the evolution of altruistic behavior rather than on efficiency. Additionally, Maynard Smith's Haystack model remains one of the simplest models of group selection. For these reasons, we believe the Haystack model will serve as a prudent starting point for our investigation.⁵

Maynard Smith's original model

Maynard Smith (1964) argued that the evolution of altruism was possible via group selection when a population is "divided into a large number of local populations" and three conditions are met. First, each local group is reproductively isolated for some time. Second, groups are started by a very small number of founders. Third, the growth of each group is not affected by the other groups.

To illustrate this possibility, he proposed the well-known "Haystack model", which is described as follows:

...suppose there exists a species of mouse which lives entirely in haystacks. A single haystack is colonized by a single fertilized female, whose offspring form a colony which lives in the haystack until next year, when new haystacks are available for colonization. At this time, mice migrate, and may mate with members of other colonies before establishing a new colony (1146).

The basic idea is that random mating will lead to the formation of some altruistic groups, which increase in size much more rapidly than egoistic groups. This, in turn, leads to an increase in the global proportion of altruism.

One important assumption here is that selection drives altruism to extinction within any mixed groups. "Only when a colony is started by an [altruistic] female fertilized by an [altruistic] male will it consist finally of [altruistic] individuals; all other colonies will lose the [altruistic] gene by selection and come to consist entirely of [egoistic] individuals" (ibid.).

Maynard Smith described the conditions for the spread of altruism formally. Let P_0 be the initial frequency of altruistic groups, r be the probability that individuals mate with members of their previous colony during migration and K be the advantage that altruism confers onto a group's reproductive success.⁶ Maynard Smith derives the following condition for an increase in the proportion of altruistic individuals:

$$r(1 + K) - (1 - P_0K^2) > 0. \quad (1)$$

⁵ Some may argue that the Haystack model is a poor model of group selection due to assumptions that are implausibly strong and not necessary. This may be true with respect to modeling the evolution of altruism, but, as we will discuss in the section "[Discussion](#)", the assumptions of the Haystack model are much more palatable in the case of the Stag Hunt.

⁶ With probability $1 - r$ individuals mate at random with a member of the population. r can also be interpreted as a simple correlation parameter representing the likelihood of mating with an individual of the same type where $r = 0$ represents purely random mating and $r = 1$ represents perfect correlation of types.

Informally we may say that, if there are enough altruists in the initial population, the differential growth of groups can lead to an increase in the global proportion of altruists. Additionally, the more likely individuals are to meet a relative (the higher r), and the greater the benefit to altruism, the easier it is for altruism to increase.

It is important to note here that the “groups” in this model refer only to sub-sets of the population that get separated into the haystacks. Hence, “groups” in this model are not collective units competing against one another, but simply a product of a certain structuring in the population. The evolutionary processes in this model should be viewed as occurring on the level of individual reproduction and how the group-structure influences this evolution.

Haystacks and the prisoner’s dilemma

For our purposes it will be convenient to re-describe Maynard Smith’s original Haystack model in terms of the Prisoner’s Dilemma. We imagine that each individual is hard-wired to cooperate (c) or defect (d) and that each haystack is initially colonized by two individuals. The payoffs are shown in Fig. 2. Since $b > a > 0$, cooperation is driven to extinction within each mixed haystack. We assume that evolution operates within haystacks sufficiently long, so that each haystack is close to the state all- c or to the state all- d . At this point, the groups from the different haystacks get mixed. Since the average fitness in cooperative haystacks is $1 + a$ whereas the average fitness of non-cooperative haystacks is 1, we assume that cooperative haystacks leave $1 + a$ times as many offspring as non-cooperative haystacks. As before, r represents the probability that each individual pairs with a member from their previous colony in colonizing the next haystack.

In this case, the condition for the spread of altruism is identical to the previous model with a in place of K in Eq. 1. It is well known that in any standard evolutionary setting (e.g., the replicator dynamics), a population of all defectors is the unique asymptotically stable state. In the Haystack model, however, cooperation can increase and take over the population if there are enough cooperators (and r and a are sufficiently high). In the next section, we develop a similar analysis which will provide a basis of comparison between the Stag Hunt in groups-structured populations and the Stag Hunt in randomly interacting populations.

Stag hunting in Haystacks

The Stag Hunt as shown in Fig. 1 is parametrized in terms of two payoff parameters, a and b , where $a > b > 0$, $b + 1 > a$. There are two strict Nash equilibria, (c, c) and

Fig. 2 The Prisoner’s Dilemma. $b > a > 0$

	c	d
c	$1 + a$	0
d	$1 + b$	1

(d , d). The first one is Pareto efficient and the second one is risk dominant. Since both are strict Nash equilibria, d and c are evolutionarily stable *sensu* Maynard Smith and Price (1973). Moreover, there is no other evolutionarily stable strategy.

We are interested in investigating how a group selection structure like the Haystack model might influence the likelihood of reaching the more efficient equilibrium. As our point of reference we choose the standard one-population replicator dynamics, which is governed by random interactions and, thus, exhibits no group structure. In this setting, the Stag Hunt has very straightforward results. Let x be the relative frequency of stag hunters in the population. There are three rest points. Corresponding to the two evolutionarily stable strategies, there are two asymptotically stable states: all- c ($x = 1$) and all- d ($x = 0$). A third one, which corresponds to the mixed Nash equilibrium, is unstable: $x = 1/(1 + a - b)$. It separates the basins of attraction of $x = 1$ and $x = 0$. The condition for the increase in c is

$$x > \frac{1}{1 + a - b}.$$

And the basin of attraction for stag-hunting in this setting is

$$1 - \frac{1}{1 + a - b}.$$

Notice that $\frac{1}{1+a-b} > \frac{1}{2}$ for our specifications of the parameters a , b . This means that the basin of attraction for $x = 1$ is smaller than the basin of attraction for $x = 0$. (It also reflects the fact that d is risk dominant for the Stag Hunt of Fig. 1.)

Let's assume, as in the case of the Prisoner's Dilemma, that groups are formed by a single pair of individuals (or equivalently, by single fertilized females) and that natural selection drives each to an equilibrium during the haystack phase. Groups that are primarily type d will reach the all- d equilibrium. Groups that are primarily c will reach the all- c equilibrium.

To be more specific, since $b + 1 > a$, d is risk-dominant and the only groups that will reach the all- c state are those that are colonized by two c individuals (as in the original model). We will assume that stag-hunting groups leave $1 + a$ times as many offspring as hare hunting groups since this corresponds to the average fitness of each haystack after within group selection. This means that the condition for the increase in cooperators is identical for both games:

$$r(1 + a) - (1 - P_0 a^2) > 0.$$

The reason for this is that, after each haystack has reached an equilibrium, the only payoffs that are compared are the diagonal payoffs of Fig. 1.⁷

⁷ If $a \geq b + 1$ we would still have a coordination game, but defection would no longer be risk-dominant. Consequently, it would be easier for cooperation to survive and increase in mixed groups. Depending on how we imagine the selection process with very small numbers, it may mean that mixed groups converge to the stag hunting equilibrium instead of the hare hunting equilibrium. However, achieving efficiency in this class of games is already more likely than not without the additional group structure, because the basin of attraction for stag hunting is larger than the one for hare hunting in this case.

We have noted above that for sufficiently high a and r , cooperation becomes stable in the Prisoner's Dilemma. We conduct an analogous analysis for the Stag Hunt by proceeding in two steps. We first consider the limiting case of no correlation, i.e., $r = 0$, and then show that it essentially generalizes to the case of $r > 0$.

No correlation ($r = 0$)

Since we wish to compare the effects of group selection with the traditional replicator dynamics, we will begin by assuming no correlation during the migration phase ($r = 0$). This means that we have

$$P_0 > 1/a^2$$

as the condition for an increase in the frequency of cooperators. Note that since we are assuming random matching, $P_0 = x^2$, where x is the proportion of c in the population immediately prior to the formation of haystacks. The condition then becomes $x > 1/a$. The size of the basin of attraction for stag-hunting (c) is

$$1 - (1/a).$$

It must be the case that $a \geq 1$ in order for stag hunting to be maintained by selection when x is large. Just as in the traditional model, this means that each stag-hunting group must contribute at least twice the population ($1 + a$) of each hare-hunting group.

Hence, the basin of attraction for stag hunting will increase in the Haystack model relative to the replicator dynamics if and only if

$$\frac{1}{1 + a - b} < \frac{1}{a}$$

or $a > 1 + a - b$, that is, if $b > 1$.

Consequently, in some Stag Hunts (when $b > 1$) the haystacks will increase the chance of reaching a cooperative state. In other Stag Hunts, haystacks will reduce the chance of reaching a cooperative state. Therefore, group selection does not always favor the evolution of efficient equilibria. It can sometimes be a detriment to the evolution of cooperation. Whether or not group selection works in favor of cooperation will depend on the specific payoffs of the game. Since b has to be sufficiently large, it will depend in particular on how well defectors do against cooperators. Higher values of b will reduce the basin of attraction for cooperation in the replicator dynamics; this, however, has no effect in the Haystack model.

The effect of group selection on reaching efficient equilibria may not be as straightforward as it seemed from Binmore's (2005) argument. However, at this point there is a significant difference between the setting Binmore describes and our model. In Binmore's argument, there was no mixing of individuals before the formation of new groups: newly formed groups simply copy the behavior of their parent groups. We can capture something similar by introducing correlation in terms of group formation ($r > 0$). Perfect copying of parent groups is just $r = 1$, but we can explore the effects of positive correlation more generally.

Positive correlation ($r > 0$)

In the general case we have

$$r(1+a) - (1 - P_0 a^2) > 0$$

as condition for the spread of cooperators. This condition is equivalent to

$$P_0 > \frac{1 - r(1+a)}{a^2}.$$

Since r is positive, $P_0 = (1-r)x^2 + rx$, where x is again the initial share of stag hunters in the population. This leads to the condition

$$(1-r)x^2 + rx - \frac{1 - r(1+a)}{a^2} > 0$$

for the spread of c . The left-hand side of this inequality is a quadratic in x which has two zeros, namely,

$$-1/a \quad \text{and} \quad \frac{1}{a} - \frac{r}{1-r}.$$

The first solution is not relevant, since x has to be non-negative. The second condition tells us that the share of stag hunters will grow if

$$x > \frac{1}{a} - \frac{r}{1-r}.$$

From this we get two immediate conclusions. Firstly, in the case of $r > 0$ the basin of attraction of c is always at least as great as its basin of attraction for the $r = 0$ case. And secondly, the basin of attraction of all- c can be made arbitrarily large. To see this, just fix a and let $r \rightarrow 1$.

As one would expect, introducing correlation into the model enhances the evolutionary prospects for cooperation in the Stag Hunt, and the more correlation the better. When r is close to 1 we can guarantee selection of the efficient equilibrium in the Stag Hunt even if a is very small. However, to more precisely judge the effect of group selection on cooperation in this setting we may again compare the Haystack model with the replicator dynamics.

This comparison can be done in two ways. One is to use a correlated replicator dynamics where encounters between types are non-random. The other is the standard replicator dynamics with random encounters.

We first consider the case of the correlated replicator dynamics.⁸ Suppose two individuals of the same type (c and c or d and d) meet each other with probability r . Then the expected payoff to stag hunters is

$$(1-r)(1+a)x + (1+a)r,$$

⁸ To our knowledge this analysis has nowhere been published before.

where x is the relative frequency of stag hunters. Likewise, the expected payoff to hare hunters is

$$(1-r)((1+b)x + 1-x) + r = (1-r)(bx + 1) + r.$$

At an interior equilibrium $x \in (0, 1)$, stag hunters and hare hunters get the same payoff, whence

$$(1-r)(1+a)x + (1+a)r = (1-r)(bx + 1) + r.$$

Hence, solving for x gives

$$x = \frac{1-r(1+a)}{(1-r)(1+a-b)}.$$

Notice that this approaches $1/(1+a-b)$ as r goes to 0.

We first note that *correlation always helps stag hunters in the replicator dynamics*. This follows from the fact that

$$\frac{1-r(1+a)}{(1-r)(1+a-b)} < \frac{1}{1+a-b}$$

since $a > 0$. We next show that stag hunters do better in the Haystack model with correlation rate r than in the replicator dynamics with correlation r if, and only if, $b > 1$. This follows straightforwardly from

$$\frac{1-r(1+a)}{(1-r)(1+a-b)} > \frac{1}{a} - \frac{r}{1-r}.$$

Hence, $b > 1$ is the relevant condition whether $r = 0$ or not. In this case, we again see that the haystack structure may either increase or decrease the likelihood of reaching the efficient equilibrium depending on the value of b .

Lastly, we can show that stag hunters do better in the Haystack model with correlation r than in the replicator dynamics with random encounters ($r = 0$) if, and only if,

$$\frac{1}{a} - \frac{r}{1-r} < \frac{1}{1+a-b}.$$

This is equivalent to

$$r > \frac{1-b}{1+(1+a)(a-b)}.$$

Notice that this condition is relevant only if $b \leq 1$, for otherwise stag hunters do better in the Haystack model independently of r . Hence, if $b \leq 1$, correlation during migration must be relatively high in order to enhance the evolution of cooperation. Relative to the standard replicator dynamics, whether the haystacks (with correlation) help stag hunting depends on both the payoffs and the correlation rate r . Therefore, even in this case, haystacks with correlation do not necessarily enhance the prospects for reaching the efficient equilibrium in the Stag Hunt.

Discussion

Extensions

Before we discuss the significance of the above results, we would like to point out several possible areas where our analysis could be extended. Firstly, less simplistic models should take account of a number of idealizing assumptions that we made. One assumption is that we assume there to be no carrying capacity for the population nor for the haystacks. This may be important since a low carrying capacity may actually be disadvantageous for faster growing groups. There is also no direct competition between groups. Introducing direct competition may enhance the prospects of efficiency.

Another restrictive assumption has to do with the growth within each haystack. We made the simplifying assumption that the payoffs of the individual interactions between cooperators were directly translatable into the population contributions of each haystack (both were $1 + a$). While this assumption may be natural, it will not hold necessarily. If there is no carrying capacity within each haystack and many generations lived during each haystack phase, then population growth within each haystack would be geometric with time. This means the contribution of each haystack would be a function of both the fitness of the interactions and the time spent in the haystack. For example, suppose that haystacks were only torn down every other year. In this case, if stag hunting haystacks were reproducing at *twice* the rate of hare hunting haystacks (per year), they would produce *four* times the population of hare hunting haystacks since they are allowed to live for two full reproductive cycles.⁹ To de-idealize the model in this respect would require introducing additional complexities.

Models which depart from the above assumptions are an interesting area for future work. We would like to emphasize, however, that our assumptions do facilitate a helpful comparison with the replicator dynamics. Generalizations may have to use another point of comparison.

There are several other areas that may warrant future investigation. Firstly, efficiency plays an important role in games other than the Stag Hunt. Binmore (2005) also considers bargaining games. In bargaining games participants may strike deals that are more or less efficient. Hence, while the Stag Hunt provides a nice starting point for assessing efficient social contracts, there are many other games that one may investigate. Secondly, it would be interesting to consider alternative evolutionary dynamics involving finite populations or stochastic evolutionary processes as different dynamics can sometimes lead to different results. Finally, as mentioned above, Maynard Smith's Haystack model is only one of many in the group selection controversy. Our results show that, even in this simple case, the effect of group selection on efficiency is not obvious. This suggests

⁹ Note that additional generations within the haystacks would be equivalent to increasing the a in the payoff matrix for purposes of calculating group contributions. This may then mean that $a > b + 1$, which would mean hare hunting is not risk-dominant.

that other models of group selection may also be interesting to examine in this respect.

Comparison with the prisoner's dilemma

Returning to the model explored above, there are some important differences between the effects of group selection in the Stag Hunt and the effects of group selection in the Prisoner's Dilemma. In neither case does the payoff to defectors against cooperators have a significant impact. However, this payoff does matter to the basin of attraction for stag hunting in the replicator dynamics. This means that, relative to the replicator dynamics, group selection can never reduce the chance of reaching cooperation in the Prisoner's Dilemma, but it can change the chance of reaching cooperation in the Stag Hunt.

There are other differences as well. Most importantly, cooperation in the Stag Hunt will not require the same restrictive assumptions that are necessary for attaining cooperation in the Prisoner's Dilemma. Consider Maynard Smith's first necessary condition for haystack-like group selection of altruism: each haystack is reproductively isolated. Introducing even a small amount of migration *between* haystacks will destroy the results for the Prisoner's Dilemma—the extreme reproductive isolation required was one of the reasons that Maynard Smith was skeptical of the model's explanatory power (Maynard Smith 1964). However, reproductive isolation is not required if the underlying game is the Stag Hunt. Stag hunting, unlike altruism, is stable and can resist invasion from mutants or migrants. Any mutant or migrant defectors will be driven out by selection within each cooperative group as long as the number of mutants is sufficiently small. Hence, although this assumption cannot be dropped entirely, it could be substantially weakened without threatening the result. This makes the reproductive isolation assumption more palatable for analyzing the Stag Hunt.

There are also differences with respect to Maynard Smith's second assumption that the groups are started by a very small number of founders. This assumption is necessary in the case of games like the Prisoner's Dilemma because homogeneous populations of altruists are needed for altruistic groups to persist at all. However, in the Stag Hunt, it is only necessary that some groups have *enough* cooperators to drive the defectors to extinction; some initially mixed groups can become cooperative groups. Thus, the strength of Maynard Smith's second assumption is arguably less important for the case of the Stag Hunt than it is for the Prisoner's Dilemma.

Group selection might be an important factor in thinking about efficiency of equilibria (as in the Stag Hunt). However, this importance is not due to group selection making efficient equilibria a moral certainty, for we have seen that the specific effect is variable and can even be detrimental to efficiency. Group selection becomes an important consideration in games like the Stag Hunt because the required assumptions needed are far more plausible than in cases like the Prisoner's Dilemma.

Haystacks and efficiency

Binmore (2005) argued informally that group selection can lead to the predominance of efficient social contracts, but left us with open questions about the details of this process. Using our model we can now ask a much more precise question: how often does a haystack-like group-structure lead to efficiency? The answer depends crucially on the degree of group cohesion. There are circumstances in which efficiency is guaranteed (when $r = 1$), but these conditions are very narrow.¹⁰ Furthermore, it should be noted that $r = 1$ would also guarantee selection of cooperation in the Prisoner's Dilemma. For r close to 1 it does not matter that (c, c) is an *efficient equilibrium* in the Stag Hunt, only that it generates a higher payoff than (d, d) , which is also the case in the Prisoner's Dilemma.

It should be emphasized, however, that for more moderate values of r , all- c remains a likely outcome of the evolutionary process in the Stag Hunt. This is good news for the Stag Hunt. The scandal of the Stag Hunt is not so much that d can be reached by an evolutionary process—after all, inefficient equilibria appear to be widespread; the scandal is, instead, that d is the likelier evolutionary outcome for a whole range of evolutionary dynamics. Our simple Haystack model thus allows for the selection of the efficient over the inefficient equilibrium. In the best of all cases (if r is sufficiently large), it even guarantees the selection of the efficient equilibrium.

As we have pointed out in “[Background](#)”, social contract theory and evolutionary biology both ask how it is possible to get to efficient states from stable inefficient ones. In both cases, one may put forward an argument that essentially rests on the logic of group selection. Our results partially vindicate such arguments. In situations like the Stag Hunt, group structure can make the efficient outcome more likely because it is also an equilibrium and can therefore not be undermined easily.

There are some important caveats, however. Firstly, even if group structure enhances the evolution of cooperation, it may not guarantee it. More importantly, group structure can be *detrimental* to the evolution of cooperation, in cases where cooperation is already fairly likely to evolve in unstructured populations.

This point is worth emphasizing. It can be tempting to think of group selection as always acting in favor of cooperation, or at least never acting against it. However, we have shown that in the case of the Stag Hunt, things are not so straightforward. Even in this simple model, we have identified cases where the group structure acts *against* the evolution of cooperation. The details matter with respect to efficiency: group selection sometimes helps the evolution of cooperation, and sometimes hinders it.

¹⁰ Maynard Smith was skeptical of the significance of this condition, remarking that “this corresponds to the case in which the population is divided into more or less permanently isolated groups, which are periodically reduced to very small numbers, and which may either become extinct or split to give rise to two groups...it is unlikely that species are often divided into a large number of small and completely isolated groups” (1146).

Acknowledgments We would like to thank Ken Binmore, Patrick Forber, James Marshall, Samir Okasha, and Kevin Zollman for many helpful comments and for discussing earlier versions of this paper.

References

- Binmore K (2005) *Natural justice*. Oxford University Press, Oxford
- Calcott B (2008) The other cooperation problem: generating benefit. *Biol Philos* 23:179–203
- Gavrilets S (2004) *Fitness landscapes and the origin of species*. Princeton University Press, Princeton
- Kandori M, Mailath GJ, Rob R (1993) Learning, mutation, and long run equilibria in games. *Econometrica* 61:15–18
- Maynard Smith J (1964) Group selection and kin selection. *Nature* 201:1145–1147
- Maynard Smith J (1976) Group selection quarterly review of biology 51:277–283
- Maynard Smith J, Price G (1973) The logic of animal conflict. *Nature* 246:1145–1147
- Okasha S (2009) *Evolution and the levels of selection*. Oxford University Press, Oxford
- Samuelson L (1998) *Evolutionary games and equilibrium selection*. MIT Press, Cambridge
- Skyrms B (2004) *The stag hunt and the evolution of social structure*. Cambridge University Press, Cambridge
- Sober E, Wilson DS (1998) *Unto others. The evolution and psychology of unselfish behavior*. Harvard University Press, Harvard
- Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the sixth international congress of genetics* 1:356–366
- Young HP (1993) The evolution of conventions. *Econometrica* 61:57–84