An evolutionary account of reactions to a wrong

Giulia Andrighetto (giulia.andrighetto@istc.cnr.it)
Institute of Cognitive Science and Technology (ISTC-CNR), via San Martino della Battaglia 44, 00185, Rome, Italy
European University Institute (EUI), Fiesole, Italy

Francesca Giardini (francesca.giardini@istc.cnr.it)
Institute of Cognitive Science and Technology (ISTC-CNR), via San Martino della Battaglia 44, 00185, Rome, Italy

Rosaria Conte (rosaria.conte@istc.cnr.it)
Institute of Cognitive Science and Technology (ISTC-CNR), via San Martino della Battaglia 44, 00185, Rome, Italy

Abstract

In this work, we propose an evolutionary account of reactions to a wrong as an integrated set. Unlike other theories, we are not interested in revenge, punishment or sanction *per se*, but in their co-existence. We posit that this variety of reactions is needed in order to achieve different goals, but it also implies an increase in cognitive costs that requires to be explained from an evolutionary perspective. Moving from the identification of the psychological traits that uniquely define each reaction, two concurrent hypotheses are suggested and discussed: either the richness of human social life requests a variety of reactions, or the benefits of single reactions at the psychological level allowed these reactions to be maintained in the social life.

Keywords: Evolution; punishment; revenge; sanction; cognitive influencing; norms; enforcing institutions social order.

Introduction

Human actions are potentially unbounded and much more opportunities are available when other people are involved. When talking about social actions we have to distinguish between actions and reactions, i.e., actions triggered by someone's else previous action. Reactions are a constitutive part of living in societies, and the ability of displaying the appropriate reaction in the right content is extremely important for our "ultra-social" species (Richerson & Boyd 1998, 2005; Hill et al. 2009). The nature and the intensity of reactions depend on both the actor and the triggering action, and it requires the capacity to forecast further reactions and to plan ahead, among other things. Humans are unique under this respect, and everyone has experienced how many reactions the same individual can display in response to the same action, even in the same context. Animals can modulate their reactions, in some cases they can also decide their behaviour on a cost-benefit analysis, but others' representations do not enter this picture (Clutton-Brock & Parker 1995; Jensen et al. 2007). Humans react because of what they believe and want, and because of what they want others to believe and of how they want them to behave.

A particularly interesting class of social reactions is that triggered in response to a *wrong*. Retaliation, revenge, punishment and sanctions have been a matter of interest since the rise of Western culture, as witnessed by the fact that the need to understand and explain motives for reacting to wrongs never ceased since pre-Homeric Greece to these days (for an analysis of the differences among these reactions see Giardini et al. 2010).

Philosophers, social scientists, political scientists, psychologists, anthropologists have been striving to answer the fundamental question: why do people react to a wrong? In many circumstances reacting is more costly than standing, it requires some kind of planning, and it also implies the possibility of suffering a counter-reaction. Even more striking, people react to wrongs suffered by strangers, intervene in others' disputes, and sanction others when failing to comply with norms that they are not supposed to enforce. Although several scholars have been interested in explaining the evolution of revenge, punishment and sanction (Lorenz, 1966; Hamilton, 1970; Boyd & Richerson, 1992; Clutton-Brock & Parker 1995; Gardner & West, 2004; Jensen, 2010), these phenomena have been usually considered in isolation and not as a rich and complex repertoire. We propose that revenge, punishment and sanction are different reactions that should not be considered in isolation but as interdependent and complementary. If we look at them as an intertwined set, we need to explain the reason why they are different, but we also need to understand why we still have more than one reaction to an offense, and how the related extra cognitive costs are compensated. Our goal is to explain the decision to apply punishment in terms of the complementary decisions to use neither revenge nor sanction, thus understanding the motives behind each and every reaction.

We propose that this variety is necessary because, unlike animals, humans' reactions do not only target the offender's behavior, but also her mental states, as well as the victim's mental states. Comparing different reactions, we highlight an evolutionary trajectory that links revenge, punishment and sanction by explaining costs and benefits of each reaction. Having the opportunity to choose among several responses means higher cognitive costs to select between actions, and to choose the most appropriate one. Therefore, a set of questions arises: Why do we have such a repertoire? Can we identify evolved mechanisms that allowed us to

distinguish among reactions and to selectively apply them depending on the context?

Moving from the identification of the psychological traits that uniquely define each reaction, we propose a complex relationship between the richness of human social life, which requests a variety of reactions (society → individual motivations), and the benefits of single reactions at the psychological level. These benefits favoured the maintenance of reactions in the social life (individual motivations → society).

The rest of the paper is organized as follows: Section 1 outlines the evolutionary model, Section 2 defines the different phenomena and Section 3 introduces the main features of our taxonomy.

An evolutionary account of reactions

Revenge, punishment and sanction are superficially similar but deeply different in terms of the evaluation of the wrong suffered (or its interpretation), the intended goal, the consequent cognitive influencing, the temporal dimension and the kind of target. Humans are usually effective in administering punishment, i.e., in selecting the best reaction, taking into account the differences and selecting, through a fast and efficient process, how to react according to the external circumstances and their internal states.

The computational demands associated with the choice are not negligible and the risks of a mismatch between the perceived wrong and the reaction are high. We propose that revenge, punishment and sanction require the evolution of specialized mental mechanisms regulating the activation of different responses to wrongs or rule violations. We suggest that humans have mechanisms designed to produce revenge, punishment and sanction that evolved because of their effectiveness in solving recurrent social problems that humans encountered during evolution (Petersen et al. 2012). Given the richness of human bonds and social life, the need for acquiring social bonds (Dunbar, 1996) and for maintaining them, also remembering who is related to whom, could have favoured the selection of different reactions that have different consequences in terms of relationships (McCullough, Kurzban, Tabak, 2012). Moreover, psychological benefits of reactions may motivate their maintenance at the individual level and thus foster their selection at the social one. On the one hand, restoring the status quo, achieving deterrence or promoting the norms are goals that cannot be achieved through a single reaction, and their related specific benefits at the psychological level may have prompted the maintenance of multiple responses. On the other hand, the costs of selecting among different reactions are not negligible, also because having more choices implies being more prone to errors, with negative consequences arising at both the individual level and the social level. In the latter case, this mismatch between the reaction chosen and the wrong suffered can be extremely dangerous, and it may challenge the social order. Avenging a wrong when there is a social norm and the related sanction, or punishing someone in a context in which revenge was expected could result in a negative judgment about the reacting agent. Failing in interpreting correctly the situation and thus applying an inappropriate reaction may lead individuals to consider the avenger/punisher/sanctioner as socially inadequate and to avoid interactions with her. At the group level, frequent failures in using the appropriate reaction may undermine the cohesion of the group and make it more vulnerable to turmoil and fights.

In evolutionary terms, the risk associated with the application of the wrong kind of reaction were compensated by the evolution of specific psychological mechanisms for selecting among reactions.

Each and every reaction involves some unintended side effects, which may prevent the agent from achieving her goals and may also make the reaction inappropriate. In revenge, making the other suffer and regaining one's sense of control, together with restoring the status quo, require the agent to evaluate the wrong suffered and to estimate how much sufferance to inflict on the offender. Since there is not any objective criterion to estimate the sufferance experienced, this evaluation can only be subjective, thus exposing the avenger to the risks of damaging his reputation because the reaction was disproportionate (too harsh or too weak), or loosing social ties, or even triggering a feud with escalation of violence. Feuds are especially costly at the group level and they may even lead to the dissolution of the group. The punisher aims at deterring the wrongdoer from further hostility (by making it a costly option). There is not a pre-established and socially shared set of rules that govern how to punish. This lack of explicit and objective regulation can have several negative consequences. If the punishment inflicted is not appropriate in quantity or in kind, this can result in perception of the punishment received as unjustified, not legitimate and unfair. When punishment is perceived inappropriate it may also become ineffective in inducing deterrence, so the punisher is not able to achieve her main goal. In addition, the punisher can acquire a bad reputation for being too harsh, and she can see some social ties severed because of his action with the consequent risk of an escalation of violence, which has consequences for the whole group.

The risks of administering an inappropriate sanction are more limited and they are mainly related to the fact that the normative message is not clearly understood by those who receive the sanction. Therefore, when the normative character of the situation is not recognized, the sanction is ineffective, and the normative belief and the normative goal will not be formed in the mind of those who receive the sanction. An inappropriate sanction may also lead to counter-reactions, either in the form of a further sanction or as a retaliatory behavior.

In what follows we will detail our model of reactions, specifying the cognitive underpinnings and the dimensions of change characteristics of each and every phenomenon, and then supporting our model with a discussion of the relevant literature.

Distinguishing among reactions

Although a number of accounts (for some representative work see Bowles & Gintis 2004; Henrich & Boyd, 2001; Henrich et al. 2006) have stressed the relevance of punishment in human societies, they suffer the flaw of considering punishment as a single behaviour. In our view, punishing actually consists in a complex behavioral repertoire in which it is useful to disentangle at least revenge, punishment, and sanction. In Giardini, Andrighetto, Conte (2010) it has been argued that this variety of punishing strategies can be differentiated on the basis of 1. their mental antecedents, 2. the way in which they influence the future conduct of others, and 3. the effects they aim to achieve. Having more than one available strategy allows humans to tailor their reactions and to achieve their goals more easily but, at the same time, this implies higher cognitive and computational costs. In fact, agents must be able to categorize actions in the correct way, meaning that the context has to be interpreted adequately, the most appropriate reaction has to be chosen on the basis of the perceived wrong, of the situation, and of the offender and other agents' mental states. This calculation leads to a significant increase in the computational costs, which should be compensated, by some sort of benefits. Revenge, punishment and sanction result from psychological adaptations that allowed to solve recurrent conflicts that humans encountered during their evolutionary history, but we still do not know why we have more than one mechanism.

If animal societies are able to cope with aggressions by using just one form of reaction (Clutton-Brock & Parker, 1995), usually termed "punishment", why do we need a collection of counter-reactions? What are the fitness benefits coming from revenge, punishment and sanction?

In our theoretical analysis of reactions to a wrong, we start by providing a preliminary list of the core elements that determine the kind of response that an individual will choose in response to an aggression (see Table 1):

- The *wrong suffered*, i.e., the cause of the response. The evaluation of the offense depends both on the intentions the aggressor (the offense was intentional *vs* the offense was not intentional), and on the nature and value of the goal(s) frustrated by the aggressor.
- The *goal* of the reaction. When deciding how to react to an aggression, individuals consider the goal(s) they want to achieve and then select the appropriate reaction.
- The kind of *influencing* the agent reacting wants to apply to achieve her goal(s). Our theory is based on the idea that different reactions are aimed to produce different changes in the mind-set of the victim. For example, the avenger is aimed at acting at the epistemic level, by changing the target's and audience's beliefs about herself. The punisher aims to act both at the epistemic and motivational levels, by generating in the victim's mind the *goal* –

usually under threat of punishment– of abstaining from doing the action that has triggered punishment again. Finally, the sanctioner wants to endow the offender with new normative knowledge and to generate in her mind the goal to comply with the norm in the future.

- The focus of the reaction refers to the agent herself (as it is in revenge), another agent (as it is in punishment), or a norm (as it is in sanction).

It is worth noticing that we do not consider reactions as clear-cut phenomena, but they are overlapping in several respects. In Table 1, we summarize the main features of each reaction, in an attempt to identify the key elements of each phenomenon.

Table 1: Dimensions of change

	Revenge	Punishment	Sanction
Wrong suffered	Intentional aggression; Frustration of a personal goal of the agent; Sufferance experienced	Intentional aggression; Frustration of a personal or social goal of the agent.	Norm violation
Goal	Making the aggressor suffer; Status quo restoration	Deterrence	Norm recognition; Norm compliance
Cognitive influencing	Beliefs	Beliefs Goals	Normative Beliefs Normative Goals
Temporal dimension	Backward-looking	Forward-looking	Forward-looking
Focus	Self	Other	Norm

Figure 1 depicts the mental path that triggers to choose a specific reaction. In the following section, a cognitive anatomy of revenge, punishment and sanction will be provided. In section 3, an analysis of the intended and unintended effects of the three reactions is presented. The latter analysis will allow us to sketch an evolutionary explanation of why we have more than one reaction to an offense, and how the cognitive extra-costs resulting from this variety of reactions are compensated.
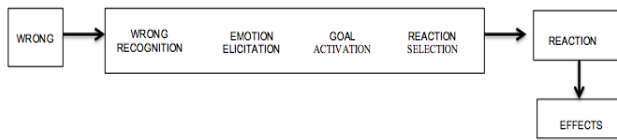
Figure1: The cognitive path of reactions to a wrong

A further consideration involves the role of emotions. The specific role that anger, but also social emotions, such as moral outrage, pride, shame, guilt, indignation, contempt, disgust, resentment, etc., (e.g. Fessler & Haley 2003; Frank 1988) play in triggering the reactions under study deserves an attentive theoretical and experimental analysis. Although crucial, this analysis is beyond the scope of the present paper and will be developed in future work.

Taxonomy of reactions to a wrong: Revenge, Punishment and Sanction

Revenge

Revenge, according to the Merriam-Webster dictionary, is "punishment inflicted in retaliation for an injury or offence". In Elster's terms (1990) it is "the attempt at some cost or risk to oneself, to impose suffering upon those who made one suffer, because they have made one suffer" (p. 862). Broadly speaking, the term 'revenge' refers to two diverse but connected phenomena.

In the first of these phenomena, revenge is a *social ritual* that requires and prescribes specific behaviours to group members to repair an offence. Ethnographic studies highlighted the transition from tribal to modern societies, in which retributive concepts of law and the creation of institutions replaced vengeance and avoided blood feuds (Boehm 1986). Posner (1980) suggests that revenge and retribution may be partially determined by historical and economic circumstances, such the private enforcement of law and high probabilities of detecting and punishing offences. When these conditions are met, a pure vengeance system may appear, although it is unlikely to be optimal. These systems are not completely extinguished, as the culture of honour in the southern United States (Nisbett 1993; Nisbett and Cohen 1996) and the Kanun in Albania demonstrate. The Kanun, a customary set of laws used mostly in northern Albania and Kosovo, disciplined people's reactions to murder (blood revenge or *gjakmarrje*) and other offences (*hakmarrje*), according to the roles and degree of kinship of all the people involved. Shirking revenge or taking it without respecting what is stated in the Kanun leads to the same result: honour cannot be restored and the whole family or clan is to blame. Apparently, the Kanun has not disappeared completely, and in some areas it is still observed, showing how an institution that is preserved in the mind can out-compete another centrally enforced institution, because the latter one is not recognized as such.

The other way of looking at revenge is to consider it as an *individual* behaviour, which is present both in human societies (Zaibert 2006), and non-human primate groups (Jensen, Call and Tomasello 2007). Turning our attention to individual factors it becomes possible to provide a cognitive anatomy of this reaction. The avenger wants to repay the damage she suffered with an equal or greater offence, no matter how risky or dangerous this retaliation is. In a sense, we can say that the avenger is a *backward-looker* who revolves around the past and acts in the present to rebalance what happened, with no concern for the future. Unlike other authors (McCullough, Kurzban, Tabak, 2012), we do not see vengeance as a means to affect the likelihood that the wrongdoer will repeat the aggression in the future, inducing her to cooperate next time or deterring her from further aggressions. Long term, strategic planning does not seem to characterize the avenger's mind, although unintended deterrence effects can be obtained.

Revenge is motivated not only by the desire to make the target suffer, but also by the goal to change the target's and audience's beliefs about the avenger, in order to restore the image that has been damaged by the aggression suffered. In this case cognitive influence is aimed at changing the *beliefs* of the wrongdoer and of the audience: the avenger aims to repay the damage she suffered with an equal or greater offence in order to change the target's and audience's beliefs about himself. Revenge is a way to regain one's position after an offence and this applies also to the symbolic dimension: the avenger wants to restore her image, damaged by the aggression suffered. Revenge is aimed to modify what the others believe about the avenger, her role and status. Presumably, the greater the offence, the more efficacious the image restoration and the effort to restore the status-quo.

Punishment

Enforcing institutions have evolved with society: starting out as simple systems of revenge and retribution imposed by the individual, family, or tribe, in modern societies they grew as institutions characterized by a higher concern for *deterrence* and rehabilitation. Institutions controlling modern societies moved from systems based on revenge to ones based on punishment. In primitive society enforcement was left to the individuals wronged, or their families, and was vindictive or retributive (Boehm 1986): in quantity and quality it would bear no special relation to the character or gravity of the offence. Gradually it arose the idea of proportionate punishment, of which the characteristic type is the *lex talionis* of early Roman law or in the Old Testament and Koran. Like revenge, also punishment refers to two distinct class of phenomena: punishment is both a *social* institution and an *individual* behavior.

As an institution, punishment serves to dissuade people from engaging in activities deemed wrong by law and by the society itself, thus reducing the frequency and likelihood of future offences. Deterrence theory suggests that punishment works by modifying the relative costs and benefits of

situation, so that wrongdoing becomes a less attractive option (Bentham 1962; Becker 1968). Punishment possibly has the effect of preventing blood feuds and giving more stability to the social order.

As an individual behaviour, punishment is a reaction intentionally aimed to minimize the chance that the aggressor will repeat the act again (Giardini, Andrighetto and Conte 2010). Unlike revenge, punishment is not inflicted in retribution for an offence or transgression. The punisher is driven by *forward-looking* considerations, and deterrence is intentionally pursued.

This enforcing mechanism, controlling modern societies, is not at all easy to distinguish from revenge (Zaibert 2006), at least from a mere behavioural point of view. Cognitive modelling allows us to disentangle them on the basis of their mental antecedents and the way in which they influence the future conduct of others. The punisher and the avenger are aimed at influencing and modifying the target and the audience's minds in different ways: unlike the avenger, the punisher has the explicit goal to deter the wrongdoer from repeating the aggression in the future. To achieve this goal, the punisher should act in such a way that the offender, and possibly the audience, generates in her mind the *goal* – usually under threat of punishment (i.e., by generating the belief in the victim's mind that future aggressions will be punished)– of abstaining from doing the action that has triggered punishment again.

Sanction

Social order can be explained as the mere result of the deterrence effect of punishment. However what makes human cooperation so spectacular with respect to all other species is the presence of social norms, efficiently orchestrating social life. When punishing institutions are able to work in tandem with social norms, they are much more viable and effective in achieving and maintaining compliance and are more robust across time (Andrighetto and Villatoro, 2011; Villatoro et al. 2011).

By analyzing a large number of spontaneously emerged institutions in different countries, the political scientist Elinor Ostrom has identified a set of characteristics that make them successful in promoting social order. She suggests that the most effective institutions are those that facilitate norms' elicitation, their spreading, and compliance (Ostrom 2005; see also Casari 2007). Punishment, when properly designed, should tell people which behaviours are acceptable, i.e., the (social) norms regulating society, and which actions will cause punishment.

We refer to punishing institutions enforcing social order through mechanisms intentionally aimed to focus people's attention on social norms and to condemn their violation as *sanction institutions*. We consider sanction institutions as the last step of the institutional evolutionary process.

As in previous work (Giardini et al. 2010; Andrighetto and Villatoro, 2011; Villatoro et al. 2011), we use sanction to indicate the enforcing individual behaviour that, in addition to imposing a cost for the wrongdoing, as punishment does,

is also intentionally aimed at *signalling* norms to the offender (and possibly to the audience) so that she will comply with them in the future.

The type of cognitive influencing sanction exerts on the offender is more complex than those in revenge and punishment. In order to deter future norms' violations, the sanctioner endows the offender with (new) normative knowledge. The sanctioner uses scolding to reign in wrongdoers, expresses indignation or blame, or simply mentions that the targeted behaviour violated a norm. Through these actions, the sanctioner aims to *focus* people's attention on different normative aspects, such as: (a) the existence and violation of a norm; (b) the causal link between violation and sanction: "you are being sanctioned because you violated that norm" (c) the probability that violations will be sanctioned; (d) the fact that the sanctioner is acting as a norm defender. As recent psychological and economic experimental evidence shows (Cialdini et al. 1990; Bicchieri 2006; Galbiati and Vertova 2008; Houser and Xiao 2010), the norm focusing effect of sanction plays an important role in eliciting norm compliance. Thus, despite punishment, we suggest that sanction has the further effect, possibly aimed at by the sanctioner, to encourage the target to ground future decisions on internal evaluative criteria, established by the norm. By facilitating the spreading, recognition and internalization of norms, sanction possibly has the effect of promoting social order in a more stable and less costly way with respect to punishment.

Concluding Remarks

In this work, we proposed an evolutionary account of reactions to a wrong as an integrated set. Unlike other theories, we are not interested in revenge, punishment or sanction *per se*, but in their co-existence. We posited that this variety of reactions is needed in order to achieve different goals, but they also imply an increase in complexity, due to the costs associated with the interpretation of the situation and the selection among reactions. We proposed that the transition from one to the other has been allowed by specific cognitive patterns, and suggesting that these mental mechanisms selected among given social structures, at the same time reinforcing and being reinforced by them.

Modifying others' actions require a set of cognitive skills that allow to represent others' mental states, considering the way in which these are harbored in one's mind, giving rise to social beliefs, namely beliefs about others' mental states (e.g. beliefs, intentions, desires, emotions). In addition, this requires also cognitive influencing, as the willingness to modify others' goals. Having a set of available reactions means that individuals should also be endowed with cognitive mechanisms to recognize which reaction is more appropriate in a given situation.

This theoretical analysis will be

Acknowledgments

References

Andrighetto, G., & Villatoro, D. (2011). Beyond the carrot and stick approach to enforcement: An agent-based model. In *European Perspectives on Cognitive Science*, edited by B. Kokinov, A. Karmiloff Smith, N.J. Nersessian. Sofia: New Bulgarian University Press.

Bentham, J. (1962). Principles of penal law. In J. Bowring (Ed.). *The Works of Jeremy Bentham*. (pp. 365-580) New York: Russell and Russell.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76: 169-217.

Bicchieri, C. (2006). *The Grammar of Society: the Nature and Dynamics of Social Norms*. Cambridge (MA): Cambridge University Press.

Boehm, C. (1986). *Blood Revenge: The Enactment and Management of Conflict in Montenegro and Other Tribal Societies*. Philadelphia (PA): University of Pennsylvania Press.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 61: 17–28.

Boyd, R. & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol*. 13: 171 – 195.

Casari, M. (2007). Emergence of endogenous legal institutions: Property rights and community governance in the Italian Alps. *The Journal of Economic History* 67(1):191-226.

Cialdini, R. B., Reno, R.R. & Kallgren, C.A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58: 1015-1026.

Clutton-Brock, T. H., & Parker, G.A. (1995). Punishment in animal societies. *Nature* 373: 209-216.

Elster, J. (1990). Norms of revenge. *Ethics* 100(4): 862–885.

Fessler, D. M. T. & Haley, K. J. (2003). The strategy of affect: emotions in human cooperation. In P. Hammerstein (Ed.). *Genetic and cultural evolution of cooperation* (pp. 7 – 36). Cambridge, MA: MIT Press.

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions*. New York, NY: Norton.

Galbiati, R., & Vertova, P. (2008). Obligations and Cooperative Behaviour in Public Good Games. *Games and Economic Behavior* 64(1): 146-170.

Gardner, A. & West, S. A. (2004) Cooperation and punishment, especially in humans. *Am. Nat.* 164: 753 – 764.

Giardini, F., Andrighetto, G., Conte, R. (2010). A cognitive model of punishment. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1282-1288). Austin, TX: Cognitive Science Society.

Gjeov, S. (1989). *Kanuni i leke dukagjinit. Albanian text collected and arranged by Shtjefn Gjeov*. New York: Gjonlekaj Publishing Company.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature,* 228: 1218–1220

Henrich, J., & Boyd, R. (2001). Why people punish defectors. weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208: 79–89.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767-1770.

Hill, K., Barton, M. & Hurtado, A. M. (2009). The emergence of human uniqueness: characters underlying behavioral modernity. *Evolutionary Anthropology,* 18, 187 – 200

Houser, D., & Xiao, E. (2010). Understanding context effects. *Journal of Economic Behavior & Organization* 73(1): 58-61.

Lorenz, K. (1966). *On aggression*. London, UK: Methuen.

Janssen, M. & Bushman, C. (2008). Evolution of cooperation and altruistic punishment when retaliation is possible. *J. Theor. Biol.*, 254, 541 – 545.

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365, 2635–2650.

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences* 104: 13046–13050.

McCullough, M.E., Kurzban, R., Tabak, B.A. (2012). A cognitive system for revenge and forgiveness. *Behavioral and Brain Sciences,* 1-58

Nisbett, R.E. (1993). Violence and U. S. regional culture. *American Psychologist* 48: 441- 449.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor*. Boulder (CO): Westview Press.

Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton (NJ): Princeton University Press.

Petersen, M. B., Sell, A., Tooby, J. & Cosmides, L. (2012) To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior*. Advance online publication.

Posner, Richard. 1980. "Retribution and related concepts of punishment." *Journal of Legal Studies* 9: 71–92.

Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. In I. Eibl-Eibisfeldt & F.Salter (Eds.). *Ideology, warfare, and indoctrinability; evolutionary perspectives* (pp. 71 – 95). London,UK: Berghahn Books.

Villatoro, D., Andrighetto, G., Conte, R., & Sabater-Mir, J. (2011). Dynamic Sanctioning for Robust and Cost-Efficient Norm Compliance. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (IJCAI 2011)

Zaibert, L. (2006). Punishment and Revenge. *Law and Philosophy* 25: 81-118.