The Stag Hunt and the Evolution of Social Structure

BRIAN SKYRMS

University of California
Irvine

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2004

Printed in the United States of America

Typeface Meridien 10/13 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Skyrms, Brian.
The stag hunt and the evolution of social structure / Brian Skyrms.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-82651-9 – ISBN 0-521-53392-9 (pbk.)
1. Social structure. 2. Social action. 3. Collective behavior.
4. Social evolution. 5. Social networks. 6. Cooperation. I. Title.

HM706.S54 2004
301–dc21 2003051530

ISBN 0 521 82651 9 hardback
ISBN 0 521 53392 9 paperback

CONTENTS

1

THE STAG HUNT

THE STAG HUNT

THE Stag Hunt is a story that became a game. The game is a prototype of the social contract. The story is briefly told by Rousseau, in *A Discourse on Inequality*: "If it was a matter of hunting a deer, everyone well realized that he must remain faithful to his post; but if a hare happened to pass within reach of one of them, we cannot doubt that he would have gone off in pursuit of it without scruple."[1] Rousseau's story of the hunt leaves many questions open. What are the values of a hare and of an individual's share of the deer, given a successful hunt? What is the probability that the hunt will be successful if all participants remain faithful to the hunt? Might two deer hunters decide to chase the hare?

Let us suppose that the hunters each have just the choice of hunting hare or hunting deer. The chances of getting a hare are independent of what others do. There is no chance of bagging a deer by oneself, but the chances of a successful deer hunt go up sharply with the number of hunters. A deer is much more valuable than a hare. Then we have the kind of interaction that is now generally known as the stag hunt.

This chapter is drawn from my APA presidential address, Skyrms (2001).

Once you have formed this abstract representation of the stag hunt game, you can see stag hunts in many places. David Hume also has the stag hunt. His most famous illustration of a convention has the structure of a two-person stag hunt game: "Two men who pull at the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other."[2] Both men can either row or not row. If both row, they get the outcome that is best for each – just as, in Rousseau's example, when both hunt the stag. If one decides not to row, then it makes no difference if the other does or does not – they don't get anywhere. The worst outcome for you is if you row and the other doesn't, for then you lose your effort for nothing, just as the worst outcome for you in the stag hunt is if you hunt stag by yourself.

We meet the stag hunt again in the meadow-draining problem of Hume's *Treatise*: "Two neighbors may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each others mind, and each may perceive that the immediate consequence of failing in his part is the abandoning of the whole project. But 'tis difficult, and indeed impossible, that a thousand persons shou'd agree in any such action."[3] In this brief passage, Hume displays a deep understanding of the essential issues involved. He sees that cooperation in the stag hunt is consistent with rationality. He sees that the viability of cooperation depends on mutual beliefs, and rests on trust. He observes that for these reasons, achieving cooperation in a many-person stag hunt is more difficult than achieving cooperation in a two-person stag hunt.[4]

The stag hunt does not have the same melodramatic quality as the prisoner's dilemma. It raises its own set of issues, which are at least as worthy of serious consideration. Let us focus, for the moment, on a two-person stag hunt for comparison to the familiar two-person prisoner's dilemma.

If two people cooperate in prisoner's dilemma, each is choosing less rather than more. In prisoner's dilemma, there is a conflict between individual rationality and mutual benefit. In the stag hunt, what is rational for one player to choose depends on his beliefs about what the other will choose. Both stag hunting and hare hunting are *Nash equilibria*. That is just to say that it is best to hunt stag if the other player hunts stag, and it is best to hunt hare if the other player hunts hare. A player who chooses to hunt stag takes a risk that the other will choose not to cooperate in the stag hunt. A player who chooses to hunt hare runs no such risk, since his payoff does not depend on the choice of action of the other player, but he forgoes the potential payoff of a successful stag hunt. In the stag hunt game, rational players are pulled in one direction by considerations of mutual benefit and in the other by considerations of personal risk.

Suppose that hunting hare has an expected payoff of 3, no matter what the other does. Hunting stag with another has an expected payoff of 4. Hunting stag alone is doomed to failure and has a payoff of 0. It is clear that a pessimist, who always expects the worst, would hunt hare. But it is also true with these payoffs that a cautious player, who was so uncertain that he thought the other player was as likely to do one thing as another, would also hunt hare. Hunting hare is said to be the *risk-dominant* equilibrium.[5] That is not to say that rational players could not coordinate on the stag hunt equilibrium that gives them both a better payoff, but it is to say that they need a measure of trust to do so.

I told the story so that the payoff of hunting hare is absolutely independent of how others act. We could vary this slightly without affecting the underlying theme. Perhaps if you hunt hare, it is even better for you if the other hunts stag, for you avoid competition for the hare. If the effect is small, we still have an interaction that is much like the Stag Hunt. It displays the same tension between risk and mutual benefit. It raises the same question of trust. This small variation on the stag hunt is sometimes also called a stag hunt,[6] and we will follow this more inclusive usage here.

Compared to the prisoner's dilemma, the stag hunt has received relatively little discussion in contemporary social philosophy – although there are some notable exceptions.[7] But I think that the stag hunt should be a focal point for social contract theory.

The two mentioned games, prisoner's dilemma and the stag hunt, are not unrelated. We will illustrate the connection in two rather different contexts – the first dealing with considerations of prudence, self-interest, and rational choice, and the second having to do with evolutionary dynamics in a model of group selection.

THE STAG HUNT AND THE SHADOW OF THE FUTURE

The first context arises in classical political philosophy. Considerations raised by both Hobbes and Hume can show that a seeming prisoner's dilemma is really a stag hunt. Suppose that prisoner's dilemma is repeated. Then your actions on one play may affect your partner's actions on other plays, and considerations of reputation may assume an importance that they cannot have if there is no repetition. Such considerations form the substance of Hobbes's reply to the Foole. Hobbes does not believe that the Foole has made a mistake concerning the nature of rational decision. Rather, he accuses the Foole of a shortsighted mis-specification of the relevant game: "He, therefore, that breaketh his Covenant, and consequently declareth that he think that he may with reason do so, cannot be received into any society that unite themselves for Peace and Defense, but by the error of them that receive him."[8] According to Hobbes, the Foole's mistake is to ignore the future.

David Hume invokes the same considerations in a more general setting: "Hence I learn to do a service to another, without bearing him any real kindness; because I foresee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me and with others."[9] Hobbes and Hume are invoking the *shadow of the future*.[10]

How can we analyze the shadow of the future? We can use the theory of indefinitely repeated games. Suppose that the probability that the prisoner's dilemma will be repeated another time is constant. In the repeated game, the *Foole* has the strategy of always defecting. Hobbes argues that if someone defects, others will never cooperate with the defector. Those who initially cooperate but who retaliate, as Hobbes suggests against defectors, have a *Trigger* strategy.

If we suppose that Foole and Trigger are the only strategies available in the repeated game and that the probability of another trial is .6, then the shadow of the future transforms the two-person prisoner's dilemma

	Cooperate	Defect
Cooperate	2	0
Defect	3	1

into the two-person stag hunt.[11]

	Trigger	Foole
Trigger	5	1.5
Foole	4.5	2.5

This is an exact version of the informal arguments of Hume and Hobbes.[12]

But for the argument to be effective against a fool, he must believe that the others with whom he interacts are not fools. Those who play it safe will choose Foole. Rawls's maximin player is Hobbes's Foole.[13] The shadow of the future has not solved the problem of cooperation in the prisoner's dilemma; it has transformed it into the problem of cooperation in the stag hunt.

GROUP SELECTION AND THE STAG HUNT

Cooperation is also a problem for evolutionary theory. How can cooperation evolve in a context of competition for survival? Darwin recognized the problem. In Darwin's own time, it was the focus of Petr Kropotkin's 1908 *Mutual Aid: A Factor in Evolution.*

More recently (1962), V. C. Wynn-Edwards revived the issue in *Animal Dispersion in Relation to Social Behavior.* He argued that many natural populations practiced reproductive restraint, which is contrary to individual "selfish" behavior, because of its benefit to the group in preserving food supply. The idea was that natural selection applies to groups, as well as individuals. The explanatory force of this sort of appeal to "group selection" was severely criticized by George Williams in 1966. Natural selection operating on populations operates at a much slower rate than natural selection operating on individuals. Williams argued that as a result, group selection would be a much weaker evolutionary force than individual selection. After the publication of his *Adaptation and Natural Selection*, many evolutionary biologists dismissed group selection as an interesting part of evolutionary theory.

But John Maynard Smith, the father of evolutionary game theory, was motivated in 1964 to find a model in which some kind of group selection could account for the evolution of altruism. He took cooperation in the prisoner's dilemma as the paradigm of altruistic behavior.

Maynard Smith imagines a large hayfield. In the fall the farmer mows hay and makes haystacks. Each haystack is colonized by two mice, drawn at random, from the ambient mouse population. Over the winter the mice play prisoner's dilemma and reproduce. In the spring the haystacks are torn down, and the mice scatter to form the ambient population for the next cycle. Haystacks full of cooperative mice produce more mice than those full of defectors, so it seems that here the group structure – where inhabitants of a given haystack are the group – should be able to sustain the evolution of cooperation in the prisoner's dilemma.

We can see how this is so in the simplest possible haystack model. (There is a whole literature on generalized haystack models, and we will illustrate principles that hold good in general.) For simplicity we will suppose that the mice pair at random within the haystack, play the prisoner's dilemma, reproduce asexually with number of offspring determined by payoff, and repeat the process for the number of generations for which the haystack remains intact.

Consider the Prisoner's Dilemma.

	Cooperate	Defect
Cooperate	2	0
Defect	3	1

If the haystack is colonized by two defectors, each gets a payoff of 1, so in the next generation there are still two defectors, and so for all subsequent generations. If the haystack is founded by a defector and a cooperator, the cooperator gets a payoff of 0 and has no progeny. The defector gets a payoff of 3 and the next generation has three defectors. At all subsequent generations the haystack has only defectors, and so the population is maintained at 3 defectors. (Don't worry about pairing.) Two cooperators produce four cooperators in generation 1, eight in generation 2, and so forth.

If the haystacks are torn down after generation 1 is born, then group selection doesn't work. The dynamics is the same as if there were no group structure and defection drives out cooperation. But if the population stays together for two generations or more, it is possible for cooperation to be sustained.

There are two complementary ways to look at this result. One is to focus on the game played within the haystacks, the prisoner's dilemma. From this point of view, the key fact is that after one generation the dynamics induces perfect correlation of types – cooperators only meet cooperators and defectors only meet defectors. Then, of course, cooperators can flourish, because it is a defining characteristic of the prisoner's dilemma that cooperators do better against themselves than defectors do against defectors. The temporary advantage of being able to defect against cooperators is gone after the initial interaction because it removes potential victims from successive generations in the haystack.

The second way of looking at the haystack model, suggested by Ted Bergstrom in 2002, is to consider the game played by founders of the haystacks. Founders are chosen at random from the ambient population. The payoffs from the game between founders are the number of progeny when the haystack is torn down. In our example, if the haystacks are torn down after two generations, the payoffs in the founders game are as follows:

	Cooperate	Defect
Cooperate	4	0
Defect	3	1

This is a stag hunt.

And, as we know, the stag hunt does not *solve* the problem of cooperation. It allows cooperation in equilibrium, but there is also the noncooperative equilibrium. If we start our two-generation haystack dynamics in a state where the ambient population is equally divided between cooperators and defectors, defection will take over the population. Group selection can transform the problem of cooperation in the prisoner's dilemma into the problem of cooperation in the stag hunt.

THE STAG HUNT AND THE SOCIAL CONTRACT

In a larger sense, the whole problem of adopting or modifying the social contract for mutual benefit can be seen as a stag hunt. For a social contract theory to make sense, the state of nature must be an equilibrium. Otherwise, there would not be the problem of transcending it. And the state where the social contract has been adopted must also be an equilibrium. Otherwise, the social contract would not be viable. Suppose that you can either *devote energy to instituting the new social contract* or not. If everyone takes the first course, the social contract equilibrium is achieved; if everyone takes the second course, the state of nature equilibrium results. But the second course carries no risk, while the first does. This is all quite nicely illustrated in miniature by the meadow-draining problem of Hume.

The problem of reforming the social contract has the same structure. Here, following Ken Binmore (1993), we can then take the relevant "state of nature" to be the status quo, and the relevant social contract to be the projected reform. The problem of instituting, or improving, the social contract can be thought of as the problem of moving from riskless hunt hare equilibrium to the risky but rewarding stag hunt equilibrium.

GAME DYNAMICS

How do we get from the hunt hare equilibrium to the stag hunt equilibrium? We could approach the problem in two different ways. We could follow Hobbes in asking the question in terms of rational self-interest. Or we could follow Hume by asking the question in a dynamic setting. We can ask these questions using modern tools – which are more than Hobbes and Hume had available, but still less than we need for fully adequate answers.

The modern embodiment of Hobbes's approach is rational choice–based game theory. It tells us that what a rational player will do in the stag hunt depends on what that player thinks the other will do. It agrees with Hume's contention that a thousand-person stag hunt would be more difficult to achieve than a two-person stag hunt, because – assuming that everyone must cooperate for a successful outcome to the hunt – the problem of trust is multiplied. But if we ask how people can get from a hare hunt equilibrium to a stag hunt equilibrium, it does not have much to offer. From the standpoint of rational choice, for the hare hunters to decide to be stag hunters, each must *change individual beliefs* about what the other will do. But rational choice–based game theory, as usually conceived, has nothing to say about how or why such a change of mind might take place.

Let us turn to the tradition of Hume. Hume emphasized that social norms can evolve slowly: "Nor is the rule regarding the stability of possession the less derived from human conventions, that it arises gradually, and acquires force by a slow progression."[14] We can reframe our problem in terms of the most thoroughly studied model of cultural evolution, the replicator dynamics.[15] This is a deterministic dynamics, intended for large populations in which the effects of chance fluctuations average out. We can ask, in this framework, how one can get from the hunt hare equilibrium to the hunt stag equilibrium; the answer is that you can't! In the vicinity of the state where all hunt hare, hunting hare has the greatest payoff. If you are close to it, the dynamics carries you back to it.

This reasoning holds good over a large class of adaptive deterministic dynamics, which generalize the replicator dynamics. Let us say that a dynamics is *adaptive* if it leads to strategies that do better than average increasing their population proportion and to strategies that do worse than average decreasing their population proportion. For any adaptive dynamics, the reasoning of the previous paragraph continues to hold good. The transition from noncooperation to cooperation seems impossible.

Perhaps the restriction to deterministic dynamics is the problem. We may just need to add some chance variation. We could add some chance shocks to the replicator dynamics[16] or look at a finite population where people have some chance of doing the wrong thing, or just experimenting to see what will happen.[17] If we wait long enough, chance variation will bounce the population out of hare hunting and into stag hunting. But in the same way, chance variation can bounce the population out of stag hunting into hare hunting. Can we say anything more than that the population bounces between these two states?

We can,[18] and in this case the analysis is very simple. It depends on the relative magnitude of the basins of attraction of the stag hunting equilibrium and the hare hunting equilibrium. Let me illustrate with our original version of the stag hunt game: Hunting hare has a payoff of 3, no matter what the other does; hunting stag with another has a payoff of 4; and hunting stag alone has a payoff of 0. If more than 75 percent of the population hunts stag, then stag hunters will take over. This is the "basin of attraction" of the stag hunting equilibrium. If less than 75 percent of the population hunts stag, then hare hunters will take over. This is the basin of attraction of the hare hunting equilibrium – which is triple the size of that of the stag hunting equilibrium.

If mutation (or experimentation) probabilities are small and independent across individuals, and the population is large, it will be much more likely for chance events to move the population from the stag hunting equilibrium into the basin of attraction of hare hunting than for the converse to happen. In the long run, the population spends almost all of its time in a state where everyone hunts hare.[19] It seems that all we have achieved so far is to show how the social contract might degenerate spontaneously into the state of nature.

Social contracts do sometimes spontaneously dissolve. But social contracts also form. People do, in fact, engage in stag hunts (and antelope hunts and giraffe hunts and pig hunts and bison hunts). Cooperative hunting is an ancient part of the human social contract that goes back to the beginning of our race. It is not so easy to infer those distant payoffs and to determine the risk-dominant equilibrium in an appropriate game model. But there is contemporary experimental evidence that people will sometimes hunt stag even when it is a risk to do so.[20] In a whole series of experiments, stag hunting is the most frequent strategy on the first round. People do not enter the laboratory with a norm of playing the risk-dominant strategy. When the game is repeated with pairwise random matching in a group of subjects, sometimes the group converges to all stag hunting and sometimes to all hare hunting, depending on the initial composition of the group. If the group starts in the basin of attraction of stag hunting, then the group almost always converges to all stag hunters. If the initial composition of the group is in the basin of attraction of hare hunting, hare hunters take over.

In a novel experiment, F. W. Rankin, J. B. Van Huyck, and R. Battalio presented subjects with a series of stag hunts in which payoffs varied from game to game and action labels were changed, so that subjects needed to evolve some rule for dealing with them. Subjects converged to payoff – dominance. Stag hunting, although it was not identified as such to the subjects, emerged as a coordination principle.[21] These experimental results, as well as our wider knowledge of the world of social interactions, suggest the need for a richer theory.