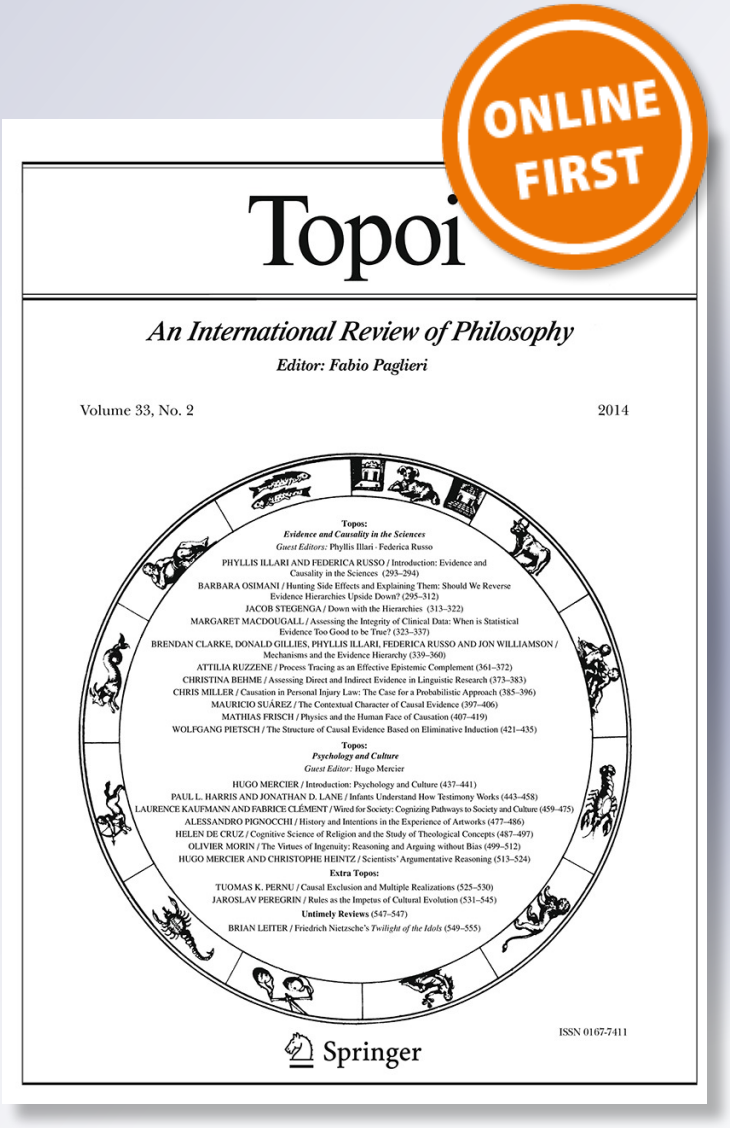
Explaining Universal Social Institutions: A Game-Theoretic Approach

Michael Vlerick

Topoi
An International Review of Philosophy

ISSN 0167-7411

Topoi
DOI 10.1007/s11245-014-9294-z



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

Topoi
DOI 10.1007/s11245-014-9294-z

Explaining Universal Social Institutions: A Game-Theoretic Approach

Michael Vlerick

© Springer Science+Business Media Dordrecht 2015

Abstract Universal social institutions, such as marriage, commons management and property, have emerged independently in radically different cultures. This requires explanation. As Boyer and Petersen (J Inst Econ 8:125, 2012) point out 'in a purely localist framework (these institutional commonalities) would have to constitute massively improbable coincidences' (3–4). According to Boyer and Petersen, those institutions emerged naturally out of genetically wired behavioural dispositions, such as marriage out of mating strategies and borders out of territorial behaviour. While I agree with Boyer and Petersen that 'unnatural' institutions cannot thrive, this one-sided explanation of universal social institutions in terms of genetic human nature is unsatisfactory. Drawing on the literature on multi-level selection and gene-culture coevolution, I argue that universal social institutions are first and foremost the products of cultural selection. They occupy fitness peaks in the landscape of cultural possibilities, much in the same way that biological adaptations occupy fitness peaks in the landscape of biological possibilities. To show this, I use game-theory. By modelling the domains of social interaction in which marriage, commons management, and property emerged as Prisoner's dilemma situations, it becomes clear how an institutional framework allows the group to move to an interactive equilibrium with a larger payoff. Institutions do so by incentivising (through punishment and/or reward) all parties to adopt a cooperative strategy. They are culturally selected ways of optimising genetically constrained domains of human social interaction.

Keywords Universal social institutions · Multi-level selection · Gene-culture coevolution · Game-theory · Prisoner's dilemma · Correlated equilibrium

1 Introduction

Institutions are a fundamental feature of human societies. As Boyd and Richerson (2008) point out, complex social institutions play a crucial role in even the simplest human societies (306). Not only do all human societies possess complex institutions, a number of institutions are shared by (virtually) all human societies. In radically different cultures similar institutions have emerged independently. This requires explanation. Indeed, Boyer and Petersen (2012) remark correctly that 'in a purely localist framework (these institutional commonalities) would have to constitute massively improbable coincidences' (3–4). Examples of such universal social institutions are marriage, commons management, 'criminal law',[1] and property.

In recent work, Boyer and Petersen (2012) explain the recurrence of social institutional features across different cultures by grounding them in our evolved psyche. According to Boyer and Petersen, 'institutions are best understood against the background of a set of human psychological dispositions that influence the effort needed to adopt and accept certain social arrangements' (4). In

M. Vlerick (✉)
Department of Philosophy B603A, University of Johannesburg, Po Box 524, Auckland Park 2006, South Africa
e-mail: michaelvlerick@gmail.com; mvlerick@uj.ac.za

M. Vlerick
9A Chichester Road, Westdene 2092, South Africa

[1] In illiterate societies this takes the form of a series of explicit norms for conflict resolution and the punishment of wrongdoing (such as in-group murder, theft, etc.; Hoebel 1964).

Published online: 10 February 2015

Author's personal copy

other words, institutions emerge naturally out of behavioural dispositions. Marriage, for instance, emerges out of mating strategies, much as borders arise out of territorial behaviour and criminal law out of moral intuitions.

In a similar vein, Sperber (1996) grounds recurring cultural facts in our innate mental apparatus. In his 'epidemiology of representations', Sperber sets out to formulate a causal explanation of cultural facts. According to Sperber some representations are more susceptible to become stable, shared representations—and therefore cultural—because those representations are favoured by our mental wiring. The psychological properties with which natural selection endowed us function as a filter, retaining certain kinds of representations and discarding others. Cultural universals—ranging from ideas (e.g. in a divinity, about the natural world) to behavioural patterns (childcare, healthcare, warfare, etc.)—are therefore in an important sense the product of our genetically determined cognitive and affective nature. Universal institutions, being a subset of cultural universals, therefore, are the product of our mental wiring, according to Sperber.

How sound is this reasoning? While Sperber, Boyer and Petersen are right in relating the features of recurring social institutions to the dispositions of our evolved mental apparatus, pressing an important point against traditional sociological and philosophical accounts of social ontology (e.g. Gudeman 1986; Searle 1995, 2010)—which turn a blind eye to human nature as a constituent in the emergence of institutions—the 'naturalness' of institutions, as Boyer and Petersen frame it, only provides us with a partial explanation of why similar institutions have emerged independently in very different cultures. They ignore the cultural side of the story.

In this paper I argue that universal social institutions (more precisely the institutions of property, marriage, and the regulation of common resources) do not merely express genetic human nature. They are also the product of a selection process at the cultural level. In the same way that biological evolution gravitates towards particular fitness peaks in the landscape of biological possibilities, cultural evolution has zoned in on a number of fitness peaks in the landscape of cultural possibilities. Universal social institutions, I will argue, occupy such cultural fitness peaks. Given that the landscape of cultural possibilities is constrained by genetic human nature (more precisely by our cognitive and affective nature), Boyer and Petersen are right in claiming that universal social institutions are grounded in human nature. Nevertheless, genetic human nature is compatible with a broader range of social institutions than the universal social institutions we actually see. In other words, human nature underdetermines the selection of universal institutions. Pointing only at genetic human nature thus provides us with an incomplete and ultimately misleading explanation of the emergence of those universal institutions. It explains why particular institutions could never thrive given our genetic nature, not why particular institutions become universal. In this regard, my account rejects both a one-sided cultural account of social institutions (ignoring human nature) and a one-sided account in terms of genetically wired dispositions of our evolved psyche (ignoring cultural selective dynamics).

In order to show this and explain the cultural dynamic behind the emergence of universal social institutions, I will draw on two strands of research. The first is the literature on multi-level selection and gene-culture coevolution, and the second is game-theory. In Sect. 2, I present the non-genetic (i.e. cultural) forces involved in shaping institutions. In Sect. 3, I introduce the game-theoretic approach. In Sect. 4, I present three case studies, in which I bring the two strands together and offer what I hope will convince the reader is a satisfactory explanation of the emergence of the proposed universal social institutions. Finally, in Sect. 5, I offer an alternative naturalistic account of the independent emergence of universal social institutions (contra Boyer and Petersen).

2 More than Genetic Expression

2.1 Multi-level Selection and Gene-Culture Coevolution

Human beings cooperate on a massive level. Indeed, all human societies, ranging from hunter gatherer groups to modern western societies, are characterised by division of labour, trade, and joint projects. Nowhere else in the history of evolution has there been a species that even comes close to our level of cooperation. As Bowles and Gintis (2011: 2) point out, what makes human cooperation exceptional, is both its scale and the fact that it extends beyond genealogical kin to include even total strangers.

Underlying this cooperation, of course, are a series of remarkable and distinctive cognitive and linguistic capacities—such as joint attentional skills, the ability to communicate rules or agreements to others and the ability to internalise these rules—in conjunction with equally distinctive social emotions such as shame, guilt, and moral outrage (Bowles and Gintis 2011; Tomasello 2009). Nevertheless, while a proximate explanation of human cooperation is rather straightforward (i.e. humans have the means and the desire to engage in altruistic cooperation), the ultimate explanation is more puzzling. The question remains how human cooperation could have evolved.

Given that survival and reproduction is the only currency in the evolutionary context, the evolution of proper altruism among non-related individuals—i.e. incurring a

Author's personal copy

fitness cost to benefit another—is baffling at first sight.[2] Self-interested individuals will outcompete altruists in any given group (they will reap more benefits and pay less costs and therefore have higher chances on successful reproduction). How then can altruism be maintained? Any altruistic group is extremely vulnerable to invading free-riders who would steadily hollow out altruistic cooperation. And conversely, altruistic behavioural traits could never invade groups made up of self-interested individuals. How, in this light, could it have spread in the first place?

Bowles and Gintis (2011) (building on Sober and Wilson 1998 among others) point at what they call multi-level selection. Not only do individuals compete with each other, groups also compete. According to Bowles and Gintis, for a number of reasons—such as dependency on large game hunting in early human environments and the extended time it takes to raise children—cooperation between members of a group endowed the group with significant advantages over non-cooperating groups. Cooperating groups therefore thrived at the expense of other groups and altruistic cooperation was able to spread through cultural transmission.

> Between-group competition for resources and survival was and remains a decisive force in human evolutionary dynamics. Groups with many cooperative members tended to survive these challenges and to encroach upon the territory of the less cooperative groups, thereby both gaining reproductive advantages and proliferating cooperative behaviors through cultural transmission. The extraordinarily high stakes of intergroup competition and the contribution of altruistic cooperators to success in these contests meant that sacrifice on behalf of others, extending beyond the immediate family and even to virtual strangers, could proliferate. (Bowles and Gintis 2011: 4).

In other words, cooperating groups tended to thrive—read expand and split—at the cost of other groups that therefore vanished. Cooperation became more and more prevalent among the human population. One question remains however, how did these cooperating groups maintain a high level of cooperation? Remember that free-riders would have an evolutionary advantage and 'freeriding' could easily invade cooperative groups.

[2] 'Kin selection'—altruism towards one's genetic kin—and 'reciprocal altruism' also called 'enlightened self-interest' ('you scratch my back and I'll scratch yours'), on the other hand, can easily be explained from a gene-centric evolutionary perspective. In the first case, altruistic behaviour favours the spread of one's genetic material, since it increases the chances of survival of one's offspring. In the second case, both parties are better off—a typical instance of mutualism (as opposed to altruism where an individual pays a cost—i.e. loses fitness—to benefit another individual—i.e. increase his or her fitness).

The answer, I contend, is institutions; or more precisely, the universal social institutions which are the subject of this paper. In an influential account, Richerson and Boyd (2005) argue that the (genetic) evolution of our altruistic tendencies on the one hand, and the (cultural) evolution of social institutions on the other hand, are fundamentally intertwined. The process is known as gene-culture coevolution. More concretely, and grossly oversimplifying, the story is as follows. First, selection at the cultural level favoured the spread of social institutions because large cooperative groups outcompeted smaller less cooperative groups *and* cooperation in the larger groups was cemented by social institutions or proto-institutions. This form of cultural (not genetic) group selection gradually weeded out less cooperative groups. The result of this cultural process is a significant modification of the environment in which human (genetic) evolution takes place. Whereas in a non-institutional environment selfishness and nepotism pays off, in an institutionalised environment, freeriding and non-cooperative behaviour is severely punished. 'Sociopaths' were identified and banished or murdered, preventing them from spreading their sociopathic genes; and more altruistic individuals (and their genetic endowment) reaped the benefits of the new social context. The genetic predisposition in the hominid gene-pool for selfish behaviour was gradually tuned down and our predisposition for cooperation and altruism took off. Social institutions changed the payoff matrix and genetic selection followed, which again reinforced social institutions etc. (Richerson and Boyd 2005: 197). In other words, whereas within a non-institutional social environment, the evolution of cooperative altruism could never evolve (since those altruists pay fitness costs and therefore impair their chances on survival and reproduction), within an institutional social environment cooperative altruistic traits evolved (since the institutional framework imposed this cooperation on the group members and defectors were punished).

2.2 Fitness Peaks in the Landscape of Cultural Possibilities

If the general tenor of the account above is correct—and I believe it is—representing institutions as an expression of our genetically wired nature is misleading. It appears that institutions shaped our genetic predisposition for prosocial behaviour rather than the other way around. So, to a certain extent, the causal arrow seems to point in the other direction, or at least points in both directions (institutions influencing genetic evolution and genetic evolution influencing institutions). Those universal social institutions are the result of intergroup competition, rather than an extended form of our genotype (such as the beaver's dam or the bird's nest). Institutions are the product of cultural

Author's personal copy

selection, rather than natural selection. They result from a group selective process, representing the most fitness inducing ways of organising human interaction, selected over tens of thousands of years.

Not surprisingly, as Boyd and Richerson (2008) contend, social institutions often benefit social groups. While there are clear and sometimes striking counterexamples, institutions in most cases function to reduce costs of interactions, limit conflict and 'predatory behavior', while boosting productivity (309). This is exactly what we would expect from the multi-level selection hypothesis outlined above.

Boyer and Petersen's (2012) one-sided explanation of universal institutions in terms of our evolved psyche, therefore, is not entirely satisfactory. The universal social institutions they refer to are not merely elaborated forms of extended phenotypes. They represent fitness peaks in the landscape of cultural possibilities, much in the same way that biological adaptations such as our opposable thumbs or the giraffe's long neck represent a fitness peak in the landscape of biological possibilities (see Wright 1932 on the notion of fitness landscape). As pointed out before, given that the landscape of cultural possibilities is constrained by our biology, Boyer and Petersen are right in claiming that 'unnatural' institutions—institutional arrangements which run against our intuitive grain—won't stand a chance. The naturalness of institutions is a necessary condition but not a sufficient one for the emergence of universal social institutions. Institutions that are deleterious to groups in the long run will be selected out even if they're a close fit with genetically wired behavioural dispositions. Following Richerson and Boyd (2005), where humans are concerned, natural and cultural selective processes are intricately intertwined. Ignoring either human nature (as in the traditional sociological and philosophical accounts of institutions) or human culture (as Boyer and Petersen do) can only provide us with a partial explanation.

Particularly revealing, in this regard, is the game-theoretic modelling of those areas of interaction in which universal social institutions evolved. After introducing game-theory in the next section, this will be the subject of Sect. 4.

3 Game-Theory: Modelling Human Interaction

Game-theory, according to Gintis (2009), is a valuable and indispensable tool in the toolbox of the social scientist. It enables us to model social interaction with mathematical precision, and gives a sound and sturdy foundation to theories which would otherwise remain mere conjectures—or to dig up a classic derogatory label in evolutionary circles—'just so stories'. As Gintis (2009) points out, social interaction in a vast array of species can be analysed with game-theory. Humans are no exception. Nevertheless, human games are peculiar since their players can play these interactive games while actually knowing the rules. This—pardon the pun—is a game changer. It allows people, or rather groups of people, to change the rules and therefore, literally, change the game.[3] This is precisely—I will argue—what institutions do. But before we get to the meat of the argument, let me briefly introduce the relevant aspects of game-theory.

Game-theory offers mathematical models of the strategic interaction between two or more players, each having two or more strategies at their disposal. The outcome of the game is a certain payoff for each player resulting from the strategies adopted and the payoff structure of the game. As an example will do more than a lengthy description, let's take a look at the classic 'Prisoner's dilemma', arguably the most well-known game in game-theory.

Two members of a gang are arrested for a crime. 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 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

Assuming both players are rational and self-interested, the outcome of this game will gravitate towards a Nash equilibrium. A Nash equilibrium is reached when all players have a best response (i.e. a response maximising their payoff) given the strategy of other players. A Nash equilibrium involves either a pure strategy (adopting one strategy) or a mixed strategy, which is a probability distribution of two or more strategies (e.g. do × 1/3 of the time and do year 2/3 of the time). In the Prisoner's

[3] Gintis et al (in press) argue that evolutionary forces steered the hominin line towards a 'political niche', to which we adapted by developing the ability (through communication and persuasion) to construct and reconstruct the social order.

Author's personal copy

dilemma the Nash equilibrium is a pure strategy equilibrium: one in which both players betray each other. This is because it is better for each player to betray the other, regardless of what the other player does. If player 1 betrays player 2, it will be better for player 2 to betray player 1, and get 2 years in prison instead of 3. If player 1 cooperates with player 2 (by remaining silent), it will still be better for player 2 to betray player 1, and get 0 years in prison instead of 1. And vice versa for player 1. Assuming self-interested players, the best response for each player—given the strategies available and the payoff structure of the game—will always be to betray. This however doesn't lead to the most preferable outcome for *both* players, either in sum or on average. The best sum or average outcome would be [−1,−1], but if both employ their best individual strategy, the outcome will be[−2,−2]. Thus the best strategy for each individual to follow leads to a suboptimal outcome for a group whose members all follow it.

The Prisoner's dilemma is not chosen haphazardly. This structure of this game, as the attentive reader may have noticed, represents the problem of cooperation presented in Sect. 2 on how cooperative behaviour could spread, given that it will always benefit an individual to defect. The Prisoner's dilemma, in other words, models the evolutionary problem of cooperation. Given that an individual is always better off adopting a self-serving strategy instead of a cooperative one, how could cooperation get off the ground?

This is where institutions come in. The 3 following case studies of universal social institutions can each be modelled as a solution to get out of the undesirable Nash equilibrium in Prisoner's dilemma contexts. The role of institutions, as will become clear in the next section, is to push the equilibrium of the interaction to the more preferable outcome for all players involved. To [−1,−1] instead of [−2,−2], to borrow the payoff structure from our example above.

4 Three Case-Studies

4.1 Case Study 1: Property

On the face of it, two strategies can be adopted to acquire goods. The first way is the 'honest' way of producing the goods or trading something in exchange. The second way is to acquire them by sheer force. History is filled with conflicts in which groups of warriors raided agrarian societies, robbing them of their resources. Even today in some war-ridden regions, characterised by an absence of institutional (legal) enforcement (such as the Kivu region in Congo), we still witness armed groups raiding villages on a regular basis.

In a non-institutionalised context—where raiding is not sanctioned—we can therefore imagine a stable mixed equilibrium, where part of the total population produces and trades goods and another part uses force to acquire them.[4] In the context of evolutionary biology, such an equilibrium has been named an evolutionary stable strategy (Smith and Price 1973). Such a strategy, once it is fixed in a population, cannot be invaded by an alternative strategy (at least if the payoff matrix remains constant). Smith and Price illustrate this with the hawk-dove game with the following payoff matrix:

	Hawk	Dove
Hawk	X, X	W, L
Dove	L, W	T, T

When fighting over resources a hawk—who is willing to fight—will dominate a dove—who merely bluffs and retreats when the other party engages. When two doves face each other, they 'thigh' and both get away without incurring any harm. When two hawks face each other they will fight and both incur serious harm. Smith and Price's (1973) insight is that in this idealised context natural selection will produce a stable proportion of hawks and doves. More hawks would make it advantageous to be a dove (and walk away from likely harm given the large population of hawks) which therefore will spread in the population. Conversely, more doves would make it advantageous to be a hawk (and dominate the large population of doves in the struggle over resources).

In the context of human strategies to acquire goods, we can imagine such an equilibrium or ESS of—for instance—2/3 producers and 1/3 takers. More takers would mean too much competition for the taking: many harmful confrontations between takers and not enough resources to be taken. Less takers would make it advantageous to become a taker given a higher amount of total resources and less competition for the taking. Imagine that in the ESS the total production amounts to 20 units produced by 2/3 of the population. If however property becomes sanctioned and

[4] I am not claiming this would actually be the case were there no sanctioning institutions. As pointed out in our discussion of gene-culture coevolution, human beings developed strong prosocial emotions. Interestingly, in this context, Gintis (2007) models the endowment effect—the fact that people value an object they possess more than the same object if they do not possess it—as respect for private property in the absence of institutional enforcement. The non-institutional state I refer to, in this regard, is a kind of hypothetical Hobbesian state of nature predating institutions and the coevolving genetic evolution of prosocial dispositions.

Author's personal copy

theft punished, the payoff matrix is changed, making the 'dishonest' strategy more costly. We can move to a scenario where almost everybody becomes a producer and almost no one a taker. This means an increase to 30 units for the whole population. Clearly a more desirable equilibrium (on average) for all parties. Compare the following payoff matrixes:

Pre-institutional game[5]:

	Honest	Dishonest
Honest	1, 1	0, 2
Dishonest	2.0	−1, −1

Given this payoff matrix, we get a mixed Nash equilibrium (the ESS) where part of the population adopts a dishonest strategy and part of the population adopts an honest strategy (or the individuals adopt a dishonest strategy part of the time and an honest strategy for the other part of the time).

Institutional game:

	Honest	Dishonest
Honest	1, 1	0, −1 (sanctioned)
Dishonest	−1.0 (sanctioned)	−2, −2 (sanctioned)

Now, it should be obvious, it will always be advantageous to adopt the honest strategy. Therefore, we get a pure Nash equilibrium where everybody adopts the honest strategy.

In this regard, the institution of property steers the outcome of the interactive game towards a preferable equilibrium. Not a mixed strategy Nash equilibrium of 1/3 dishonest and 2/3 honest (the ESS), but a pure strategy Nash equilibrium where everybody adopts the honest strategy (the cooperative equilibrium). This maximises the amount of available resources and ultimately yields a bigger payoff for all parties compared to the equilibrium reached when strategies are determined by the participants without any constraints. It therefore makes sense that groups adopting such an institutional framework would have a selective advantage (more resources and less costly conflicts) over groups that do not, and that through a group-selective process, the institution of property would spread throughout the human population.

4.2 Case Study 2: Commons Management

Common resources or common-pool resources, such as water reservoirs and fish or timber stocks, are meant for the use of the whole community. Given the nature of these resources—they must be pooled and can be depleted by opportunistic use (Ostrom 1990)—they are extremely vulnerable to free-riders. Again we are presented with a Prisoner's dilemma in which the non-correlated Nash equilibrium is the suboptimal situations in which the players don't cooperate ([1,1] in the payoff matrix hereunder).

	Cooperating	Freeriding
Cooperating	3, 3	0, 5
Freeriding	5, 0	1, 1

Hardin (1968) referred to this as 'the tragedy of the commons'. A group of rational and self-interested individual will eventually deplete common resources to the detriment of all. Indeed, it will always benefit the individual to behave against the whole group's common interest, given that it yields a personal payoff to do so. Good contemporary examples are the issues of pollution and global warming. It serves a company's or a country's economic interest (at least in the short term) to refrain from restricting pollution and therefore produce in an unhindered fashion, even though in the long run it is detrimental to all.

In order to move to the more desirable equilibrium in which all parties cooperate, we need to change the payoff matrix by imposing punitive measures to freeriding. In other words, rights and obligations need to be introduced by institutionalising the use of commons by means of explicit rules. Not surprisingly, this is exactly what happens. Ostrom's analysis (1990) of common resources, sets out a number of principles which are necessary to preserve commons: authorised users need to be determined, the rules of usage must be defined, the observance of rules must be monitored, and the violation of rules must be *punished*. Without this institutional framework, a lapse back into the suboptimal Nash equilibrium seems inevitable. By adopting the institutional framework and punishing violations, the payoff matrix is changed to:

[5] Note that the hawk-dove game or the pre-institutional property game is—strictly speaking—not a prisoner's dilemma. In a Prisoner's dilemma the Nash equilibrium is a pure strategy equilibrium in which both players defect. In the hawk-dove game, the Nash equilibrium is a mixed equilibrium: the ESS. The problem to be solved however remains to same: to get all parties to adopt a cooperative strategy thereby increasing their payoff in sum and on average.

Author's personal copy

	Cooperating	Free riding
Cooperating	3, 3	0, −1 (sanctioned)
Free riding	−1, 0 (sanctioned)	−1, −1(sanctioned)

This results in the more preferable outcome of [3,3], in which everybody cooperates and the common resource is preserved. Once again, groups who adopt this institutional framework will have a selective advantage over groups who do not. It should not surprise therefore that institutional frameworks regulating the use of common resources are found in all societies, from rules regulating fishing during breeding season to prevent fish stock to plummet in small fishing communities, over rules preventing overgrazing in pastoral communities, to the global directives formulated in the Kyoto protocol to attenuate global warming. Interestingly, experimental evidence suggests that we developed a very keen eye to detect cheating or freeriding. So keen in fact, evolutionary psychologists Tooby and Cosmides (1992) argue that we have evolved a separate cognitive module for it, a so-called 'cheater detection module'. This comes to show again that it is not so much a case of nature versus nurture, but both are—in the case of human social interaction—inextricably intertwined.

Boyer and Petersen (2012: 11–12) actually discuss the case of common management. They trace the institution in our 'competencies and motivations for fair exchange that are part of our evolved cognitive equipment'. While this is not strictly false, we did evolve these competencies and motivations, it is misleading. As pointed out, following Richerson and Boyd's (2005) account of gene-culture co-evolution, Boyer and Petersen seem to point the causal arrow in the wrong direction. Moreover, in this regard, it appears mistaken to take human nature as a given and see culture as a mere expression of that nature. Introducing the game-theoretic modelling of this area of interaction, it becomes clear how institutionalised commons management solves an important problem (the tragedy of the commons) posed by the pooled nature of particular resources and the strategies available to the members of a group. Reducing the institution to an expression of human nature, therefore, does little in explaining its emergence and prevalence.

4.3 Case Study 3: Marriage

While there are important differences in mating strategies between men and women, both sexes are innately predisposed to seek mating exclusivity from a partner with whom they are involved in a long-term mating relationship (mediated by the emotion of jealousy), all the while being tempted by mating opportunism (Buss 2006). The adaptive rationale behind this strategy is obvious, although different for both sexes. Men 'benefit'[6] from mating opportunism, given that this increases their offspring and might even yield offspring with better genetics (i.e. genetic material enhancing the offspring's chance on survival and reproduction). Women, on the other hand, 'benefit' from mating opportunism because it might yield genetic improvement in their offspring, and/or because it enables them to switch to a better mate/caretaker. The benefits of ensuring mating exclusivity from one's long term partner, on the other hand are: reducing the risk of raising offspring of a rival for men, and losing the care-taker to a rival for women. These are both very important drawbacks from an evolutionary perspective. The strategic interaction can be modelled as follows:

	Monogamous	Opportunistic
Monogamous	3, 3	0, 5
Opportunistic	5, 0	1, 1

Given the payoff matrix, it will always be more beneficial for either player to be opportunistic. The Nash equilibrium therefore is the one in which both players are opportunistic. The desirable equilibrium however is the one in which both players are monogamous ([3,3] instead of [1,1]). The institution of marriage, a 'contract' in virtue of which each player renounces opportunism to expect monogamy in return, enables all parties to move to the more desirable outcome.[7] Once again, the institution influences the strategies of the players involved which overall lead to a better payoff than the Nash equilibrium reached when both parties develop their strategy independently.

Interesting, in this regard, is Lewis's (1969) and Young's (1995) view of conventions. According to both authors, conventions are adhered to by virtually all members of a group, because the relevant behaviours are 'mutual best responses conditional on the expectation of similar behaviors by most others' (Bowles and Gintis 2011: 111). Marriage, in this sense, does exactly that. In contrast

[6] 'Benefit' here is intended in a purely evolutionary sense. Following Dawkins's (1976) gene centric point of view, it might be more accurate to state that genes rather than the individual benefit.

[7] In the case of polygamy, women trade desisting mating opportunism for caretaking. Given that the primary evolutionary 'interest' for women is ensuring that the care-taker will not desert the nest and withdraw resources and protection, the commonness of polygamy should not surprise. Polyandry, on the other hand, is extremely rare as can be expected from the evolutionary stakes outlined above.

Author's personal copy

to the first two cases, the preferable equilibrium is reached not by means of changing the payoff matrix externally (unless of course adultery is punished—which sometimes is the case), but by giving both parties extra guarantees that the other party will adopt the cooperative strategy.

According to Gintis (2009), this preferable equilibrium is a correlated equilibrium. A correlated equilibrium is a solution concept in game theory developed by Aumann (1987). It can be informally defined as an assignment of actions, by an external source (the 'choreographer') of correlated action recommendations to agents, so that nobody would want to deviate. Take traffic lights. They coordinate the actions of drivers meeting at a crossroad. If the light is green for me, I wouldn't want to stop which would cause me to lose time. If it's red, however, I wouldn't want to go either, which could cause me to crash, a negative outcome worse than losing some time waiting. Therefore nobody would want to deviate from the action recommendations of the traffic light—the 'choreographer'.

In a similar vein, Gintis (2009: 132–135) argues, social norms often function as correlating devices or 'choreographers'. Marriage is no exception. Given the prior knowledge that if either party breaches the contract and adopts an opportunistic strategy, he or she can expect that the other party will no longer uphold his or her part of the bargain, a rational player will not deviate from the social norm (in this case the norm enforcing monogamy in a marital relationship). This norm therefore implements a 'correlated equilibrium'. It correlates both strategies in such a way that it yields the best possible payoff for both players.[8]

Notice that even in the absence of punishment, the payoff matrix is affected. The mere fact that both players know that the cooperative strategy of the other player depends on their cooperation, adds a serious cost to the opportunistic strategy. Again, we can see why the institution of marriage (and the coevolving emotions of jealousy and love) emerged—often independently—in almost all human societies. In an evolutionary sense, both parties benefit from the trade-off. The institution of marriage, therefore, provides those adopting it with a selective advantage and can be expected to spread through the human population (again, together with and enforced by the emotions of love and jealousy).

[8] I'm very well aware that the evolutionary dynamics behind male and female reproductive strategies and therefore ultimately the institution of marriage are more complex than represented here. My account doesn't factor in the limited reproductive window for women for instance. Also in many cultures there are external punishments (be it only in the form of social punishment) imposed on those adopting the opportunistic strategies (especially for women). Nevertheless, this simplified account offers a neat way of showing that the outcome of an interactive game can be changed by 'correlating' the strategies of both parties.

It is important to note that the emergence of those basic social institutions is typically not the result of a conscious political decision, following a cost-benefit analysis. Institutions such as marriage and property predate modern and literate society, and can be expected to have emerged without any deliberation. More often than not, societies—at least the early human societies in which these institutions emerged and evolved—are blind to the exact payoff matrixes involved in these areas of interaction. In a very real sense, the causal mechanism that pushes groups to adopt those institutions, is not conscious reasoning and explicit design by the members or leader of the group, but the payoff structure inherent to those domains of interaction. Rather than implementing them, societies gravitate towards these institutions.

Boyd and Richerson, in this regard, argue that those universal, basic institutions evolved as the result of myopic responses of individuals to the incentives they experience (Boyd and Richerson 2008: 311). They illustrate this with the following example. Take two systems of inheritance: equal partition among brothers and primogenitor, only the oldest brother inherits. 'When brothers agree, they receive a higher payoff than when they disagree, because disputes are costly. This means that once either system becomes common, people with the more common belief achieve a higher payoff on average' (311). A given cultural group therefore can gravitate towards a particular model of interaction without conscious design. Universal social institutions, in this regard, are typically not products of top to bottom regulation, but bottom up emergence. And if they regulate the interaction between members of a group in a beneficial way (i.e. a way that increases the fitness of the members), they will be selected and spread throughout the human population.

5 Institutions are Game Changers

In all three case studies, institutions pushed the game to a new equilibrium. They change the outcome either by changing the payoff matrix by introducing punishment of non-cooperative behaviour) and/or by motivating both players to cooperate, by making the desired cooperation of the other party depend on one's own cooperation (marriage case). Institutions, in other words, incentivise. Social institutions, more particularly, incentivise us to cooperate and deter us from freeriding. They solve the evolutionary problem of cooperation (see Sect. 2) and do so by affecting 'the rewards and penalties associated with particular behaviors' (Bowles and Gintis 2011). Through the incentives they create, social institutions are able to steer the outcome of the interaction in certain domains to the cooperative equilibrium with the larger payoff.

Author's personal copy

This view of institutions as incentivising entities has been proposed by Buekens et al. (2013). With their theory, they target an influential account by Searle (1995, 2010), which maintains that institutions are irreducible to natural processes. According to Searle, the existence of an institution depends on the 'collective acceptance' (recognition) that an object X has a certain status-function Y (e.g. paper as money, lines as borders, passport booklets as providing legal entrance to certain foreign countries). Searle argues that these institutional facts cannot be explained in fully non-institutional terms. In other words, we cannot explain money or marriage without referring to other institutions such as states.

Recently, Buekens et al. (2013) have challenged this view. Drawing on Lewis's (1969) influential work on conventions, they developed the incentivised action view of institutional facts. According to their preferred view, institutions can be understood in virtue of the specific actions associated with the institution and the way we are incentivised to perform these actions. A border, for instance, divides two areas to which we are incentivised in a different way, money is an object that we are incentivised to acquire for exchange purposes, and so on. Note that this approach avoids invoking other institutions such as states and therefore escapes Searle's definitional circle. This approach, they argue, can be applied to all institutional objects. If institutional objects fail to incentivise, they cease being institutional objects. Bits of paper that nobody is incentivised to acquire for purposes of exchange stop being money, pieces of laminated paper that nobody is incentivised to acquire in order to certify one's right to drive on public roads stop being driver's licenses, etc. In other words, contra Searle, institutional facts such as money don't exist because of other institutional facts such as states (and this ad infinitum), but exist because of the incentives they create.

This account of institutions fits very well with the explanation of universal social institutions I'm proposing in this paper. In the three case studies presented, the institution emerged to regulate strategic interaction and push the outcome to a better equilibrium. It does so, every time, by incentivising the players to adopt a certain strategy (i.e. a different strategy than a self-interested and rational player would adopt in a non-institutionalised context). This incentive is created by changing the payoff matrix of a domain of interaction by either deterring players to adopt a certain strategy (negative incentive in the form of punishment) or motivating players (for instance by providing the guarantee that the other player will cooperate and cooperation can be maintained as long as nobody defects, as in the case of marriage). Universal social institutions, therefore, are game-changers. They change the payoff matrix of domains of social interaction by incentivising (through punishment and/or reward) the players to adopt a certain strategy. They change the rules of the game and consequently change the outcome of the game. They solve the evolutionary problem of cooperation and that—I argue—is why they emerged and prevailed.

6 Conclusion

In response to Boyer and Petersen (2012), I argue that it is misleading to frame universal social institutions as mere 'fits' with human nature. At best, this provides us with a partial explanation. These recurring social institutions should be seen as fitness peaks in the landscape of (biologically constrained) cultural possibilities, rather than cultural expressions of human nature. Game-theoretic models are particularly informative in the area of human social interaction. They clearly show how cultural evolution solved the problem of cooperation by changing the payoff matrix of Prisoner's dilemma situations, pushing the outcome from the original undesirable Nash equilibrium in which all players defect to the desirable cooperative equilibrium. Add to that the evolutionary dynamic of multi-level selection that favoured groups which reaped the larger payoff through those social institutions, and we have a clear account of the dynamics involved in the emergence and prevalence of a number of core social institutions. When it comes to universal social institutions, therefore, it is not a matter of nature *or* nurture, but very much one of nature *with* nurture.

References

Aumann R (1987) Correlated equilibria and the expression of Bayesian rationality. Econometrica 55:1–18

Bowles S, Gintis H (2011) A cooperative species: Human reciprocity and its evolution. Princeton University Press, Princeton

Boyd R, Richerson P (2008) Gene-culture coevolution and the evolution of social institutions. In: Engel C, Singer W (eds) Better than conscious? Decision making, the human mind, and implications for institutions. The MIT Press, Cambridge, pp 305–324

Boyer P, Petersen M (2012) The naturalness of (many) social institutions: evolved cognition as their foundation. J Inst Econ 8:125

Buekens F, Smit J, du Plessis S (2013) Developing the incentivised action view of institutional reality. Synthese. Available online-first

Buss M (2006) Strategies of human mating. Psychol Top 15(2):239–260

Dawkins R (1976) The selfish gene. Oxford University Press, New York City

Gintis H (2007) The evolution of private property. J Econ Behav Organ 64:1–16

Gintis H (2009) The bounds of reason: game theory and the unification of the behavioral sciences. Princeton University Press, Princeton

Author's personal copy

Gintis H, van Schaik C, Boehm C (in press) Zoon Politikon: the evolutionary origins of human political systems. Curr Anthropol

Gudeman S (1986) Economics as culture: models and metaphors of livelihood. Routledge & Kegan Paul, London

Hardin G (1968) The tragedy of the commons. Science 162:1243–1248

Hoebel E (1964) The law of primitive man: a study in comparative legal dynamics. Harvard University Press, Cambridge

Lewis D (1969) Conventions: a philosophical study. Harvard University Press, Cambridge

Ostrom E (1990) Governing the commons: the evolution of institutions for collective action. Cambridge University Press, Cambridge

Richerson P, Boyd R (2005) Not by genes alone: how culture transformed human evolution. The University of Chicago Press, Chicago

Searle J (1995) The construction of social reality. Penguin Books, London

Searle J (2010) Making the social world: the structure of human civilization. Oxford University Press, Oxford

Smith M, Price G (1973) The logic of animal conflict. Nature 246(5427):15–18

Sober E, Wilson D (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Harvard

Sperber D (1996) Explaining culture: a naturalistic approach. Blackwell, Oxford

Tomasello M (2009) Why we cooperate. The MIT Press, Cambridge

Tooby J, Cosmides L (1992) The psychological foundations of culture. In: Barkow J, Tooby J, Cosmides L (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, Oxford

Wright S (1932) The Roles of mutation, inbreeding, crossbreeding, and selection in Evolution. Proceedings of the sixth international congress on genetics 355–366

Young H (1995) Conventions. J Econ Perspect 10:105–122