

# Natural Social Contracts

Brian Skyrms

Received: 20 March 2012 / Accepted: 5 November 2012 / Published online: 25 April 2013  
© Konrad Lorenz Institute for Evolution and Cognition Research 2013

**Abstract** There are two fundamental problems for instituting a social contract. The first is cooperating to produce a surplus; the second is deciding how to divide this surplus. I represent each problem by a simple paradigm game, a Stag Hunt game for cooperating to produce a surplus, and a bargaining game for its division. I will discuss these simple games in isolation, and end by discussing their composition.

**Keywords** Dynamic networks · Signals · Social contract · Stag Hunt

*“For they (philosophers) conceive of men, not as they are, but as they themselves would like them to be. Whence it has come to pass that, instead of ethics, they have generally written satire...”*

—Spinoza, *Tractatus Politicus* (1677).

## Introduction

Spinoza’s targets are not restricted to predecessors and contemporaries. New instances have continued to pop up. But there is also a naturalistic tradition in philosophy. Social contracts evolved. The view goes back to the great Greek atomist Democritus (Cole 1967, Verlinsky 2005)—perhaps the first thoroughly scientific philosopher. Democritus’ works are lost, and we know his theories from secondary sources. But the naturalistic view of the social

contract finds clear expression in the work of another great naturalist, David Hume:

Two men who pull the oars of a boat, do it by an agreement or convention, although they have never given promises to each other. Nor is the rule concerning the stability of possessions the less derived from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it. On the contrary, this experience assures us still more, that the sense of interest has become common to all our fellows, and gives us confidence of the future regularity of their conduct; and it is only on the expectation of this that our moderation and abstinence are founded. In like manner are languages gradually established by human conventions without any promise. In like manner do gold and silver become the common measures of exchange, and are esteemed sufficient payment for what is of a hundred times their value. (Hume 1739)

Study of the evolution of the social contract has both an empirical and a theoretical side. The empirical side ranges across the social and also the biological sciences. Philosophers should not lose sight of the fact that social contracts have evolved in many species. We should follow Darwin in *The Descent of Man* (and indeed, Democritus) in placing man in proper biological context.

This paper will focus mostly on the theoretical side of a naturalistic view, although some references to relevant experimental evidence will be included. A full-blown theory of the social contract for any society, for any species, is out of reach. I will use a sketch of general problems of the social contract from which to hang various theoretical considerations. This is just a cartoon and nothing more.

---

B. Skyrms (✉)  
Department of Logic and Philosophy of Science, University  
of California, Irvine, CA, USA  
e-mail: bskyrms@uci.edu

B. Skyrms  
Department of Philosophy, Stanford University, Stanford,  
CA, USA

There are two fundamental problems for instituting a social contract. The first is cooperating to produce a surplus; the second is deciding how to divide this surplus. A more detailed cartoon would add other features, for instance division of labor, but here I will stick with the two-part scheme. I represent each problem by a simple paradigm game, a Stag Hunt game for cooperating to produce a surplus, and a bargaining game for its division.

We should take care in discussing each of the problems in isolation, since solutions may have coevolved. I will briefly return to this point later. But I think that it is best to try to understand the simple games first, and then move on to their composition.

### The Stag Hunt

The Stag Hunt is introduced by Rousseau in the *Discourse on Inequality* in 1755, although by other names it appears earlier—notably in David Hume’s *Treatise of Human Nature: Being an Attempt to introduce the Experimental Method of Reasoning into Moral Subjects* (1739). Hare hunters are solitary. They rely on no others in pursuing their small prey. Stag hunters are collective hunters. If they successfully cooperate then they can bring home a lot of meat. But if a stag hunter attempts to hunt with a hare hunter, the stag is not caught and the stag hunter gets nothing. Essential elements of this scenario can be encapsulated in a two-person game:

	Stag	Hare
Stag	3	0
Hare	2	2

(Payoffs are for row played against column.)

There are two pure equilibria in the game,  $\langle$  Stag, Stag  $\rangle$  and  $\langle$  Hare, Hare  $\rangle$ , representing the social contract and the pre-contract state of nature, respectively.

The Stag Hunting equilibrium produces a surplus, but to achieve it requires a measure of trust. A stag hunter runs the risk that his partner may defect and chase a hare, ruining the hunt. A hare hunter runs no such risk. The equilibrium where all cooperate to hunt stag is the *payoff dominant* equilibrium; that where all hunt hare is the *risk dominant* equilibrium.

Two-player games are easiest to analyze, but most instances of cooperative hunting involve more than two hunters. How should the Stag Hunt game be generalized to multiple players? There are various possibilities. We might require all of  $N$  players to cooperate to get any success,

with any defection totally ruining the hunt. But this is clearly an extreme case. Perhaps some minimum is required, with extra cooperators being superfluous. Perhaps a minimum is required for some chance of success, but extra cooperators increase the probability of a good catch. We will take the hallmark of an  $n$ -person Stag Hunt to be the existence of both a non-cooperation equilibrium, and a (possibly partial) cooperation equilibrium, with cooperation bringing a greater payoff.

If we move from two to three players, an example is Taylor and Ward’s (1982) very interesting game of “three in a boat, two can row.” The title is self-explanatory. If no one rows, the boat goes nowhere. If one person rows, the boat goes in circles, and she works hard for nothing. If any two row, the boat gets to its destination. A third rower doesn’t add anything. There are two kinds of pure equilibrium. A no-row equilibrium, and three payoff-dominant equilibria where two row and one does not. Cooperation is certainly possible in equilibrium, but here the desire to free ride is an additional impediment to coordination on such an equilibrium.

Public goods provision games with a threshold provide general examples of  $n$ -player Stag Hunts (Pacheco et al. 2009). There is a “Hare Hunt” equilibrium where no one contributes to the public good. There are partial cooperation equilibria where the public good is produced. There is a temptation to free ride. These are much richer models than the much-discussed  $N$ -person Prisoners Dilemma.

The *fundamental problem of the institution of the social contract* in the context of the Stag Hunt game is: **How do we get from a non-cooperative equilibrium to a cooperative equilibrium?**

How indeed? *Hunt hare* and *hunt stag* are both evolutionarily stable strategies in the two-person Stag Hunt game. They are both attractors in the replicator dynamics. In a large, random mixing population you can’t get from the non-cooperative equilibrium to the cooperative one by differential reproduction. Adding a little mutation (i.e., moving to replicator-mutator dynamics) doesn’t really change this.

What about a small population with random encounters and mutation between stag and hare hunters? Now if you wait long enough you can get from any state of the population to any other. So you *can* get from hare hunting to stag hunting. But for small mutation rates, the population will spend almost all its time at the hare hunting equilibrium. Hare hunting is the unique *stochastically stable strategy* (Foster and Young 1990; Kandori et al. 1993). What we have here is a model of the devolution of the social contract rather than its evolution!

That is all right. Social contracts sometimes do come apart. And we want to model man as he is, rather than as we want him to be. But social contracts also sometimes

evolve. This also needs to be explained. So there is something missing.

Here I will discuss three mechanisms that can promote the evolution of cooperation: *signals*, *local interaction*, and *adaptive networks*.

### Signals

We preface our Stag Hunt game with an initial exchange of signals. Individuals can then condition their decision to hunt stag or hare on the signal received. Signals cost nothing, and have no initially assigned meaning. Meaning can coevolve with strategy. This is surely a modest step, and one that is realistic for humans and many other organisms. Nevertheless, it completely changes the evolutionary dynamics. Signals can destabilize the inefficient non-cooperative equilibrium (Robson 1990). Cooperators can use a signal as a “secret handshake” that they can cooperate with each other, and play like the natives against the natives. Just the addition of two signals reverses the size of basins of attraction in the Stag Hunt game, with evolution usually leading to hunting stag (Skyrms 2002, 2004). More signals are better. Santos et al. (2011) show that in a finite population with mutation, in a Stag Hunt with lots of available signals, the population spends almost all of its time in the Stag Hunting equilibrium. A little cheap talk (i.e., costless signaling) reverses the previous results!

Is experiment consistent with theory? A large human experimental literature shows significant effects of costless preplay communication in a variety of games. With respect to Stag Hunts, two papers deserve special mention. Cooper et al. (1992) find that two-way communication in a two-person Stag Hunt almost guarantees cooperation. Blume and Ortmann (2007) investigate a variety of 9-person generalized Stag Hunt games. Preplay communication promotes coordination on the payoff-dominant equilibrium.

### Local Interaction

The population may not be random mixing. Instead individuals may interact with their neighbors on a spatial grid or, more generally, on some network structure. Interactions determine reproductive fitness, which is played out locally. Or, for cultural evolution, payoffs drive local differential imitation rather than differential reproduction. The shift from random mixing to local interaction completely changes the dynamics (Pollock 1989; Nowak and May 1992; Eshel et al. 1998; Alexander and Skyrms 1999; Skyrms 2004; Alexander 2007).

In Stag Hunt games, it is possible for a few contiguous cooperators to grow and take over the whole population. But it is also possible for the expansion of cooperation to get stuck, so patches of cooperators and patches of non-

cooperators coexist. It depends on the payoffs, with different Stag Hunts yielding quite different dynamics. (This is investigated in detail in Alexander 2007.) Bearing Spinoza in mind, we see this as a plausible result. Self-generating patches of different social norms may appear naturally. But changing the payoffs can tip the dynamics in the direction of everyone hunting stag.

In the foregoing, the interaction neighborhood, where the game is played with neighbors, and the imitation neighborhood, in which payoffs are observed and successful strategies are imitated, are the same. (For biological evolution, imitation neighborhood corresponds to the neighborhood in which progeny are dispersed.)

Eshel et al. (1998) point out that these neighborhoods need not be the same, and that differences can have significant effects on the evolutionary dynamics. If the imitation neighborhood is larger than the interaction neighborhood, so that hare hunters can see inside a group of successful stag hunters and then imitate their success, the phenomenon of a small group of contiguous stag hunters taking over a large population of hare hunters becomes robust (Skyrms 2004).

### Adaptive Networks

Interaction networks need not be fixed. They can coevolve with the strategies played in interactions with neighbors on the network. Suppose that a small group of individuals, some stag hunters and some hare hunters, begin by interacting at random to play Stag Hunt games. And suppose that they then modify their probabilities of choosing with whom to interact by simple reinforcement learning based on their payoffs. Such a model is investigated in Skyrms and Pemantle (2000) for a two-person Stag Hunt game. (This is extended to three-person Stag Hunts in Pemantle and Skyrms 2004.) It is shown reinforcement leads players to self-segregate, so that stag hunters always meet stag hunters and hare hunters always meet hare hunters. Then a strategy-revision dynamics is added. At random times individual players decide to imitate the strategy with the greatest average success in the population.

The outcome of this co-evolutionary process depends on the relative rates of the two adjustment processes. Fast network dynamics and slow strategy revision leads to cooperation taking over the population. Stag hunters find each other, prosper and are imitated. Converts to stag hunting quickly associate with other stag hunters. Slow network dynamics is sluggish in moving away from random interaction. Then, more often than not, it is the hare hunters that fare better and are imitated. (For further development see Santos et al. 2006).

Several laboratory experiments have been run in which individuals have been given opportunities to change those

with whom they associate between rounds of playing a game, e.g., Page et al. (2005). The closest experiment to the model of Skyrms and Pemantle was recently carried out by Rand et al. (2011). Results of the experiments support the theoretical predictions. Fluid networks favor cooperation.

### Bargaining

If individuals have cooperated to produce a surplus, they then need to decide how to divide it. We model this as a Nash bargaining game. This abstracts from the sequential structure of bargaining by focusing on the players' bottom-line demands.

Each demands some fraction of the good to be divided. If their demands do not exceed the whole amount of the good, they get what they demand. Otherwise no bargain is struck and they get nothing. Other bargaining games could be considered, but I will focus on this one here.

In fact, I will consider only the case where the players' positions are completely symmetrical. Other cases may be more interesting, but this one already presents its own puzzles. We consider discrete versions of the game, so that there are only a finite number of possible demands. For example, if the possible demands are  $1/3$ ,  $1/2$ ,  $2/3$ , the payoffs of row against column are:

	$1/3$	$1/2$	$2/3$
$1/3$	$1/3$	$1/3$	$1/3$
$1/2$	$1/2$	$1/2$	0
$2/3$	$2/3$	0	0

There is one evolutionarily stable strategy, demand  $1/2$ . But there is another evolutionarily stable state of the population, in which half of the population demands  $1/3$  and half demands  $2/3$ . If we let possible demands draw a finer line, then more evolutionarily stable states appear. For instance, there is an evolutionarily stable state where most of the population demands 1 % and a few demand 99 %. These evolutionarily stable states are attractors in the replicator dynamics applied to the discrete bargaining game.

That is a model of pure differential reproduction in a large random mixing population. Let us see how the modifications of the assumptions that we applied to the Stag Hunt affect this bargaining game. They all make a difference.

### Noise

Let us first add a little noise, due to mutation in a finite population, or alternatively due to some sort of exogenous

shocks to the differential reproduction. In the Stag Hunt game this led to the devolution of the social contract. Here, in this symmetric Nash bargaining game, it leads to the equal split. The equal split—each demands half—is the unique stochastically stable state. (See Young 1993 for a more precise statement and proof of the result.)

### Signals

Pre-play signals favor the equal split. If only 3 pre-play signals are added to the mini-bargaining game introduced above and the population evolves according to the standard replicator dynamics, then the basin of attraction of the equal-split equilibrium expands from 62 to 98+ % of the space of possible population proportions (Skyrms 2004).

### Local Interaction

Instead of bargaining with strangers, individuals may bargain with neighbors on a square lattice. Then just a small cluster of demand- $1/2$  players will expand and take over the whole population. Players who demand more than  $1/2$  get in each other's way and rapidly die out. Players who demand less get less. This result is robust over various discretizations of the bargaining game (Alexander and Skyrms 1999; Alexander 2007).

### Adaptive Networks

Here, as before, timing is of essential importance. Greedy players—those who demand more than  $1/2$ —will form links to modest players with whom their demands are compatible. But then modest players will see greedy players doing well and switch to being greedy themselves. Demand- $1/2$  players will link with both modest players and other demand- $1/2$  players. They have more opportunities to link with compatible players. But successful greedy players may tempt them also to switch to greedy strategies. When the modest players convert, greedy players can find no compatible partners. Depending on timing, any of the strategies may be the last one standing. There is no clear-cut advantage for the equal split here. (See Alexander 2007 for further discussion and for N-player bargaining.)

### Experiments and Empirical Data

In human laboratory experiments where players are in a symmetric situation, the equal split is exceptionally robust (Nydegger and Owen 1974; Roth and Malouf 1979; Van Huyck et al. 1995).

Of course perfectly symmetrical positions represent a very special case. But the equal split seems to survive as a norm across a range of situations where we do not have

such symmetry. See, for instance, the study of sharecropping arrangements in Young and Burke (2001). However, much of the experimental literature investigates asymmetries in bargaining situations that produce asymmetrical divisions. For a review see Roth (1995).

### Stag Hunt + Bargaining

Cooperation to produce a surplus and division of the surplus may coevolve. This suggests analysis of a composite game. Players can either hunt hare or hunt stag. If at least one chooses to hunt hare, no successful stag hunt is possible—and the payoffs are as in the Stag Hunt game. But if both choose to cooperate in a stag hunt, they enter a subgame in which they bargain over how to divide the stag. This composite extensive form game can be analyzed in all the ways that we have analyzed its composite pieces. The first place to start is with simple replicator dynamics. Wagner (2012) does just this, and shows that in this setting, evolution of Stag Hunting + Equal Split can be easier than either Stag Hunting or Equal Split when considered separately.

### Conclusion

The foregoing are just elementary considerations concerning natural social contracts. However, two important general themes are illustrated. One is the importance of the move from random interactions to correlated interactions, where one's probability of meeting a strategy is dependent on one's own strategy. Signaling, local interaction, and dynamic networks all serve as basic correlation devices. There are others. A developed society is full of institutions that generate correlation. Correlation plays a central role in shaping the social contract.

The second theme is the importance of dynamics. Static analysis tells us very little about these interactions. There are typically multiple stable equilibria. To get any understanding about equilibrium selection we need to look to the dynamics. Sometimes the interaction of several dynamical processes is important.

Dynamics interacts with correlation. Correlation is not static, but waxes and wanes according to the dynamics of interaction. In some cases there is no correlation left at the end of evolution, with one strategy taking over the population. But transient correlation plays a part in determining which strategy takes over. In other cases, persistent correlation sustains a polymorphic population. In a developing theory of natural social contracts I believe that dynamics and correlation will continue to be of central importance.

### References

- Alexander JM (2007) *The structural evolution of morality*. Cambridge University Press, Cambridge
- Alexander JM, Skyrms B (1999) Bargaining with neighbors: is justice contagious? *J Philos* 96:588–598
- Blume A, Ortmann A (2007) The effects of costless pre-play communication: evidence from games with Pareto-ranked equilibria. *J Econ Theory* 132:274–290
- Cole T (1967) *Democritus and the sources of Greek anthropology*. American Philological Association, Cleveland
- Cooper RC, De Jong D, Forsythe R, Ross T (1992) Communication in coordination games. *Quart J Econ* 107:771–793
- Eshel I, Samuelson L, Shaked A (1998) Altruists, egoists and hooligans in a local interaction model. *Am Econ Rev* 88:157–179
- Foster D, Young HP (1990) Stochastic evolutionary game dynamics. *J Theor Biol* 38:219–232
- Hume D (1739) In: Norton DJ, Norton MJ (eds) (2011) *A treatise of human nature*. Oxford University Press, Oxford
- Kandori M, Mailath G, Rob R (1993) Learning, mutation and long-run equilibria in games. *Econometrica* 61:29–56
- Nowak M, May RM (1992) Evolutionary games and spatial chaos. *Nature* 359:826–829
- Nydegger RV, Owen G (1974) Two-person bargaining: an experimental test of the Nash axioms. *Int J Game Theory* 3:239–249
- Pacheco J, Santos F, Souza M, Skyrms B (2009) Evolutionary dynamics of collective action in N-person Stag Hunt dilemmas. *Proc R Soc B* 276:315–321
- Page T, Putterman L, Unel B (2005) Voluntary association in public goods experiments. *Econ J* 115:1032–1053
- Pemantle R, Skyrms B (2004) Network formation by reinforcement learning: the long and the medium run. *Math Soc Sci* 48:315–327
- Pollock G (1989) Evolutionary stability of reciprocity in a viscous lattice. *Soc Netw* 11:175–212
- Rand D, Arbesman S, Christakis N (2011) Dynamic networks promote cooperation in experiments with humans. *PNAS* 108:19193–19198
- Robson A (1990) Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *J Theor Biol* 144:379–386
- Roth AE (1995) Bargaining experiments. In: Kagel J, Roth AE (eds) *Handbook of experimental economics*. Princeton University Press, Princeton, pp 253–348
- Roth AE, Malouf M (1979) Game theoretic models and the role of information in bargaining. *Psychol Rev* 86:574–594
- Rousseau JJ (1755) *Discourse on inequality*. Oxford University Press, Oxford
- Santos FJ, Pacheco J, Lenaerts T (2006) Cooperation prevails when individuals adjust their social ties. *PLoS Comput Biol* 2(10):e140
- Santos F, Pacheco J, Skyrms B (2011) Coevolution of pre-play signaling and cooperation. *J Theor Biol* 274:30–35
- Skyrms B (2002) Signals, evolution and the explanatory value of transient information. *Philos Sci* 69:407–428
- Skyrms B (2004) *The Stag Hunt and the evolution of social structure*. Cambridge University Press, Cambridge
- Skyrms B, Pemantle R (2000) A dynamic model of social network formation. *Proc Natl Acad Sci USA* 97:9340–9346
- Taylor M, Ward H (1982) Chickens, whales and lumpy goods: alternative models of public goods provision. *Political Stud* 30:350–370
- Van Huyck J, Battalio R, Mathur S, Ortmann A, Van Huyck P (1995) On the origin of convention: evidence from symmetric bargaining games. *Int J Game Theory* 24:187–212
- Verlinsky A (2005) Epicurus and his predecessors on the origin of language. In: Frede D, Inwood B (eds) *Language and learning*:

- philosophy of language in the Hellenistic age. Cambridge University Press, Cambridge, pp 56–100
- Wagner E (2012) Evolving to divide the fruits of cooperation. *Philos Sci* 79:81–94
- Young HP (1993) An evolutionary model of bargaining. *J Econ Theory* 59:145–168
- Young HP, Burke M (2001) Competition and custom in economic contracts: a case study in Illinois agriculture. *Am Econ Rev* 91:559–573