



Article

Michael Vlerick

The Evolution of Social Contracts

<https://doi.org/10.1515/jso-2019-0041>

Abstract: Influential thinkers such as Young, Sugden, Binmore, and Skyrms have developed game-theoretic accounts of the emergence, persistence and evolution of social contracts. Social contracts are sets of commonly understood rules that govern cooperative social interaction within societies. These naturalistic accounts provide us with valuable and important insights into the foundations of human societies. However, current naturalistic theories focus mainly on how social contracts solve coordination problems in which the interests of the individual participants are (relatively) aligned, not competition problems in which individual interests compete with group interests (and in which there are no group beneficial Nash equilibrium available). In response, I set out to build on those theories and provide a (more) comprehensive naturalistic account of the emergence, persistence and evolution of social contracts. My central claim is that social contracts have culturally evolved to solve cooperation problems, which include both coordination and competition problems. I argue that solutions to coordination problems (which I spell out) emerge from “within-group” dynamics, while solutions to competition problems (which I also spell out) result largely from “between-group” dynamics.

Keywords: Social contracts; Game theory; Cultural evolution; Coordination problems; Competition problems; Within-group dynamics; Between-group dynamics; Cultural group selection.

1 Introduction

In important and influential work, Young (1998), Sugden (2005), Binmore (1994, 1998, 2005, 2007) and Skyrms (1996, 2004) have developed naturalistic accounts

Michael Vlerick, Department of Philosophy, Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands; University of Johannesburg, Auckland Park, South Africa,
e-mail: michaelvlerick@gmail.com, m.m.p.vlerick@uvt.nl
<https://orcid.org/0000-0002-0664-1509>

Open Access. © 2020, Michael Vlerick, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 Public License.

of the emergence, persistence and evolution of social contracts. Social contracts are sets of commonly understood rules that govern cooperative social interaction within societies. These rules range from commonly known but implicit rules of interaction (such as the obligation to share meat after a successful hunt in hunter-gatherer societies) to explicit laws. Social contracts are the backbones of societies: they underlie social interaction within the group and orchestrate the unrivaled large-scale and flexible forms of cooperation humans engage in.

The chosen tool to analyze the emergence and cultural evolution of social contracts or social structure is evolutionary game theory (see also Gintis 2009). Game theory enables us to model the outcome of strategic interaction both in human and non-human groups (Maynard-Smith and Price 1973). It is, as Gintis (2009) points out, an indispensable tool in the toolbox of the social scientist. This evolutionary approach, I believe, is valuable and important. It seeks to provide a scientific understanding of the foundations of human societies (see also Vlerick 2016).

Drawing from David Lewis' (1969) seminal work "Convention", Young, Sugden, Binmore and Skyrms argue that social contracts emerge spontaneously out of the social interaction within the group and can only persist in Nash equilibria of what Binmore (1994, 2005, 2007) calls the "game(s) of life" we play with each other. This is a key insight and the central point of naturalistic theories of social contracts. In such equilibria, everybody has a best response to what everybody else is doing.

While current game-theoretic accounts of social contracts have much merit, they focus primarily on how evolutionary dynamics drive the interaction towards group beneficial equilibria *available* in the various games of life being played within the group. This leaves us with a somewhat incomplete account of what social contracts do. It provides an account of how social contracts solve *coordination problems* in which the interests of the individual participants are (relatively) aligned, not *competition problems*¹ in which individual interests compete with group interests (and in which there is therefore no group beneficial Nash equilibrium available).

Tellingly, Binmore defines a social contract as a "self-policing agreement between members of society to *coordinate* on a particular equilibrium in the game of life" (1994, p. 35, my italics) and as "sets of common understandings that allow

¹ What I call a "competition problem" is often referred to as "the problem of cooperation" (or "the hard problem of cooperation") or the "free-rider problem" in the literature (e.g. Eriksson and Strimling 2012). I prefer to use the general term "cooperation problem" to include both coordination problems and free-rider or competition problems, since both kinds of problems need to be solved in order to cooperate successfully, not merely the latter kind.

the citizens of a society to *coordinate* their efforts” (2005, p. 3, my italics). In a similar vein, Skyrms (2004, p. vii) argues that prisoner’s dilemma problems – the archetypical competition problems – were not typically encountered in human (evolutionary) history. It is the “stag hunt” – which is a coordination problem – he claims, that is “the key to the evolution of cooperation, collective action, and social structure”.

The aim of this paper is to build on those important naturalistic theories and provide a (more) comprehensive naturalistic account of the emergence and evolution of social contracts by including how social contracts solve competition problems and why these solutions emerged in human societies. My central claim, spelled out in Section 2, is that social contracts have culturally evolved to solve cooperation problems. Cooperation problems include both coordination problems – which, as pointed out, have been getting most of the attention – and competition problems. In Sections 3 and 4, using game-theoretic models and drawing from various strands of empirical research, I will outline how social contracts solve coordination problems and how they solve competition problems. Both kinds of problems, as I will show, are solved in a different way.

Finally, in Section 5, I will outline the underlying cultural dynamics that drive the emergence and evolution of social contracts. Those are “within” and “between-group” dynamics. The former drive social contracts to stable points in the games of life – (Nash) equilibria – while the latter select efficient equilibria. Between-group dynamics are the result of competition between groups. They have driven the cultural evolution of social contracts in societies *directly* and the genetic evolution of human (social) psychology *indirectly* through what Richerson and Boyd (1985, 2005) have called “gene-culture co-evolution”.

2 Cooperation Problems

2.1 What are Cooperation Problems?

Cooperation problems are problems that a group of individuals needs to overcome in order to reap the benefits of mutual cooperation. They occur in positive sum games in which the sum of the payoff of the players increases when they cooperate. What could stand in the way of cooperation (the cooperation problems) is the failure of individuals to coordinate their actions and/or the failure to prevent individuals from free-riding (not contributing to the cooperative effort). In the first case, we have a coordination problem. In the second case, a competition problem.

Cooperation problems do not include “conflict problems” in which the interests of the individuals involved are diametrically opposed to one another. In such “zero sum games” one individual’s gain is balanced by another individual’s loss (hence “zero sum”: the sum of the payoffs is zero). Consequently, cooperation can never emerge. Take the “matching pennies game” for example. In this game, two participants are asked to reveal a penny simultaneously. If the pennies match – i.e. both have heads or tails facing up – player 1 keeps both pennies. If they do not, player 2 keeps both pennies. In such a context, the players – provided that they are rational and act out of self-interest – will never cooperate (e.g. by coordinating their choices to yield matching or non-matching pennies).

2.2 Coordination vs. Competition Problems

A coordination problem arises when there are different ways to produce a commonly desired outcome and the participants need to agree on how to proceed. The typical example of a pure coordination problem is the “driving game”. There are two optimal ways in which traffic can be regulated: either everybody drives on the left of the road or everybody drives on the right of the road. Drivers do not care either way (both solutions are equally optimal) but they all want to coordinate on one of those two solutions (i.e. all drive on the same side of the road).

A competition problem arises when the immediate interest of individuals within the group is pitted against (competes with) the interest of the group (or the compounded interest of all the individuals within the group). In these contexts, participants are better off on average and in total when all participants cooperate, but each individual participant is better off when he does not cooperate (when he “defects” in game-theoretic parlance). Two famous “games” that illustrate competition problems are the “prisoner’s dilemma” and the “tragedy of the commons”.

The prisoner’s dilemma goes as follows. Two suspects of a crime are arrested. They are kept in different cells and interrogated separately. Each member has a choice to betray the other by testifying that the other committed the crime or to cooperate with the other by remaining silent.

- If A and B both betray the other, they both serve 2 years in prison
- If A betrays B but B remains silent, A will be set free and B will serve 3 years in prison (and vice versa)
- If A and B both remain silent, both of them will only serve 1 year in prison

This gives us the following payoff matrix:

	Cooperate	Betray
Cooperate	-1, -1	-3, 0
Betray	0, -3	-2, -2

The best outcome on average and in total for both players is the one in which they both cooperate (remain silent) and serve just 1 year in prison. However, each player always has an incentive to betray the other. If the other does not betray him, he goes free instead of serving 1 year in prison. If the other also betrays him, he serves 2 years in prison instead of three. When both players act rationally and out of self-interest, they will both betray each other. Mutual defection [-2, -2] is the Nash equilibrium of the prisoner's dilemma – the outcome in which no player can improve his payoff by unilaterally changing his strategy – but it is not the optimal outcome, which is mutual cooperation [-1, -1].

The same problem occurs in the tragedy of the commons (Hardin 1968). Suppose there is a pasture that can sustain a hundred sheep. Ten shepherds use the pasture and each shepherd has ten sheep. Adding more sheep would cause the pasture to be overgrazed and ultimately turn into a wasteland. Nevertheless each shepherd will still benefit from adding another sheep. This generates a substantial increase in income (10% more wool) and only a relatively small cost (1% less grass for his sheep). Shepherds aiming to maximize their individual payoff will therefore increase their livestock beyond the capacity of the pasture leading to the demise of this common good. The problem is that the cost of adding another animal is shared by the entire group, while the benefit of adding another animal is reaped by each individual shepherd. As in the prisoner's dilemma, individual interests clash with group interests. Everybody is better off if everybody cooperates, but each individual is always better off by defecting (betraying in the prisoner's dilemma and adding livestock in the tragedy of the commons).

2.3 Solving Cooperation Problems with Social Contracts

Real world examples of solutions to coordination and competition problems abound. Traffic rules, signs and lights are all successful solutions to coordinate traffic. Social conventions, such as the convention that the caller is to call back when a phone call is interrupted (avoiding the undesirable situations in which both call back at the same time or wait in vain for the other to call back), is an equally effective solution to a coordination problem (Lewis 1969, p. 5). Finally,

even language is a solution to a coordination problem (Lewis 1969). By implicitly agreeing to attach a certain meaning to certain utterances or sets of symbols, we are able to realize the commonly desired outcome of successful communication.²

Successful solutions to competition problems are equally prevalent in societies. Throughout human history, groups have imposed rules on how individuals may and may not use common goods or common-pool resources in order to prevent a tragedy of the commons (Ostrom 1990). Think of restrictions on the use of grazing pastures (e.g. no allowing grazing in certain periods of the year) in pastoral societies, rules to maintain water reserves in arid regions, and rules to keep city parks and neighborhood playgrounds clean. Today, climate change presents us with the threat of a tragedy of the commons on a global scale and people around the world are trying hard (albeit it with somewhat limited success so far) to prevent this by imposing worldwide rules on the emission of CO₂ and other greenhouse gasses.

The same goes for prisoner's dilemma situations. Sport federations regulate the use of performance enhancing substances. If they would not do so, athletes would be tied in an arm's race to use substances to get an edge over their competitors. This would be costly for the athletes and could have a negative impact on their health. By regulating the use of substances and punishing infractions, federations try to prevent that "the game" gravitates towards the Nash equilibrium of a prisoner's dilemma in which everybody is worse off: a level playing field in which all athletes dope. Similarly, international treaties aimed at curbing the proliferation of weapons of mass destruction try to prevent a situation in which countries have to build up their arsenal in response to other (possibly antagonistic) nations doing so. Again, the result of such an arm's race would make all parties worse off: countries would have to bear important economic costs and – worse – increases the likelihood of being the victim of a "nuclear Armageddon". Finally, enforcing property prevents individuals from adopting predatory strategies to acquire goods by ruse or by force, making society better off.³

² Note that we also solve coordination problems on the fly. When doing so, we do not rely on pre-established rules or agreements. Think of the way people form a line and pass buckets with water to extinguish a fire, or send and fill in a doodle to determine a place and time for a meeting, or divide tasks on the spot when jointly cooking a meal that all will enjoy.

³ Technically this is a solution not to a prisoner's dilemma but to another competition game namely a "hawk dove game" or a "chicken game" which has a mixed Nash equilibrium in which part of the population adopts the predatory strategy (the "hawk" strategy) and the other part the cooperative strategy (the "dove" strategy). See Vlerick (2016).

All of these solutions to cooperation problems come from social contracts: sets of commonly recognized rules or norms in a group (a society) that govern certain domains of interaction. They enable people to coordinate their actions to achieve commonly desired outcomes (solving coordination problems) and protect the group interest from free-rider behavior (solving competition problems). In the next two subsections, I will outline in more detail how social contracts solve both kinds of cooperation problems.

3 Solving Coordination Problems

3.1 Signaling

Coordination problems are epistemic problems. They are typically solved by communication. Since the interests of senders and receivers are aligned, successful cooperation depends on the successful transfer and interpretation of information. If senders and receivers communicate well, cooperation follows. Solving coordination problems is by no means the prerogative of humans. In fact, at all levels of biological organization organisms have evolved to solve coordination problems. They do so by signaling: transferring information to each other (Skrvms 2010, p. 6). Female baboons signal their fertility to male baboons with a red, swollen bottom. Bees signal the presence and location of food to other bees in the hive with an intricate dance. Even lifeforms as simple as bacteria have developed complex signaling systems (e.g. quorum sensing) to coordinate their behavior (e.g. to coordinate gene expression in the light of the density of the population).

These signaling systems evolve because they benefit both senders and receivers. They create pathways of information that enable senders and receivers to coordinate their actions and increase their respective fitness (their chances of successful reproduction). Of course, sender and receivers must not be aware of the coordination they achieve, nor must they be consciously aware of the signal. Most non-human organisms are hard-wired to send and react on signals. Humans, on the other hand, are often aware of the signals they send and receive and understand the coordination they achieve. A nod of the head lets our interlocutor know we understood him and that she can continue the story. A longer than usual affectionate look in the eyes of a potential partner lets that person know we're romantically interested. And, of course, we constantly send linguistic signals to one another in order to coordinate our behavior. Whereas other species are limited in the signals they can send and interpret, we possess open-ended communication systems (human languages) enabling us to coordinate in an

equally open-ended way. This endless repertoire of signals enables us to coordinate in an inexhaustible number of ways for an equally inexhaustible number of purposes. That makes human coordination unique.

3.2 Public Knowledge

When it comes to coordinating with others, we also have another trick up our sleeves. Contrary to other species, humans do not have to communicate with each other to achieve coordination. As pointed out above, we coordinate successfully with other drivers on crossroads not by waving hands or flashing headlights (i.e. sending signals) but by heeding traffic lights or observing priority rules. In this case, we rely on public knowledge. Something is public knowledge among a group of people, when everybody does not only know X (that would merely be “shared knowledge”), but when everybody also knows that everybody else in the group knows X.

Public knowledge enables successful coordination in the absence of direct communication. Going back to our example above, we coordinate our driving behavior successfully at crossroads by having public knowledge of traffic light rules. We go at green not merely because we know we are allowed to. We go at green because we expect the light to be red for oncoming drivers and know (or at least expect with a high degree of confidence) that these drivers also know the traffic light rules and will stop at the red light. In this scenario, public knowledge of traffic light rules enables drivers to coordinate successfully without communicating with each other.

The rules of social contracts that solve coordination problems, do so by virtue of being public knowledge. Lewis (1969) refers to such rules as “conventions”: arbitrary rules that coordinate our behavior. They are arbitrary since participants could coordinate successfully in other ways (e.g. the rule to drive on the right side of the road is arbitrary since we could equally well all drive on the left side of the road). Such rules only need to be (publicly) known, they do not need to be enforced (at least not when we are dealing with pure coordination problems in which the interests of participants are fully aligned), since nobody would profit from breaking these rules (as long as they expect others to follow them). Even if the police would not punish driving on the “wrong” side of the road, (the vast majority of) people would not be tempted to do so, since they do not want to get into car accidents.

Lewis (1969) famously argued that the epistemic requirement for successful coordination on conventions is “common knowledge”. Common knowledge is “an infinite recursion of shared mental states, such that A knows X, A knows

that B knows X, A knows that B knows that A knows X, ad infinitum” (Thomas 2015, p. iii). However, experimental evidence suggests that such a demanding epistemic requirement (an infinite recursion of shared mental states) is not necessary for successful coordination. In an ingenious experiment, Devetag and colleagues (2013) actually found that second degree mutual knowledge (I know that you know and you know that I know) is sufficient for successful coordination. In a similar vein, Binmore (2008) has argued that common knowledge is not a necessary epistemic condition for successful coordination.

4 Solving Competition Problems

4.1 The Free-Rider Problem

Unlike coordination problems, competition problems are not merely epistemic problems. Getting information across or having common knowledge of a coordination solution does not suffice. The problem that needs to be dealt with is not only how we can coordinate our actions in order to cooperate but also: how we can prevent free-riding (“defecting”) from eroding cooperation. Given the payoff structure of competition problems, receivers have an incentive not to coordinate on the information received and senders have an incentive to send deceptive information.

Regarding the former, consider the following scenario. There is a water shortage in the village and the council determines that all households should not exceed a daily ration of 100 l per day, otherwise the reserves will dwindle. I need 150 l per day for normal consumption and irrigating my crops. We have a competition problem on our hand: a tragedy of the commons. I will benefit by exceeding my daily allowance in order to save my crop. Merely having the information (the ration) and being able to do my part in a coordinated effort for all of our long-term benefit (preventing a total drought), does not guarantee that I will do so.

Regarding the latter (sending deceptive signals), take an “honesty bar”. An unattended fridge is filled with beverages, a price list is tacked on the fridge and there’s a cash box in which customers are supposed to deposit the money for their consumption. The honesty bar is a cheap and efficient way to provide drinks (there’s no bartender on the payroll). The system benefits all customers since the price for the drinks is lower than it would be if there were a proper bar (the cost of which would reflect on the prices of the beverages). The honesty bar is located in a club and is only accessible to club members, all of whom promised to be honest. Nevertheless, members still have an incentive to cheat: retrospectively having sent a deceptive signal to abide by the rules.

4.2 Costly Signaling

One often invoked way to ensure honest signaling is “costly signaling” (Zahavi 1975, 1977; Nur and Hanson 1984; Grafen 1990; Roberts 1998; Smith and Bliege Bird 2000; Leimar and Hammerstein 2001; Gintis et al. 2001). The male peacock’s bright tail is a costly signal – it requires a lot of resources to keep it brightly colored – and therefore honest signal that the bird is fit and healthy. Only fit and healthy birds could forage enough to keep the tail brightly colored. Similarly, buying a love interest expensive gifts is a costly (and therefore honest) signal that one is wealthy and a good provider (it is honest since a poor person could not afford it and a bad provider would not want to).

In contexts where individuals could benefit from sending deceptive signals (i.e. in the context of competition problems), imposing costly signals is a way to prevent free-riding. Take religious communities. Typically, such communities offer support to their members in need by imposing altruistic obligations on their members (such as giving alms to the poor, taking care of the sick, etc.). They are therefore inherently vulnerable to free-riding (joining the community to reap the benefits of this religious altruism but not contributing anything oneself) (Iannacone 1992). That is why, according to Iannacone (1992) and other scholars of religion (Irons 2001; Sosis 2003, 2006; Bulbulia and Sosis 2011), costly signals or “credibility enhancing displays” (Henrich 2009) of commitment to the religion (and the prosocial norms it imposes) are so rampant in many religious communities. Such costly signals range from fasting and pilgrimages to fire walking, self-flagellation and even reenacted crucifixion. Their function is to keep free-riders out. Professing one’s belief in a god is easily done, but walking on hot coals to prove one’s commitment not so much.

However, imposing such costly signals may work in religious communities where participants are highly motivated to show their commitment, but it does not always work for solving everyday competition problems. In order to be effective in keeping free-riders out, signals must be costly enough to completely offset the benefits of free-riding. Otherwise free-riders can still benefit by sending the costly signal in a dishonest way. The problem is that imposing signals that are costly enough to prevent free-riding would also keep out potential cooperators since the payoff they get from mutual cooperation would equally be offset by the costly signal(s) they are required to send. Take the honesty bar example. We could install a locker on the fridge with a code that is only known to members having sent a signal costly enough to offset any benefit they could reap by free-riding. For example, they could be asked to donate a sum to the club that is substantial enough so that they could not make up for it by helping themselves to free drinks. However, that would defeat the whole

purpose of having cheap access to drinks. Nobody would want to pay the large sum to get access to the code.

What would work is a signal that is honest not in virtue of being costly but because it is hard – or even better, impossible – to fake. If I lift a barbell weighing 200 kg, I honestly signal that I am strong. There is no faking it, a weak person could not lift it. Similarly, the springbok “stotting” or “pronking” – performing acrobatic leaps in the air – when stalked by a predator to signal that it is fit (and the predator should not bother chasing it), sends a “hard-to-fake” signal. A sick or old animal could not perform the leaps. The problem is that such hard-to-fake signals are not always available. A good deceiver may often fool us by deceptively sending us all the “right” signals that she is honest.

4.3 Punishment

Since signaling often comes short in solving competition problems, we need another mechanism. That mechanism is punishment. More particularly “prosocial” punishment in which individuals are punished for harming group interests, rather than being retaliated against by individuals who were harmed (Fehr and Fischbacher 2004). Prosocial punishment occurs in all human societies. In his famous list of human universals, Brown (1991) includes “sanctions for crimes against the collectivity”. While such sanctions have always existed in human societies in some form or other, they became much more prevalent since the Neolithic transition from homogenous, small-scale societies to larger and more complex societies. Whereas in the former cooperation is largely maintained by kinship ties – explained by the mechanism of inclusive fitness (Hamilton 1964) – and personal exchanges – explained by the mechanism of reciprocal altruism (Trivers 1971) – in the latter these mechanisms are powerless to maintain high levels of cooperation since many interactions now take place between strangers (Powers and Lehmann 2014; Powers et al. 2016).

Prosocial punishment ranges from subtle signs of disapproval displayed by group members, over being ridiculed or scolded to being fined, incarcerated, banished or killed as a last resort. It is a very effective way to protect group interests in those domains where free-rider problems surface. Both experimental studies (Ostrom et al. 1994; Fehr and Gächter 2002; Falk et al. 2005; Gürerk et al. 2006; Fudenberg and Patac 2009; Chaudhury 2011) and field data (Ostrom 1990; Boehm et al. 1993; Henrich et al. 2006) show the effectiveness and the central role of punishment in solving competition problems. Going back to my honesty bar example, we could just install a camera and fine people (a sum exceeding the price of the stolen consumption) who fail to pay for their drink. No one would

benefit from cheating any longer (provided all cheaters are detected and fined). The competition problem would be solved: individual interests would no longer compete with group interests.

Given the central role of norms backed up by prosocial punishment in solving competition problems across human societies, it is surprising that they do not feature more prominently in Young's, Sugden's, Binmore's and Skyrms' naturalistic accounts of social structure. Two reasons, I believe, may explain this hiatus. The first is that the game-theoretic approach they adopt may lead them to implicitly assume that players in the various games of life are stuck with the payoff matrixes they are initially given (meaning that they cannot be changed by the players involved in these games). Secondly, as these authors are aware of, introducing prosocial punishment may not solve the free-rider problem after all because the act of punishing is costly to the punisher. So who is going to prevent the punisher from free-riding his obligation to punish? Any account invoking prosocial punishment to solve free-rider problems, must address this important challenge. In response, I will firstly argue that payoff matrixes of the games of life group members play with each other can and have indeed been changed by the participants. Then I will deal with the free-rider problem inherent to prosocial punishment.

4.4 Game Changers

Norms backed up by punishment solve competition problems by aligning individual interests with the interest of the group. They do so by imposing a cost on free-riding. That cost consists of (the perception of the actors of) the negative payoff yielded by the punishment for committing the infraction, given (the perception of the actors of) the chance of actually being punished for the infraction. The cost hinged on the free-riding strategy serves to no longer incentivize individuals to adopt a free-riding strategy. When the expected payoff of free-riding is lower than the expected payoff of cooperating, the competition problem is solved. Individual interests no longer compete with group interests.

Of course the perception of the cost of free-riding and of the gains one could get out of it, varies from individual to individual. The same punishment can deter one person but not another. So implementing punishment will rarely prevent everybody in the group from breaking the norm. However, in order to be effective, it should deter the vast majority of doing so. Throughout human history and across cultures, norms backed up by punishment have emerged to do just that. Take commons management. As Ostrom (1990) points out, in order to preserve common-pool resources from depletion, the use of such resources must be

regulated and monitored *and* infractions must be punished. Such norm enforcement can be found in all human societies [see Vlerick (2016) for an extensive account].

Note that the (endlessly) repeated prisoner's dilemma is not a competition problem. Sugden (2005), Skyrms (2004) and Binmore (2005) rightly point out that cooperation can emerge and be maintained in such an iterated game (without changing the payoff matrix), because cooperators can retaliate against free-riders by refusing to cooperate with them in future rounds. In such an iterated game, it is in the interest of participants to cooperate as long as others do so. The iteration, as Skyrms (2004, p. 4) points out, turns the game into a stag hunt. Like the stag hunt, the game has two Nash equilibria: the cooperative and the uncooperative equilibrium. In order to get to mutual cooperation, cooperators only need to identify and coordinate with other cooperators. Once cooperators interact with cooperators, there is no more incentive for free-riding. The immediate benefit of free-riding would be quickly offset by the retaliation that would follow in the future. It is therefore very much a coordination problem, not a competition problem.

Many real life prisoner's dilemma and common-pool resource problems, however, are not endlessly repeated. This is especially true in larger, more complex societies, in which many interactions between individuals are one-offs and anonymous. In those cases there is no "shadow of the future" (Trivers 1971; Axelrod 1984; Binmore 2005) looming over the participants and sanctions must be imposed in order to align individual interests with group interests. Unsurprisingly, in this light, while prosocial punishments occur in all human societies, the extent to which they occur correlates with their size and complexity (Marlow and Berbesque 2008). However, invoking pro-social punishment to solve free-rider problems may not make the free-rider problem disappear. Punishing, it is often pointed out, is costly for the punisher. So what is to prevent punishers from free-riding their obligation to punish?

4.5 The Cost of Punishment

Prosocial punishment is often referred to as "altruistic" punishment. It is deemed altruistic because the punisher incurs a personal cost, while the benefits are not reaped by the punisher in particular but by the whole community. The cost of punishment includes the time spent and the resources devoted to monitoring the behavior of others and punishing infractors as well as the personal risk one incurs when punishing someone who might lash back or take revenge later. Given that punishing is itself costly, the free-rider problem does not disappear by introducing prosocial punishment. It resurfaces on another level. The question now

becomes: how to enforce people to enforce the rules that prevent free-riding? Or as Binmore (2005, p. 85) puts it: “who guards the guardians”? Other guardians who would need to be guarded? We seem caught in an infinite regress. Prosocial punishment, nevertheless, is a prominent feature in societies (and doesn’t require second – let alone third or fourth order – norm enforcement). How can we account for the emergence and maintenance of this core feature of social contracts?

The short answer is: prosocial punishment is typically not as altruistic as it may seem. According to Boehm (1997) and Bowles and Gintis (2011, p. 5), in small-scale societies throughout human history – which were typically egalitarian – the possession of projectile weapons dramatically reduced the cost of punishing norm violators. Such projectile weapons enabled groups of people to collectively punish norm violators (e.g. by banishing or murdering them) at relatively low risk to each individual punisher. This, in effect, turns the cooperation problem of altruistic punishment from a competition problem into more of a coordination problem, i.e. the coordination of the act of collective punishment. This is also what Guala (2012, p. 9) points out. Ethnographic data suggests that prosocial punishment in small-scale societies typically is not (so) costly to the individual punishers, precisely because it takes the form of a coordinated punishment by a coalition against an individual. Moreover, as Boyd and colleagues (2003) point out, once punishing norm violators is done effectively in a society, norm violation becomes less frequent (people are deterred) and the cost of prosocial punishment decreases further (since there is less punishment to be handed out).

In large-scale societies on the other hand – according to Singh, Wrangham and Glowacki (2017) – the cost of enforcing group beneficial norms and the benefits reaped by enforcing these norms often vary dramatically from individual to individual. The reason for this is that these societies typically morphed into highly hierarchical societies (Powers and Lehmann 2014). Powerful individuals within such societies have much to gain from norms yielding high levels of cooperation since they find themselves at the receiving end of these cooperative endeavors *and* they have the means to enforce these norms with relative ease given the power they wield. Prosocial punishment is in those cases typically not altruistic but self-serving, since the enforcers often have more to gain than they stand to lose by enforcing such norms.

Moreover, in large-scale societies prosocial punishment is typically heavily institutionalized and does not require the actual punishers to be altruistic. They are professional punishers who are incentivized to carry out their duty in order to keep their livelihood and are provided with the means (arms) to do so efficiently and at relatively low risk. Professional policing is by no means a modern phenomenon. Such law enforcers existed in the ancient Egyptian, Greek and Roman societies.

Together with the burden on cognition to detect free-riders, the cost of punishment may explain why solutions to competition problems have rarely evolved in non-human social species (as opposed to signaling systems solving coordination problems) (Raihani et al. 2012). Humans, in contrast, have stumbled on ways to (radically) reduce the cost of prosocial punishment and evolved the cognitive abilities to detect free-riding.⁴ This makes us very apt at solving competition problems. So much for *how* human groups (could) solve cooperation problems. The question remains *why* these solutions (social contracts) emerged. That is the subject of the next section.

5 The Emergence and Evolution of Social Contracts

Which cultural dynamics underlie the emergence, persistence and evolution of social contracts? In this section, I will argue that social contracts are the outcome of two cultural dynamics: “within-group” and “between-group” dynamics [see also Vlerick (2020) where I apply these cultural dynamics to the cultural evolution of institutional religions]. Within-group dynamics refer to the interaction between individuals in the group, between-group dynamics to the interaction between groups.

5.1 Within-Group Dynamics

The focus of current game-theoretic accounts of social contracts is on these kinds of dynamics (Binmore 1994, 1998, 2005, 2007; Skyrms 1996, 2004; Young 1998; Sugden’s 2005). Within-group dynamics refer to the dynamic interplay of individual strategies of group members. They drive the outcome of the games of life we play with each other to stable points: Nash equilibria (in which everybody has a best response to whatever everybody else is doing). If the strategies of all participants do not hold each other in equilibrium (if everybody does not have a best response to what everybody else is doing), some participants can be expected to change their strategy and the outcome of the social interaction will be upset until it reaches an (Nash) equilibrium.

⁴ According to Cosmides and Tooby (1992) we are endowed with a specialized cognitive module to do so: a “cheater detection module”, and language enables us to exchange information regarding the trustworthiness of others in the group (Dunbar 1998).

Social contracts – sets of commonly known rules that govern the interaction of group members – occupy such equilibria. They don't need any glue, as Binmore (2005, p. 4) points out. They do not require anyone to sacrifice his or her self-interest for the benefit of the group.⁵ If this were not the case, people would (often) not adhere to the rules making up social contracts and the social contract would fall into desuetude. This is the central and important point of game-theoretic accounts of the emergence and evolution of social contracts.

Not all equilibria are equally likely to be occupied by social contracts, however. On the one hand, between-group dynamics (as I will explain in the next section) select efficient equilibria over less efficient equilibria. On the other hand, within-group dynamics often select psychologically salient equilibria. This is an important point made by Schelling (1960) and Lewis (1969). Psychologically more salient coordination solutions (equilibria in coordination games) are more likely to emerge than less salient solutions.

Take the driving game. The two salient coordination solutions are to all drive on the left or to all drive on the right. There are however an infinite number of other (and equally effective) coordination solutions. We could decide to drive on the left on uneven days of the month and on the right on even days, or left on week days and right in weekends, or left between 8 am and 8 pm and right between 8 pm and 8 am, etc. All these rules are proper equilibria⁶ – nobody has an incentive to deviate from these rules if everybody else follows them. However, they are unnecessarily complicated. It should not surprise us that all countries coordinate on one of the two most salient solutions of the driving game.

Within-group dynamics explain why (salient) coordination rules emerge. When it comes to solving competition problems, however, between-group dynamics play a major role. They select game changing norms (norms that affect the payoff related to the available strategies through punishment or reward to solve free-rider problems) which create better equilibria than the ones originally available.

5.2 Between-Group Dynamics

Social contracts are not only shaped by the interaction of individuals within the group (each coming up with a best response to what the others are doing), they

⁵ Note that when social contracts solve competition problems by imposing norms with sanctions, they do not require self-sacrifice. They change the payoff matrix so as to get self-interested strategies in line with group interests.

⁶ In fact they are what Aumann (1974) calls “correlated equilibria” in which all participants determine their strategy by following a public signal (such as date or time of the day in the examples above).

are also shaped by the interaction between groups. Whereas the former move the game to (salient) equilibria, the latter select efficient equilibria.⁷ As Binmore (2005, p. 5) puts it: “a social contract must be internally stable, or it would not survive. It needs to be efficient or it would not compete with the social contracts of other societies.” In other words, if a social contract is not stable, it will be dismantled by within-group dynamics. If it is not efficient (i.e. yielding a large payoff for group members), it will be dismantled by between-group dynamics.

Competition between groups selects for norms (and punishments) that enable cooperation in the context of competition problems (Aviles 2002; Boyd et al. 2003; West et al. 2007; Puurtinen and Mappes 2009).⁸ The selective pressure arising from group competition is what Boyd, Richerson, Henrich and others refer to as “cultural group selection” (Boyd and Richerson 1985; Richerson and Boyd 1999; Boyd and Richerson 2002; Henrich 2004; Richerson et al. 2016). Cultural entities (such as technology, knowledge and – in the context of this paper – social contracts) that provide the group with an advantage over other groups are likely to spread (i.e. they are likely to be culturally selected).

Several important factors underlie the cultural selection of group beneficial social contracts. Conflict between groups, competition between groups over scarce resources, demographic expansion of successful groups, migration to successful groups and imitation of successful groups all result in the spread of group beneficial norms and customs (Bowles and Gintis 2011, p. 50). How does that work? Very briefly (and grossly oversimplifying), in a direct conflict between groups, groups that cooperate well often have a (military) edge over groups that do not cooperate as well. The former groups end up winning conflicts against the latter, who vanish or are absorbed in the more cooperative group (and brought under their social contract). The same logic applies to

⁷ Although, as Singh et al. (2017) rightly argue, within-group dynamics can also lead to group beneficial equilibria. This is the case when the interests of powerful individuals (who are able to enforce the individually costly norms that lead to such equilibria) are aligned with the interests of the group. In that case, we do not need to invoke between group dynamics to explain the emergence of group beneficial norms, merely the strategic actions of individuals aimed at maximizing their personal payoff.

⁸ This does not mean that mutual cooperation in competition problems is produced or sustained by competition between groups. As pointed out in the previous section, it is typically sustained by prosocial punishment. It means that in a context of group competition, groups with high levels of cooperation will often outcompete groups with lower levels of cooperation and consequently the social contracts enabling these high levels of cooperation will thrive (more on this below).

competition for scarce resources between groups: the more cooperative groups have an advantage over less cooperative groups who find themselves at the losing end of the competitive interaction and ultimately vanish (while the cooperative groups thrive and expand). Furthermore, since efficient social contracts yield a higher (average) payoff to the individuals within the group, such groups have expanded historically. They did so for two reasons: firstly, more resources translated in greater reproductive output – more mouths could be fed (today this trend is reversed – that is known as the “economic demographic paradox”). Secondly, members of less successful groups migrated to more successful (read wealthier) groups (this trend holds true today – more than ever in fact because of increased mobility – see Collier 2013). Finally, efficient social contracts do not only spread because groups possessing them outcompete groups not possessing them. They also spread because they are being copied by other groups. Less successful groups imitate more successful groups, taking over some of their customs, innovations and social contracts (Boyd and Richerson 1985, 2002; Henrich 2004).⁹

5.3 Gene-Culture Co-Evolution

Interestingly, between-group dynamics have not only shaped human cultures, but indirectly also human nature. They had an (important) impact on human genetic evolution. In particular, they have shaped our genetically wired social psychology. That is known as “gene-culture co-evolution” (Boyd and Richerson 1985; Henrich 2004; Richerson and Boyd 2005; Gintis 2011).

In short (and again grossly oversimplifying), social contracts punishing free-riding and other asocial behavior create an environment in which individuals who are genetically predisposed to such behavior have a reproductive disadvantage (read they are often banished or killed and prevented from spreading

⁹ The cultural group selection hypothesis advanced by Boyd, Richerson and Henrich is influential but not without its critics. Some see it as a repackaged version of old and debunked (biological) group selection arguments (e.g. Pinker 2012). Others question the assumption that group conflict was pervasive throughout human history. As Sterelny (2016) (who nevertheless concludes that cultural group selection may well have been a major factor in the transition to large-scale, hierarchical societies in the Holocene) points out: ethnographic evidence by itself does not establish that conflict between groups was as prevalent as is often assumed. However, as explained above, direct conflict between groups is but one of the mechanisms producing cultural group selection.

their sociopathic genes) compared to individuals endowed with a more prosocial nature. This, as Henrich (2010) points out, led to “the self-domestication” of our species: we became increasingly more altruistic (towards group members) and prone to follow social norms. In turn, this reflected on the social contracts in the groups of our ancestors. Those contracts imposed increasingly more altruistic behavior on its member (and severe punishments on defectors), which again sharpened the genetic selective pressure on altruistic, prosocial, norm-abiding individuals, etc.

One important consequence of this is that human beings will often cooperate in competition situations (and so forsake the higher payoff they would get out of free-riding) in the total absence of sanctions (social, economic or other). As Bicchieri (2005, p. x) points out, humans often act prosocially even if they are not coerced to do so. (That is why honesty bars without cameras work). This is clearly shown by behavioral game-theoretic experiments (dictator games) in which people regularly divide a sum fairly even if they could have kept everything for themselves without any negative consequence (Camerer and Thaler 1995). The acute sense of fairness that prompts people to act altruistically seems to an important extent innate. Young children, who could not have acquired this sense of fairness culturally, already exhibit it (Warneken and Tomasello 2009). It evolved in response to social environments characterized by increasingly severe prosocial norms and makes *Homo sapiens*, as Bowles and Gintis (2011) have put it, a remarkably “cooperative species”.

6 Conclusion

Social contracts emerged, persisted and evolved in response to the various cooperation problems humans have faced throughout history. Cooperation problems comprise both coordination and competition problems. The former are solved by creating common knowledge of coordination rules within the group. These solutions emerge from within-group dynamics. The latter are solved by norms that punish free-riding and doing so align individual interests with the group interest. They are largely (but not exclusively) culturally selected as a result of between-group competition. Current naturalistic accounts of the emergence and evolution of social contracts have focused mainly on how within-group dynamics lead to solutions to coordination problems. A more complete picture emerges when we view social contracts as solutions to cooperation problems – not merely coordination problems – and include between-group dynamics that select solutions to competition problems.

Bibliography

- Aumann, R. (1974): “Subjectivity and Correlation in Randomized Strategies”. In: *Journal of Mathematical Economics* 1. No. 1, p. 67–96.
- Aviles, L. (2002): “Solving the Freeloaders Paradox: Genetic Associations and Frequency-Dependent Selection in the Evolution of Cooperation among Nonrelatives. In: *Pnas* 99. No. 22, p. 14268–14273.
- Axelrod, R. (1984): *The Evolution of Cooperation*. New York: Basic Books.
- Bicchieri, C. (2005): *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.
- Binmore, K. (1994): *Playing Fair: Game Theory and the Social Contract I*. Cambridge, MA: The MIT Press.
- Binmore, K. (1998): *Just Playing: Game Theory and the Social Contract II*. Cambridge, MA: The MIT Press.
- Binmore, K. (2005): *Natural Justice*. Oxford: Oxford University Press.
- Binmore, K. (2007): *Game Theory: A Very Short Introduction*. Oxford: Oxford University Press.
- Binmore, K. (2008): “Do Conventions Need to be Common Knowledge?” In: *Topoi* 27. No. (1–2), p. 17–27.
- Boehm, C. (1997): “Impact of the Human Egalitarian Syndrome on Darwinian Selection”. In: *American Naturalist* 150, p. 100–121.
- Boehm, C., H. Barclay, R. K. Dentan, M. Dupre, J. Hill, S. Kent, B. Knauft, K. Otterbein and S. Rayner (1993): “Egalitarian Behavior and Reverse Dominance Hierarchy”. In: *Current Anthropology* 34. No. (3), p. 227–254.
- Bowles, S. and H. Gintis (2011): *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.
- Boyd, R. and P. Richerson (1985): *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Boyd, R. and P. Richerson (2002): “Group Beneficial Norms can Spread Rapidly in a Structured Population”. In: *Journal of Theoretical Biology* 215, p. 287–296.
- Boyd, R., H. Gintis, S. Bowles and P. Richerson (2003): “The Evolution of Altruistic Punishment”. In: *Pnas* 100. No. 6, p. 3531–3535.
- Brown, D. (1991): *Human Universals*. New York: McGraw-Hill.
- Bulbulia, J. and R. Sosis (2011): “Signaling Theory and the Evolution of Religious Cooperation”. In: *Religion* 41. No. 3, p. 363–388.
- Camerer, C. and R. Thaler (1995): “Anomalies: Ultimatums, Dictators and Manners”. In: *Journal of Economic Perspectives* 9. No. 2, p. 209–219.
- Chaudhuri, A. (2011): “Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature”. In: *Experimental Economics* 14, p. 47–83.
- Collier, P. (2013): *Exodus: How Migration is Changing our World*. Oxford: Oxford University Press.
- Cosmides, L. and J. Tooby (1992): “Cognitive Adaptations for Social Exchange”. In: J. Barkow, L. Cosmides and J. Tooby (Eds.): *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press, p. 163–228.
- Devetag, G., H. Hosni and G. Sillari (2013): “You Better Play 7: Mutual Versus Common Knowledge of Advice in a Weak-Link Experiment”. In: *Synthese* 190. No. 8, p. 1351–1381.
- Dunbar, R. (1998): “The Social Brain Hypothesis”. In: *Evolutionary Anthropology: Issues, News, and Reviews* 6. No. 5, p. 178–190.

- Eriksson, K. and P. Strimling (2012): “The Hard Problem of Cooperation”. In: *PLoS One* 7. No. 7, e40325.
- Falk, A., E. Fehr and U. Fischbacher (2005): Driving Forces Behind Informal Sanctions. IZA Discussion Paper No. 1635. Available at SSRN: <https://ssrn.com/abstract=756366>.
- Fehr, E. and S. Gächter (2002): “Altruistic Punishment in Humans”. In: *Nature* 415, p. 137–140.
- Fehr, E. and U. Fischbacher (2004): “Third-Party Punishment and Social Norms”. In: *Evolution and Human Behavior* 25. No. 2, p. 63–87.
- Fudenberg, D. and P. Pathak (2009): “Unobserved Punishment Supports Cooperation”. In: *Journal of Public Economics* 94, p. 78–86.
- Gintis, H. (2009): *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton: Princeton University Press.
- Gintis, H. (2011): “Gene-Culture Coevolution and the Nature of Human Sociality”. In: *Philosophical Transaction of the Royal Society* 366, p. 878–888.
- Gintis, H., E. Alden Smith and S. Bowles (2001): “Costly Signaling and Cooperation”. In: *Journal of Theoretical Biology* 213, p. 103–119.
- Grafen, A. (1990): “Biological Signals as Handicaps”. In: *Journal of Theoretical Biology* 144. No. 4, p. 517–546.
- Guala, F. (2012): “Reciprocity: Weak or Strong? What Punishment Experiments do (and do not) Demonstrate”. In: *Behavioral and Brain Sciences* 35, p. 1–59.
- Gürker, O., B. Irlenbusch and B. Rockenbach (2006): “The Competitive Advantage of Sanctioning Institutions”. In: *Science* 312, p. 108–111.
- Hamilton, W. (1964): “The Genetical Evolution of Social Behaviour I and II”. In: *Journal of Theoretical Biology* 7, p. 1–16 and 17–52.
- Hardin, G. (1968): “The Tragedy of the Commons”. In: *Science* 162, p. 1243–1248.
- Henrich, J. (2004): “Cultural Group Selection, Coevolutionary Processes and Large-Scale Cooperation”. In: *Journal of Economic Behavior & Organization* 53. No. 1, p. 3–35.
- Henrich, J. (2009): “The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and their Implications for Cultural Evolution”. In: *Evolution and Human Behavior* 30. No. 4, p. 244–260.
- Henrich, J. (2010): *The Secret of our Success: How Culture is Driving Human Evolution, Domesticating our Species and Making us Smarter*. Princeton: Princeton University Press.
- Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer and J. Ziker (2006): “Costly Punishment across Human Societies”. In: *Science* 312, p. 1767–1770.
- Iannaccone, L. (1992): “Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and other Collectives”. *Journal of Political Economy* 100, p. 271–291.
- Irons, W. (2001): “Religion as a Hard-to-Fake Sign of Commitment”. In: R. Nesse (Ed.): *Evolution and the Capacity for Commitment*. New York: Russell Sage Foundation, p. 290–309.
- Leimar, O. and P. Hammerstein (2001): “Evolution of Cooperation through Indirect Reciprocity”. *Proceedings of the Royal Society of London B* 268, p. 745–753.
- Lewis, D. (1969): *Conventions: A philosophical Study*. Cambridge, MA: Harvard University Press.
- Marlow, F. and J. C. Berbesque (2008): “More ‘Altruistic’ Punishment in Larger Societies”. *Proceedings of the Royal Society of London B* 275, p. 587–590.
- Maynard Smith, J. and G. Price (1973): “The Logic of Animal Conflict”. In: *Nature* 246, p. 15–18.
- Nur, N. and O. Hasson (1984): “Phenotypic Plasticity and the Handicap Principle”. In: *Journal of Theoretical Biology* 110. No. 2, p. 275–297.

- Ostrom, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press. Retrieved from: https://www.edge.org/conversation/steven_pinker-the-false-allure-of-group-selection.
- Ostrom, E., R. Gardner and J. Walker (eds.) (1994): *Rules, Games and Common-Pool Resources*. Michigan: The University of Michigan Press.
- Pinker, S. (2012): “The False Allure of Group Selection”. In: *Edge* Jun 19, 2012.
- Powers, S. and L. Lehmann (2014): “An Evolutionary Model Explaining the Neolithic Transition from Egalitarianism to Leadership and Despotism”. In: *Proceedings of the Royal Society B* 281. No. 1791, p. 20141349.
- Powers, S., C. van Schaik and L. Lehmann (2016): “How Institutions Shaped the Last Major Evolutionary Transition to Large-Scale Human Societies”. In: *Philosophical Transactions of the Royal Society B* 371. No. 1687, p. 20150098.
- Puurtinen, N. and T. Mappes (2009): “Between-Group Competition and Human Cooperation”. In: *Proceedings of the Royal Society B* 276, p. 355–360.
- Raihani, N., A. Thornton and B. Redouan (2012): “Punishment and Cooperation in Nature”. In: *Trends in Ecology and Evolution* 27. No. 5, p. 288–295.
- Richerson, P. and R. Boyd (1999): “Complex Societies: The Evolutionary Origins of a Crude Superorganism”. In: *Human Nature* 10. No. 3, p. 253–289.
- Richerson, P. and R. Boyd (2005): *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Richerson, P., R. Baldini, A. V. Bell, K. Demps, K. Frost, V. Hillis, S. Mathew, E. K. Newton, N. Naar, L. Newson, C. Ross, P. E. Smaldino, T. M. Waring and M. Zefferman (2016): “Cultural Group Selection Plays an Essential Role in Explaining Human Cooperation: A Sketch of the Evidence”. In: *Behavioral and Brain Sciences* 39, p. e30.
- Roberts, G. (1998): “Competitive Altruism: from Reciprocity to the Handicap Principle”. In: *Proceedings of the Royal Society of London B* 265, p. 427–431.
- Schelling, T. (1960): *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Skyrms B. (1996): *Evolution of the Social Contract*. Cambridge, UK: Cambridge University Press.
- Skyrms, B. (2004): *The Stag Hunt and the Evolution of Social Structure*. Cambridge, UK: Cambridge University Press.
- Skyrms, B. (2010): *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press.
- Singh, M., R. Wrangham and L. Glowacki (2017): “Self-Interest and the Design of Rules”. In: *Human Nature* 28, p. 457–480.
- Smith, E. and R. Bliege Bird (2000): “Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signaling”. In: *Evolution of Human Behavior* 21, p. 245–261.
- Sosis, R. (2003): “Why aren’t we all Hutterites? Costly Signaling Theory and Religious Behavior”. *Human Nature* 14. No. 2, p. 91–127.
- Sosis, R. (2006): “Religious Behaviors, Badges, and Bans: Signaling Theory and the Evolution of Religion”. In: P. McNamara (ed.): *Where God and Science Meet*. Westport, CT: Praeger, p. 61–86.
- Sterelny, K. (2016): “Cooperation, Culture, and Conflict”. In: *British Journal for the Philosophy of Science* 67. No. 1, p. 31–58.
- Sugden, R. (2005): *The Economics of Rights, Co-Operation and Welfare*. London: Palgrave MacMillan.
- Thomas, K. (2015): *The Psychology of Common Knowledge: Coordination, Indirect Speech, and Self-Conscious Emotions* (Doctoral dissertation). Harvard: Harvard University, Graduate

- School of Arts & Sciences. (Retrieved from: <http://nrs.harvard.edu/urn3:HUL.InstRepos:17467482>).
- Trivers, R. (1971): “The Evolution of Reciprocal Altruism”. In: *Quarterly Review of Biology* 46, p. 35–57.
- Vlerick, M. (2016): “Explaining Universal Social Institutions: A Game-Theoretic Approach”. In: *Topoi* 35. No. 1, p. 291–300.
- Vlerick, M. (2020): “The Cultural Evolution of Institutional Religions”. In: *Religion, Brain & Behavior* 10. No. 1, p. 18–34.
- Warneken, F. and M. Tomasello (2009): “Varieties of Altruism in Children and Chimpanzees”. In: *Trends in Cognitive Sciences* 13. No. 9, p. 397–402.
- West, S., A. Griffin and A. Gardner (2007): “Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection”. In: *Journal of Evolutionary Biology* 20. No. 2, p. 415–432.
- Young, P. (1998): *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.
- Zahavi, A. (1975): “Mate Selection – a Selection for a Handicap”. In: *Journal of Theoretical Biology* 53. No. 1, p. 205–214.
- Zahavi, A. (1977): “The Cost of Honesty (Further Remarks on the Handicap Principle)”. In: *Journal of Theoretical Biology* 67. No. 3, p. 603–605.

Article note: This paper is original work. It has not been published elsewhere, nor is it under consideration at another journal.