

**An Evolutionary Psychology Examination of the
'Tragedy of the Commons'**

Simon Jonas Hadlich

Psychology

Dr. Eddy de Bruijn

March 8, 2011

An Evolutionary Psychology Examination of the 'Tragedy of the Commons'

"Freedom in a commons brings ruin to all", Hardin (1968) concluded on the fate of common-pool resources. His statement has become a dogma for many, with far-reaching consequences in various areas of policy, but particularly environmental protection. In recent decades, it has been widely used as an argument in support of neoliberal privatization of common-pool resources. More recently, however, the Nobel Prize given to Elinor Ostrom has shed light on her research highlighting successful management of common-pool resources as commons, and cast doubt on the inevitability of tragedy in Hardin's scenario.

Undoubtedly, psychological factors are highly relevant for the feasibility and efficiency of different property rights regimes, including the commons. Hardin's conclusion is based on the fact that natural selection favours selfishness, while it seemingly cannot explain altruistic behaviour. However, since Hardin's essay was first published there have been manifold advances in the study of human cooperation. Trivers (1971) has modelled the circumstances under which reciprocal altruism can be selected for, and Axelrod and Hamilton (1981) have shown via game theory that cooperation can be an evolutionary stable strategy in the scenario described by Hardin.

This essay examines Hardin's 'tragedy of the commons' from the perspective of evolutionary psychology. It focuses on the question whether the tragedy, as Hardin predicts, is inevitable, and reviews evidence that cooperative behaviour is rooted in evolutionary adaptations. In the following, I will present the analytical framework of this essay, encompassing Hardin's (1968) initial claims, as well as Axelrod and Hamilton's (1981) theory of the evolution of cooperation. I will then review evolutionary adaptations which are relevant for managing common-pool resources, in particular selfishness, reciprocal altruism, and mutual coercion.

ANALYTICAL FRAMEWORK

The 'Tragedy of the Commons'

In his essay, "The Tragedy of the Commons", Hardin (1968) paints a bleak future for any resource open for all to use. He asks his readers to imagine "a pasture open to all", used by herdsmen to graze their cattle. At a certain point, an equilibrium will be reached between the growth of grass and its loss due to grazing. But in trying to maximize his individual gain, Hardin argues, each herdsmen will add more and more cattle to his herd, inevitably leading to overexploitation of the commonly held resource. "Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons", Hardin warns, concluding that "[f]reedom in a commons brings ruin to all."

Hardin argues that this tragedy arises because each herdsmen, seeking to maximize his gain, will evaluate the utility of adding further cattle to his herd. He will come to the conclusion that this utility has a negative and a positive component: The value of having another animal versus the increased depletion of the commonly shared pasture. The individual herdsman will gain all the benefits from adding one more animal, whereas the costs will be born by the community. From the perspective of the individual herdsman, the positive component ("+1") of the utility will thus always outweigh the negative component ("-1/n"). Hardin argues that this makes the addition of further cattle the only "sensible" choice for each herdsman, but because this conclusion is reached by all members of the community, the commonly shared resources will inevitably be depleted. "Therein is the tragedy", Hardin writes, "[e]ach man is locked into a system that compels him to increase his herd without limit - in a world that is limited."

Defining Common-pool Resources

In order to evaluate Hardin's arguments, it is important to distinguish between the intrinsic nature of a resource and the property regime under which it is held. Feeny et al. (1990) provide a useful definition of common-pool (or "common-property", as they call them) resources and the regimes under which

these can be held. They characterize common-pool resources through two features. The first is a lack of excludability, which is to say that restricting access to a resources is costly to an extent that it is not feasible. The second is subtractability, or rivalry, which denotes that exploitation of a resource by one user adversely affects the ability of others to make use of it. However, there are good arguments for treating virtual goods such as knowledge as non-rivalrous common-pool resources (but this distinction is of minor relevance for the subject of this essay). Furthermore, the authors define four categories of property rights which can be applied to common-pool resources. These are free access - a lack of property rights -, private property, communal (or common) property - owned by a clearly defined group excluding outsiders -, and state property.

Hardin describes "commons" as rivalrous (subtractable) resources open to access by all, i.e. not subject to well-defined property rights. This has led to considerable confusion, as it confounds the intrinsic nature of the resource and the regime under which it is held, and applies the term "commons" wrongly. As Hardin later acknowledged, much argument could have been avoided had he initially written of the "unmanaged commons". More exactly, the subject of his essay are rivalrous common-pool resources held under an open access regime.

Game Theory: The Evolution of Cooperation

Game theory has been proven to be particularly useful to formalize models of human cooperation (Axelrod & Hamilton, 1981). As it turns out, the tragedy of the commons is a case of one of the most well-known games, the 'prisoners' dilemma' (Cartwright, 2008, p. 207). In this model, two players each have the option to either cooperate or to defect, with outcomes depending on both their actions. The term 'prisoner's dilemma' describes a particular application of this game's logic. Assume that two suspects are questioned separately by the police. The existing evidence is only sufficient to sentence both for a minor part of the crime, so they are individually offered freedom in case they turn 'king's evidence' and inform (defect) on the other. If both defect, they will receive full jail time for the shared crime, while in case they both cooperate, each receives a smaller sentence.

The prisoners' dilemma illustrates that rational behaviour in Hardin's sense - i.e. maximizing one's own gains - can lead to the worst possible outcome (Cartwright, 2008, p. 202), while cooperation would yield a far more favourable result. Hardin's (1968) tragedy of the commons extends the classical prisoners' dilemma to a multi-person game theoretical problem. If we apply the game logic of the prisoners' dilemma to his metaphorical village, each herdsman, by maximising the count of his cattle, in defecting on his peers; and as each herdsman defects, in the end all are worse off.

However, Axelrod and Hamilton (1981) have shown that iterative games can lead to different results than one-off situations, and have argued that this better represents human social life (p. 1392). They argue that, if players meet repeatedly, they can learn to benefit from the greater rewards cooperation yields, i.e. there would be a way for cooperation to evolve. In a contest aimed at finding an 'evolutionary stable strategy' (ESS) - which is resistant to displacement by alternative strategies if pursued in a population - it occurred that 'tit for tat' provides a powerful strategy. With this strategy, a player would always cooperate, except to retaliate once in the face of a previous defection. While 'tit-for-tat' can never win over an opponent, it only loses with small margins, whereas more hawkish strategies (such as "always defect"), when fighting each other, accumulate high negative scores (Cartwright, 2008, pp. 203-204). Axelrod and Hamilton (1981, pp. 1394-1395) conclude that "if the probability that interaction will continue is great enough, then TIT FOR TAT is itself evolutionary stable" (emphasis original), which is to say that "cooperation based on reciprocity can get started in a predominantly noncooperative world, can thrive in a variegated environment, and can defend itself once fully established".

MANAGING COMMON-POOL RESSOURCES

Human Selfishness

Underlying Hardin's (1968) reasoning on the tragedy of the commons is an understanding of human behaviour as inherently selfish. Indeed, evolutionary psychology lends much credibility to the assumption that humans generally act

selfish. Behind this behaviour is natural selection. Contrary to what is sometimes believed, selection at the level of populations or whole species is marginal compared to selection at the individual or, more correctly, gene level. As Axelrod and Hamilton (1981, p. 1390) make clear, there is "no sound basis for a pervasive group-benefit view of selection". Subsequently, humans can be expected to always act in a way that promotes their reproductive success, i.e. which increases the survival chances of their genes¹.

It is worth noting that this selfishness is located at the gene level, which is to say that the unit of selection is not the individual human (who acts merely as a "vehicle"), but the gene ("replicator"). One consequence is kin selection: individuals might behave in a way that enhances the reproductive success of close kin, at the cost of their own reproductivity. According to inclusive fitness theory, this will be the case when the cost (in reproductive success) to the helping gene is more than compensated for by the potential increase of success of the gene in its close kin (Cartwright, 2008, pp. 42-44).

Reciprocal Altruism

While altruism at the gene level seems implausible given this selfishness rooted in natural selection, Trivers (1971) has shown that reciprocal altruism even between members of different species can well be selected for, without assuming a group advantage. Reciprocal altruism describes a form of time-delayed discrete mutualism, i.e. aid is given to another in hope that it will be returned. As Trivers writes, "under certain conditions natural selection favors these altruistic behaviors because in the long run they benefit the organism performing them."

Trivers gives the example of a drowning man whose chance of survival is greatly enhanced (from 50% to 95%) if another man comes to his rescue (at a risk of 5% of drowning himself). In an isolated incident, "it is clear the rescuer should not bother to save the drowning man" (p. 36), but if the drowning man later reciprocates at similar odds of survival, "[e]ach participant will have traded a

¹ Or, to be more correct: Humans will act in a way that would have promoted their reproductive success in the environment of evolutionary adaptedness (EEA), 20000-40000 thousand years ago. Maladaptive behaviour, i.e. adaptations which are detrimental to reproductive success in our modern environment, further complicate the picture.

one-half chance of dying for a one-tenth chance” (ibid.). This scenario is also crucial for Axelrod and Hamilton's (1981) later development of their theory on the evolution of cooperation.

Cartwright (2008, p. 200) gives three conditions for reciprocal altruism to occur. First, reciprocity should be granted, i.e. there needs to be a reasonable chance of meeting aid receiver again. Second, reciprocal altruists need to be able to recognize each other and to detect free-riders who don't return altruistic acts. And third, "the ratio 'cost to donor/benefit to receiver'" must be low to make reciprocal altruism feasible even when reciprocation is uncertain.

If these conditions are given, Trivers (1971, p. 36) argues that cheating (i.e. non-reciprocative behaviour) can be selected against if it has negative consequences for the cheater, e.g. because altruists respond to this by curtailing all altruistic behaviour towards this individual. As Trivers (ibid.) writes, [a] assuming that the benefits of these lost altruistic acts outweigh the cost involved in reciprocating, the cheater will be selected against relative to individuals who, because neither cheats, exchange many altruistic acts.

Trivers (1971) argues selection will favour a “complex, regulating system” for reciprocal altruism. He finds that some features of the human psychological system might be “important adaptations to regulate the altruistic system”. Each individual has altruistic as well as cheating tendencies, whose expression is dependent of environmental variables (p. 48), and normally distributed.

Friendship, Triver writes, means “liking those who are themselves altruistic”, and will be selected for “as the immediate emotional rewards” which motivate the forming of altruistic partnerships. This means that emotions of friendship and hatred “may evolve *after* a system of mutual altruism has appeared” (emphasis original), because they help regulating the system.

Moralistic aggression acts as “protective mechanism” for altruists who are vulnerable to cheaters exploiting their positive emotions. According to Trivers (1971, p. 49), it serves the needs for a counterbalance to altruistic tendencies in case of lacking reciprocation as well as a means of “educating the unreciprocating individual”, and perhaps of even directly selecting against it.

Gratitude, Trivers (ibid.) argues, has been selected for as a means of

regulating humans response to altruistic acts with regard to its cost/benefit ratio, as well as “a function of the plight of the receiver”, signalling possible high reciprocity.

Guilt can be seen as a function of “reparative altruism” (Trivers, 1971, p. 50). Since cheating has negative consequences (the cut-off of an altruistic, i.e. beneficial, relationship), the cheater will try to avoid these (which is also beneficial to the individual cheated on). As Trivers finds, the cheater “should be selected to make up for his misdeed [and] to make a reparative gesture”.

While selection favours behaviours that support the altruistic system, it might also lead to the development of *subtle cheating*, i.e. the ability to mimic signals of altruism, e.g. through “sham guilt” (ibid.). Trivers (1971, p. 50-51) argues that this favour the development of detection mechanisms, such as “distrusting those who perform altruistic acts without the emotional basis of generosity or guilt”.

Mutual Coercion

Trivers (1971, p. 52) also argues that, as humans live in close-knit social communities, selection should favour multi-party interactions, including learning from others' experiences, mutual help against cheaters, and the formation and regulation of multi-party exchange systems. In particular, there is strong evidence that evolutionary adaptations also support mutual coercion as a means of keeping free-riders from exploiting common-pool resources at the disadvantage of others. Hardin (1968) himself initially proposed the use of "mutual coercion, mutually agreed upon by the majority of the people affected" as a means of managing common-pool resources. Indeed, social pressure in form of e.g. shame and moralizing is a can work to prevent the tragedy of a commons. For example, Milinski et al. (2002) have shown that people were more likely to conserve a commonly held resources when their reputation was at stake.

Human cooperation relies on indirect reciprocity to enforce individual restraint, for example through reputation and social pressure. Mutual coercion is based on each individuals interest to detect free-riders who don't reciprocate altruistic acts or defect otherwise on their peers. While evolution has made

humans selfish, it has also made them susceptible to social pressure in various forms, providing a regulatory mechanisms to prevent free-riding. When resources are governed as commons, in particular, these mechanisms are institutionalized to protect the resources from overexploitation.

CONCLUSION

This paper has shown that the tragic end of each common-pool resource does not have to be. While natural selection made humans (and indeed all organisms) inherently selfish, it has also equipped us with capacities that enable cooperation and make a successful management of common-pool resources possible.

Axelrod and Hamilton (1981) have provided evidence that an evolutionary stable strategy (ESS) of interactions might be based on cooperation, i.e. that humans could have learned to behave cooperatively. This is based on the assumption that cooperative behaviour yields higher gains than defection (i.e. provide a fitness benefit), as assumed in Hardin's outline of the tragedy of the commons.

Parts of the human psychological system might have evolved to support means of cooperative behaviour. Reciprocal altruism and mutual coercion are based on humans' susceptibility to social pressure. Cooperation is enabled by the ability to detect reciprocating altruists as well as free-riders, which is based on emotional responses.

The tragedy of the commons is not inevitable. Selfishness is selected for, but so are traits which enable cooperative behaviour - in humans just as well as in other organisms. Institutions and property rights which take these mechanisms into account can be well-suited to govern common-pool resources and prevent their overexploitation.

BIBLIOGRAPHY

Axelrod, R., & Hamilton, W.D. (1981). The Evolution of Cooperation. *Science*

21(1), 1390-1396.

Cartwright, J. (2008). *Evolution and Human Behaviour, 2nd ed.* Basingstoke & New York, NY: Palgrave Macmillan.

Feeny, D., Berkes, F., McCay, B.J., & Acheson, J.M. (1990). The Tragedy of the Commons: Twenty-Two Years Later. *Human Ecology* 18(1), 1-19.

Hardin, G. (1968). The tragedy of the commons. *Science* 162, 1243-1248.

Milinski, M., Semmann, D., Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature* 415, 424-426.

Trivers, R.L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology* 46(1), 35-57.