The evolution of conscience – from parent-offspring conflict to morality

Eckart Voland

Institut für Philosophie, Universität Giessen, Germany

Abstract: In this paper, two theories regarding the biological evolution of morality with conscience as its central regulatory agency are compared and contrasted. One theory (“navigator theory”) interprets conscience as a strategically operating agency for the optimum balance between selfish and altruistic behavioral tendencies to maximize gains in cooperation in view of social complexity. From this standpoint, conscience serves the evolved self-interest of the person having a conscience. In contrast hereto, the second theory (“helper theory”) locates the evolutionary origins of conscience on the battlefield of the parent-offspring conflict through intrafamilial demands for altruism. Functions of conscience, and thus human morality in a narrower sense, evolved during the transition of hominines to cooperative breeding and the novel helper conflict emerging through this evolution. The “helper theory” of the evolution of conscience can resolve some of the theoretical and empirical inconsistencies of the conventional “navigator theory”, in particular, the contradiction between the consequentialistic regulation of altruistic behavior and the non-consequentialistic nature of the judgment of conscience. And in contrast to the “navigator theory”, it is compatible with the observation that behavior guided by a conscience is not infrequently disastrous for one’s own fitness outcome.

Keywords: parent-offspring conflict, cooperative breeding, helper conflict, non-consequentialism

Introduction

“…strange if judgment remains, where reason is forgotten, it is conscience”

Charles Darwin 1838

During the last decades, a great deal of work has been done with the aim of achieving a better understanding of under what biographical, social, and ecological constraints cooperative and altruistic behavior maximizes lifetime reproductive success and therefore endures in a Darwinian world of the “survival of the fittest”. Clearly, cooperative and altruistic strategies no longer appear to be those paradoxes which must have caused Darwin many headaches, but in accordance with Hamilton’s turn in biology, most of them fit into existing evolutionary theory quite smoothly.
The essential ideas on the evolution of altruism and cooperation are referred to in the literature under the headings of “mutualism”, “biological markets”, “reciprocal altruism”, “indirect (or strong) reciprocity”, “altruism as a costly signal”, and “nepotistic altruism” (Kurzban et al. in press; Price 2011). All these strategies have in common that – on average – they offer an optimal solution to the problem of cooperation that benefits the reproductive fitness of the altruistic individuals. Now the question is: How much of human altruistic behavior can be explained by these strategies, which are alleged to be successful from an evolutionary standpoint, because on average they lead to personal fitness benefits in the lifetime balance, even if they coincide with short-time costs? Those who maintain that altruistic behavior can be explained completely by these well-described strategies would argue that consequently the biological evolution of the behavioral component of human morality is understood in principle, because morality, as widely conceived, manifests itself as altruistic behavioral tendencies. However, human morality has not only a behavioral, but also a cognitive component. People are able to make moral judgments. They process a competence the evolution of which is anything but clear. Moral judgment, at least moral judgment with a reference to one’s self, is processed and has an impact on behavior through an agency that has been called “conscience”. Strangely enough, this term is almost completely missing in the subject indexes of the textbooks on evolutionary psychology. This omission could be related to the fact that an evolutionary understanding of the performance of conscience is definitely balky, for one reason alone: The production of altruistic behavior obeys conditional rules, e.g., by being contingent on the probability of reciprocation, or on kin relationships or the presence of a watching audience. The moral judgment generated by conscience does not correspond to these conditional rules; instead, it is unconditional and impartial. In some sense, the judgment of conscience is non-consequentialistic, because individuals making a moral judgment are mostly in possession of the relevant information regarding the consequences of their actions, especially for themselves. Nevertheless, they judge (and possibly act) irrespective of possible consequences for themselves.

This observation motivates the central question of this paper. It is the question of how – in a Darwinian world, in which the features of the organisms are selected according to the consequences that these features have for their own genetic basis – actors who operate non-consequentialistically can, at least occasionally, arise and successfully persist in such a world although they face severe risks when behaving this way. In this paper I am going to explain a possible answer to this question, as originally developed in Voland and Voland (2014).

From Conditionality to Non-Conditionality, from Altruistic Behavior to Moral Judgments

What is Conscience and what Does It Do?

When reference is made to “conscience” in the following, the mental agency characterized below is meant, which plays a significant role in moral self-regulation. Please note that contextual manifesta-
tions of conscience are not dealt with here. Of the latter, it may be assumed that they show a certain ethno-historic variability. Instead, the biological evolution of this agency is going to be discussed here, not the history of its mode of action.

Despite the fuzziness inherent in the term “conscience”, there appears to be a consensual semantic cluster – at least from the standpoint of everyday psychology – with which conscience can essentially be paraphrased: The central descriptions revolve around expressions such as “inner policeman”, “self-regulation” and “torment”. In detail, the following traits are attributed to conscience:

Asymmetric evaluation system: The internal torment evoked by a bad conscience is not paralleled by any similarly effective emotional composure of a good conscience. The reward for a good conscience is obviously only comprised of having escaped the torment coming from a bad conscience. “Personal guilt is in fact the default condition of the conscience”, is how Strohm (2011:34) puts it and “A good conscience is only the absence of a bad conscience”. Not being rebuked is enough of a reward for a good conscience, whereby the rebuke is experienced as personal guilt and shame. Guilt is a moral emotion which plays a role in social transactions with others, whereas shame results from difficulties in accepting one’s being as it is. It is inherent to both guilt and shame that they lead to a reduction in the feeling of self worth (Tyson and Tyson 1993).

A second perspective: The ability to experience moral guilt and shame presupposes the capability of taking a second look, namely a look at ourselves from a different perspective. This second perspective makes us, as Nietzsche expressed it in the preface to Zur Genealogie der Moral, “strangers to ourselves”. This aspect is even reflected in the etymology of this term, because the meaning of the Greek term “syneidesis” describes joint knowledge, and interestingly enough, initially in prosecutable, i.e., culpable contexts. In Latin, this term was translated into “conscientia”. The Old High German word “gewiazza-ni” was, in turn, borrowed from “conscientia”. Thus conscience is the gateway for the perspectives and concerns of others, albeit unpersonal others, which massively challenges our certainty of personal integrity and intimacy (Strohm 2011).

Domain generality: Conscience is not associated with genuine content. It can become active in different behavioral domains, and accordingly motivate various types of actions. Attempts to determine the content of conscience fail due to their dependence on the subject (Tesak 2003). Therefore, the impact of conscience is contingent upon history, culture, and especially biography. It can produce an extremely high degree of moral solidarity, but also pure terror.

Potential for nonconformity. Conscience does not deplete itself in a bundle of internalized social norms which generate social conformity (Tesak 2003). It can also motivate social rebellion. Democratic legal systems base a right to resist on conscience and accept it as the final moral decision-making agency. Therefore, there must be a surplus function which goes beyond securing social conformity.

Motivational pressure: The pressure exerted by conscience requires action. The inner voice of conscience can not be pacified as long as it has not led to a certain action. There is obviously no even-
tual habituation to feelings of guilt and shame (Bischof 2012).

Non-cognitivism: The results of the social and cognitive neurosciences have, in recent years, provided knowledge about the development of moral judgments, which tend – overall – to reinforce a non-cognitivistic position. Accordingly, the moral judgment coming from conscience is not the result of a consciously controlled, rational process, but of one that is automated, intuitive, and guided by feelings (Haidt 2012). It judges without input through rational arguments and may be deaf to legal or philosophical expertise.

Non-consequentialism: The moral judgment of conscience tends to be oriented to rules instead of results. This significantly increases the risk that the fitness outcome of a certain conscientious behavior might be negative. In everyday social practice, this risk is relatively rare, however, because conscience operates mostly in a stand-by mode. However, it goes into operation occasionally. These are the crucial moments for the functioning of conscience, in which the takeover of net fitness losses are on the agenda. Probably due to this non-consequentialism, Freud sees in the super-ego a potential for “hypermorality” and speaks of its “mercilessness” (Tyson and Tyson 1993).

Egocentrism. In contrast to moral emotions, which also become perceivable given third-party actions and omissions, the activity of one’s conscience remains restricted to one’s own actions (Tesak 2003). As a member of a moral society, one can evaluate the behavior of others ethically – something that happens constantly in fact– but this occurs in a more distanced and uninvolved way, unless one’s own immediate concerns are affected. A conscience with a threat of self-punishment is not required to speak about something as right or wrong in others. Conscience only monitors one’s own actions and omissions, i.e., that arena where guilt and shame make their appearances. The inner policeman is only responsible for oneself.

This brief paraphrase designates aspects of conscience which are likely to be not unimportant when looking at the evolution of human morality. If very obviously the central function of conscience comprises the threat of self-punishment for non-compliance with moral rules, the question arises of what special problem it has adapted to during evolution. Why can other animals (presumably) live without a conscience even though they are also capable of altruism? What caused the increase in complexity of human social life so that evolution had to produce a regulatory agency of its own? Briefly: What is conscience good for?

Where is conscience located within Darwinian Theory?

Natural selection is necessarily a consequentialistic principle. The features of the organism – including behavioral features and life strategies – are selected in accordance with their net contribution to lifetime reproductive success, i.e., according to their consequences for genetic fitness. Now the cost-benefit ratio of a behavioral strategy can vary individually, locally and temporally, which is why evolution has provided organisms with conditional behavioral strategies. With reference to human moral practice, this means that we have to expect moral flexibility: Value relativism, moral opportunism and “double standards” – or in the words of Batson (2008): “moral masquerades” and “hypocrisy” – are log-
The evolution of conscience

ical results of a consequential selection of conditional strategies. According to Boehm (2012:32) conscience also fits into this picture: “[T]he conscience also has its Machiavellian functions, for it can guide us to take a flexible approach to being moral that allows us to profit from having a decent reputation and at the same time judiciously cut the occasional not-too-serious corner and profit from doing so. How should a reproducively useful conscience be designed, then? First, in the Darwinian competition among individuals a conscience shouldn’t be too weak because this can lead to personal disaster. Nor should it be too strong, for the internalization of rules shouldn’t be too inflexible”. With this adaptive interpretation of conscience, Boehm (2012) extends the point of view originally contributed to the discussion by Alexander (1979:133). Conscience, he said, was “the still small voice that tells us how far we can go in serving our own interests without incurring intolerable risks. It tells us not to avoid cheating but how we can cheat socially without being caught”. Regardless of how the social pressure for the adaptive formation of conscience is reconstructed in detail, these considerations are based on the conviction that in the final analysis, conscience has to be associated with fitness benefits for the person acting.

In contrast to these authors, I see the role of conscience in the evolution of morality as anything but clear. The kinds of altruistic behavior to be observed in non-human animals obviously occur without a conscience. Even though punishing and policing can also be observed in non-human animals (Clutton-Brock and Parker 1995; Ratnieks and Wenseleers 2007), there are still no reliable indications of internalized norms and rules – with the possible exception of chimpanzees (De Waal 2006). Conscience, which engages in self-monitoring as an internal policeman and which evokes shame and guilt with moral transgressions, appears to be a new evolutionary trait, even though the adaptive problem of finding the best possible balance between selfish and altruistic tendencies is as old as social life itself. Although the production of moral behavior is consequentially oriented to a large extent – completely in line with Alexander’s/Boehm’s point of view – conscience, however, is characterized by its deontological power. The moral judgment generated by conscience is precisely not what one would expect in a consequential cosmos, namely a strategically clever advisor with flexible instructions on how to best achieve one’s personal goals. In contrast, a moral judgment is impartial and non-consequentialistic. An evolutionary understanding of morality must, therefore, remain deficient, as long as the specific contribution of conscience in producing altruistic behavior is not understood.

This diagnosis points to an interesting state of affairs. A moral judgment guided by conscience and the emotions producing morality do not necessarily concur. Do humans have two moral systems perhaps? One would be evolutionarily ancient and people are barely aware of it; it operates intuitively and efficiently in the self-interested regulation of altruistic tendencies. Typical scenarios for this would be cooperative reciprocity, nepotism, and altruism as costly signal. The second system would be evolutionarily more modern, deontological, and more present in one’s awareness. This system would spawn a morality which necessarily requires a conscience.
This finding points to a significant gap in the understanding of the evolution of morality (DeScioli and Kurzban 2009; Kurzban et al. 2012). To date, there is no convincing theory that could explain how a moral judgment (or an ex-ante commitment) can exist in a Darwinian world, which pays more attention per se to a rule than to the consequences of such a rule. This theoretically unclear situation feeds the surmise that—as suggested in the foregoing—two different moral systems are intertwined in human moral practice. One of these two systems operates consequentially, completely as if the likely inclusive-fitness consequences decide on the probability of altruism. This system does not represent an insoluble problem with regard to an evolutionary explanation. It is the system which we find in the animal kingdom and which, moreover, constitutes a large portion of human moral practice. The second moral system is organized non-consequentially; its “director” is the conscience, the function of which—as it appears—has not yet been captured satisfactorily by the Darwinian theory of evolution.

In the following, I shall refer to the bundle of hypotheses which sensu Alexander, Boehm and others regard conscience as being personally beneficial, because it guides its owner in how to personally optimize the balance between selfishness and altruism, as the “navigator theory” of the evolution of conscience. Further below I shall contrast this with a “helper theory” of the evolution of conscience and attempt to defend this.

**Cui bono?**

Because its default option is a bad conscience, significant disadvantages are associated with a conscience which more or less lastingly depresses the quality of life as a perception of guilt and shame. Persons with a strong, persistent pressure of conscience not infrequently suffer from self-esteem problems which can lead to suicidal inclinations if taken to an extreme. Conscience can be lethal. It can also lead to a ritualized behavioral rigidity as is observed in the context of mental obsessiveness, i.e., in the context of a pathological condition (DeScioli and Kurzban 2009). In other words, conscience definitely does not only cause trivial costs.

In an evolutionary analysis, it is, however, not the consequences of behavior for life quality that stands in the foreground, but its consequences for the solution to the fundamental problems of life, namely maintenance and reproduction, i.e., ultimately the consequences for genetic fitness. With regard to this criterion as well, it appears to be extremely doubtful to attribute a function that essentially promotes fitness to a non-consequentialistic conscience. These doubts are nurtured for theoretical reasons, because non-consequentialism certainly undermines the opportunistic functional logic of natural selection. Empirically, this doubt is nurtured because behavioral rigidity guided by conscience is often associated with damaging, sometimes disastrous personal consequences. Ample anecdotal evidence (e.g. Atran 2012) supports the view that altruistic self-sacrifice can hardly be explained as an outcome of inclusive fitness-maximizing strategies. Saints and heroes, on average, pay too high a price for their genes to be counted among the winning models in Darwinian competition. Yet how can the evolution of
The evolution of conscience

conscience be envisaged differently, as a result of a selection maximizing personal fitness?

There is sociobiological theory that is capable prima facie of explaining an altruism that does not promote personal fitness but is still evolutionarily stable, namely the concept of the parent-offspring conflict presented by Trivers (1974). The theory according to which the evolution of conscience possibly stems from the dynamics of parent-offspring conflict was already proposed by Voland and Voland (1995). Since then, our anthropological knowledge has noticeably expanded in some of the domains relevant for this theory, so that it appears to be worthwhile to subject the early considerations to an updating revision and refinement.

According to Trivers’ (1974) genetic model of parent-offspring conflict, parents should expect more altruism from their children, for “selfish-gene” reasons, than the offspring would be willing to provide on their own. Although the reproductive interests of parents and their offspring are identical for the most part, because after all, with sexual reproduction, 50% of their alleles are identical by parentage, yet for the same reason, the interests of parents and their offspring do not completely match. Parents and their offspring will develop basically different opinions about appropriate behavioral tendencies; especially in as far as the degree of intrafamilial altruism is concerned.

**Situational conflicts about parental investment: the weaning conflict, for example**

The behavioral manifestations of the parent-offspring conflict have experienced an evolution which is characterized by an increasing significance of the consequences for the lives of offspring. In the simplest case of the parent-offspring conflict, the diverging interests of both sides clash, and an “open exchange of blows” decides who the winner and who the loser is. The weaning conflict is a typical example here. Female mammals will nurse their offspring for as long as the cost-benefit ratio does not become negative, i.e., it requires more losses in reproductive potential than it leads to fitness. If this occurs as the infant grows older, it is to be expected that a mother will end lactation and instead commence with the investment into another offspring. On the other hand, the infant will continue to insist on being nursed. Only when the costs for this are twice as high as the benefit, is it to be expected that the infant will stop pressing its claims. If this infant is still demanding milk and thus preventing the production of a full sib, it would then be impeding its own “selfish gene” interests. After all, a full sib receives half of the infant’s own genetic endowment on average, due to the common parentage, and thus contributes to the spread of this genotype. Therefore, preventing one’s mother from having another baby for purely blind selfishness would not promote inclusive fitness. In other words, there is a limited transitional phase, during which the interests of the mother and her offspring diverge with regard to the weaning process. This is the timeframe for the typical weaning conflict, which can also cause dramas in human families as well (Blurton Jones and Da Costa 1987; Fouts et al. 2005). In such conflicts, mothers do not develop any further interest in a lasting manipulation of their offspring. The conflict is solved by an ad-hoc regulation.
The temporary helper conflict: conflicts over cooperation within the kin group

In recent years, our ideas concerning the evolution of human familial systems have clearly expanded. Whereas the role of fathers as paternal investors, and thus the functional significance of pair-bonds, has been relativized and tendentially devalued, other family members, such as maternal grandmothers, maternal sisters and older children of a family group have increasingly become the focus of attention on the part of evolutionary anthropologists when studying reproductive systems. Human are now characterized as “cooperative breeders” (Hrdy 2005, 2009) which means that the work of raising offspring is not solely the mother’s responsibility – as is the case in the other apes. She can delegate some of the tasks to helpers or “alloparents”, as they are also called. Reproduction thus becomes the collective enterprise of a kin group.

Hrdy (2009) sees in cooperative breeding the evolutionary engine for developing the typically human cognition with the ability to empathize as its very special trait, something that is presumably unparalleled in the animal kingdom. Cooperative breeding is also likely to have provided fertile soil for the evolution of human morality – at least that is the thesis of this paper – because with the implementation of cooperative breeding, battlegrounds come into being for new intrafamilial conflicts related to the production and utilization of kin support. Within kin groups, unselfishness is expected and eventually asserted; after all, these are primary concerns of morality. Such a conflict occurs, e.g., if under certain socio-ecological conditions, parents can increase their lifetime reproductive success if they are able to motivate some of their offspring to assist with their parents’ reproductive activity, instead of dismissing them to their own reproductive careers.

The presence of offspring represents an energetic cost, rather than a benefit, to parents and in hunter gatherers are a net draw on family resources up to their mid-teens at a minimum (Kaplan et al. 2000). As a result, parents can be expected to have evolved an interest in cutting their costs by requesting their offspring to contribute to the family economy. This offers a battleground for what one could call the “helper conflict”. Parental interest in obtaining support from their offspring clashes with the latter’s interests in their own affairs.

The helper conflict is more basic and more sustained than the weaning conflict, because parents can not win it spontaneously, but only through a systematic and enduring redirection of the offspring’s own interests to a kin interest. In particular, this is achieved by manipulating the life contexts of their offspring. When parents influence the pay-off matrix of the strategies of their offspring, they can eventually ensure that helping appears to be the best of all possible available strategies for their offspring. This is why in many cooperative breeding species the helper conflict is settled in an overtly costly manner, e.g., through egg destruction, nest destruction, or infanticide (Cant 2012).

Help from offspring (and other kin) and the consequences of this for parental reproduction have also been investigated for human societies with varying cultural backgrounds (see Kramer 2010; Mace and Sear 2005; Sear and Mace 2008). For a rural population in Grande Anse, Trinidad, for example, Flinn (1989) was able
to show how the co-existence of fecund mothers and daughters in one household led to the improved reproduction of the older generation on the one hand and on the other, to the delayed reproduction of the younger, helping generation. This competition was essentially regulated by the antagonistic interaction between mother and daughter leading to reproductive suppression on part of the daughter. Thus, monogyny, i.e., the reproduction of only one women in Trinidadian households seems to be the outcome of what is described here as the “temporary helper conflict”.

With the helper conflict, the parent-offspring conflict experienced an expansion. Now the conflict is no longer about merely differences in interests with regard to the amount and the duration of parental investment, but acquires a pronounced social dimension because this means a takeover of different roles in the social network of the kin group. The parent-offspring conflict has changed from being an investment conflict, as it is typically described (Haig 1999; McDade 2001; Michalski and Euler 2007; Parker et al. 2002; Povinelli et al. 2005; Salmon 2007; Schlomer et al. 2011), and becomes a role conflict.

The lifelong conflict: conflicts over reproductive pathways

The parent-offspring conflict experiences a further escalation when lasting, even lifelong, altruistic performance is demanded of the offspring, something that had already been referred to by Trivers (1974:249). “Conflict in some species, including the human species, is expected to extend to the adult reproductive role of the offspring: under certain conditions parents are expected to attempt to mold an offspring, against its better interests, into a permanent non-reproductive.” Of course, the investment conflict and the conflict over reproductive pathways can be subtly interwoven with one another, because the “parent-offspring conflict over early PI [parental investment] may itself involve parent-offspring conflict over the eventual reproductive role of the offspring” (Trivers 1974:262).

A well-known life-determining family conflict is inherent in pre-modern farming societies, namely the conflict related to the issue of which offspring will inherit the farm and which offspring will have to waive their claims. In terms of behavioral ecology, this is an issue of who will receive the benefit of a more or less secure and productive breeding ground, by virtue of which, this person can marry and have children, and who instead will either be forced to a accept a decline in social status, emigration, or permanent celibacy (cf. Beise and Voiland 2008; Strassman and Clarke 1998). For this conflict to secure the patriline in European social history, there is also its logical counterpart, namely the conflict over the takeover of various familial roles to strengthen the matriline (e.g., Taylor 2004 for the Thai Khon Muang). Such lasting parent-offspring conflicts which characterize the whole biography of the parties involved can no longer be won by the parents through the direct suppression of childish selfish impulses. A lasting assertion of parental interests can only be achieved by subtly manipulating the offspring, at the end of which, a lifelong psychic fixation of the reproductive roles with typical behavioral tendencies is generated.

Trivers (1974:260) clearly recognized that his model had to lead to a completely new assessment of the socialization
process: “One is not permitted to assume that parents who attempt to impart such virtues as responsibility, decency, honesty, trustworthiness, generosity, and self-denial are merely providing the offspring with useful information on appropriate behavior in the local culture, for all such virtues are likely to affect the amount of altruistic and egoistic behavior impinging on the parent’s kin, and parent and offspring are expected to view such behavior differently”. According to the hypothesis of this paper, the agency in which conflicts of the type discussed here are negotiated is a child’s conscience. As a result, the authorship for moral judgments generated by conscience lies not solely with the offspring, but they are also an expression of parental (and possibly alloparental) interests that have been asserted during the course of the parent-offspring conflict. Thus the functions of conscience precisely do not, as assumed by Alexander (1987), Boehm (2012) and others, guide the bearers of a conscience through the snares of their culture in their own interests. Instead, there is much in favor of the idea that the function of conscience, with its unmerciful, non-consequentialistic rigidity, works in the interests of those persons who are able to fill the conscience with specific norms and attitudes. These are the persons of the cooperative breeding group, above all the parents.

The “slave scenario” of the parent-offspring conflict

Depending on how reproductive success is contingent upon economic success, it may be worthwhile to raise offspring to become more or less hard-working family members, instead of preparing them for their own reproduction. An example of what this conflict may have looked like in the lives of families is illustrated by a study of a Swabian rural community (Germany) in the 18th and 19th centuries. It is stated there that:

“In many families, children were systematically kept uneducated or even crippled. This not only eliminated the competition, but also tied a daughter to the house, for example, so that she could look after the aged members of her family. An uneducated daughter was also able to do the housework and did not make any demands. Such simple-minded persons were found in many families. The technique of producing them is not found in the historic sources, but the fact that this process did not function as naturally as often appeared to be the case, is something that people remained aware of; they don’t say that ‘such and such is feeble-minded’, but there is always talk of ‘having been made dim-witted’ or more clearly, ‘as being held as a dim-wit’” (Illein and Jeggle 1978:76, my translation).

The fact that the strategic production of helpers is part of the repertoire of parental strategies is also obvious in the statistics on differential infant and child mortality. For instance, data from rural Punjab (India) show that the traditional sex bias in this region shows itself in the survival chances in favor of boys only in the second and following birth orders. Here the girls in the sample studied were exposed to the risk of overmortality of 13% to 15%. Among firstborns, on the
The evolution of conscience

other hand, relatively more boys died than girls. Their overmortality amounted to 32% and was thus significantly larger than would have been expected physiologically (Das Gupta 1987). Very obviously, the neglect of daughters is selective. The first daughter played a special role in the family formation strategies. There are plenty of indications that this was the “helper role”, completely in the sense of the “slave scenario” discussed here. The strategic production of helpers is also especially discernible in the Roma living in Hungary. Here too, there is increased investment in firstborn daughters, from whom it is traditionally expected that they will assume an intrafamilial helper role (Bereczkei and Dunbar 2002).

The “slave scenario” has a second variant. One finds, of course, celibate life histories which are not primarily compelled by the exploitation of productive work performance, but are the result of wealth concentration strategies. In aristocratic and upper-class families, the preservation of dynastic interests not infrequently compelled the neutralization of the reproductive potential of some of one’s descendants, mostly daughters. They did not necessarily have to directly take over functions as helpers, but merely had to withdraw to cloisters and convents. This happened ostensibly to achieve the preservation of family property (Qirko 2002). Under the conditions of an extreme displacement competition, such as prevailed under the dominant and influential dynasties of Europe, the concentration of property can much more efficiently contribute to genetic persistence than the exploitation of all of the generative potentials of a family. If an undivided estate formed the basis of a family’s power and competitiveness, as was the case among the nobility, then the preservation of this estate required the uneven allocation of life opportunities to one’s offspring. Any payable inheritance and dowry payments that would automatically decrease the family assets would be able to be averted in this manner. European convents and monasteries were not infrequently supplied with new members from these laterborn heirs and heiresses, from whom lifelong celibacy and childlessness were required in the interests of a family solidarity aimed at wealth concentration (Hill 1999). According to all of the models of behavioral ecology, breeding is always better than helping under unconstrained conditions, so that the reproductive neutralization of the daughters from the “better families” could not have lain in their own fitness interests. Such situations perfectly reflect the parent-offspring conflict, and the fact that so many heirs ultimately submitted to parental authority documents the fact that this conflict was very frequently won by the parents.

The “tax scenario” of the parent-offspring conflict

The basic idea of a “tax scenario” for the parent-offspring conflict is that adaptively coping with life not only depends on the immediate economic circumstances to which one’s offspring can contribute and who are therefore prone to be exploited by their parents as described in the “slave scenario”. In addition, lifetime reproductive success also depends on the infrastructure of the social community in which the parents live. Common goods and collective actions might be crucial for maintenance and reproduction. It may be worthwhile for families to invest in public goods, if they are thus able to improve their own opportunities in life. It may be worthwhile to sacrifice a certain part of
current fitness in favor of a longer-term interest in common goods, i.e. to pay taxes so to speak. The currency of this tax is the altruism of one’s offspring. Under certain conditions, such a strategy can be fitness-maximizing for parents, namely when the costs incurred by the altruism of their offspring are more than compensated for by improved opportunities for survival and reproduction. As this happens at the expense of the children, a parent-offspring conflict arises and conscience has evolved to regulate this conflict. What shape this takes surely depends to a very high degree on the socio-ecological conditions under which the population lives, and depends on which public goods are important locally.

On the evidence of human history, which is described by Alexander (1987) as a history of constant competition among autonomous kin groups, the ability to put up a fight is likely, for example, to be a public good which has consistently decided the fate of human communities; the investment into this ability is likely to have been of great interest to the families affected. However, in the earlier milieus of human history, paying taxes for the ability to fight meant producing warriors who were willing to assume the role of combatants. This scenario does not require that the warrior’s fitness outcome is positive on average. Warriors who would also give their lives to secure the existence and the survival chances of their community factually number among the evolutionary losers (even if their heroism were to be honored in society). Producing warriors, however, might be an expensive but unavoidable measure for parents to preserve opportunities allowing them to further participate in the evolutionary game. Even if under historical conditions a child was certainly difficult to replace, one’s own life, and the genetic persistence of one’s own lineage could never be replaced. It can only be indicated here that modern phenomena, such as the existence of suicide bombers, could also be explained using this logic (Voland and Voland 2014).

Of course, the warrior role is an extreme example, because the costs and benefits of the life and death of one’s offspring are offset. In everyday social practice, there are clearly even less spectacular manifestations of helper behavior, which are to be interpreted as an expression of the “tax scenario”. Whether participating in maintaining irrigation systems, avalanche protection measures, or constructing bridges over the nearby creek – the actors involved invest in the local infrastructure and therefore exhibit behavior which actually should not exist, according to the logic of the N-persons prisoner’s dilemma. That such behavior does exist and that it is evolutionarily stable, is an indication for the fact that collective actions are not necessarily engaged in for personal advantages (even if they do occasionally arise: Rusch, in press), but for third-party advantages. This interest in investments in public goods lies with the cooperative breeding group. Paying “taxes” in form of altruistic contributions increases the probability of persistence in view of the hazards of life; therefore, altruistic helpers are needed for this purpose.

Neither the “slave scenario” nor the “tax scenario” are to be confused with the “kin selection” scenario, as there is a grave difference here: kin-selected altruists maximize their own inclusive fitness, i.e., they are acting in his or her own evolved interest. On the other hand, altruists who have evolved in accordance with the “slave” or “tax” scenario and
The evolution of conscience

263

who are guided by their conscience runs the risk of becoming the losers of the evolutionary game since their behavior is strongly conditioned by interests that are not intrinsic to themselves. This is why very special barriers have to be overcome to assert these demands for altruism and for that purpose the evolutionarily novel institution of conscience takes over its adaptive role. One could call the outcome of the helper conflict “moral altruism”. This comes about when parents subordinate their offspring to their dynastic interests.

The perspective of the offspring or the improbability of disobedience

The scenarios discussed here imply that offspring are under selection to ignore their parents’ manipulative attempts. They should stave off the “selfish-gene” interests of their parents and not allow the “inner policeman”, as an agency of parental dominance, to take over, in order to gain the maximum degree of autonomy when pursuing their own “selfish-gene” oriented life strategies. It seems, however, that this kind of counter-selection is not very impressive. With respect to the evolution of conscience at least, it seems that Alexander’s prognosis (1974) applies, whereupon parents are likely to win the conflict with their offspring. Why is this so? I would like to put forward two aspects for discussion, the interaction of which might provide the evolutionary reasons for why the parent-offspring conflict was won by the parents in the hominine line, and why a conscience was established as a species-specific adaptation of Homo sapiens despite its inherent costs.

Costs of docility

Children are exposed to a fateful dilemma. They come into the world as dependents and have to rely on their caregivers. Only during the course of their socialization do they acquire the necessary and culturally-specific knowledge in order to be able to implement and follow the local rules for life. Children need both specific models of learning and support as well, and in view of the partially genetic identity of parents and their children, parents appear to be better at this than all of the other members of the community. This is why children are selected for being able to trust their parents more and for learning more from them than from all others, at least during their early years.

Docility is the adaptive strategy with which naïve individuals are able to learn the culturally relevant modes of behavior, attitudes, norms, etc., without, however, being themselves able to evaluate the impact of this cultural input on their own fitness. Clearly, what is described here is the child’s world of bounded rationality. Simon (1990) has been able to demonstrate very convincingly that altruism is the price for the personal benefits of docility: “Because docility – receptivity to social influence – contributes greatly to fitness in the human species, it will be positively selected. As a consequence, society can impose a ‘tax’ on the gross benefits gained by individuals from docility by inducing docile individuals to engage in altruistic behaviors. Limits on rationality in the face of environmental complexity prevent the individual from avoiding this ‘tax’” (Simon 1990:1665). For Simon, society is the tax collector, whereas in the argument of this paper, the tax collectors are the parents and a conscience, with its non-consequentialistic mode of function-
ing, is the tax notice. Interestingly, the word “conscience” does not even appear in Simon’s article (1990), even though it was presumably implicitly thought of, because one page later we read: “Guilt and shame, although perhaps genetically independent of docility, also serve most people as strong motivators for accepting social norms. Guilt is particularly important because it can operate independently of the detection of nonconformity” (p. 1667), and further “Moreover, guilt and shame will tend to enforce even behavior that is perceived as altruistic. Hence the docile individual will necessarily also incur the cost, c, of altruism” (p. 1667). What Simon is referring to here is the functions of conscience. That is why I consider it to be justified to transpose his considerations from the context of society as a whole to the context of the family, without this coinciding with a significant loss in meaning. This is not to deny, however, that “society as a whole” benefits from the performance of conscience in its members – but that is the result of the parent-offspring conflict, not the direct cause of self-interested altruistic tendencies on the part of the offspring.

At the same time, Simon (1990:1667) refers en passant to an interesting ancillary aspect of his considerations, which gain more importance in the context of our argument on the evolutionary possibility of non-consequentialistic moral judgments than was attributed to this by Simon himself in only two sentences: “Moreover, much of the value of docility to the individual is lost if great effort is expended evaluating each bit of social influence before accepting it. Acceptance without full evaluation is an integral part of the docility mechanism and of the mechanisms of guilt and shame”. Non-consequentialism, as indicated here, can not be kept at bay by the children as an expensive, but unavoidable by-product of the otherwise functional adaptation of the children’s docility. Owing to their bounded rationality children should be especially motivated to rely fully on their primary caregivers.

Empirical findings support this surmise. Up to a certain age, children are fairly uncritical about what they are being offered to learn. Harris and Corriveau (2011) report on various experiments conducted in various labs which demonstrate with what ease children are willing to abandon correct, but self-generated knowledge, if adults make different claims. This holds even if all of the obvious evidence concerning, for example, the location of an item, very clearly shows the inaccuracy of the claims made by adults. Children of that young age simply are not competent enough to verify the learnings that they are being offered.

If, however, the possibility for a critical evaluation of a teaching is missing, then the second-best strategy is implemented, namely selective trust – however, selective trust only for those persons for whom at least a partial overlap of evolved interests can be assumed; and these are the members of the kin group, especially mothers. Selective trust is generated by mental mechanisms in which the history of the interactions between the children and the adults in their environment play a crucial role and are reflected in the development of attachment. Empirical research shows that children are especially receptive to information that comes from informants to whom there is a secure or an ambivalent attachment. On the other hand, if the relationship has to be classified as “avoidant”, then mothers do not enjoy any bonus of trust compared
to outside learning models (Harris and Corriveau 2011). Attachment research interprets this type of attachment as the outcome of the constant experience of the children, namely that their wishes are regularly and predictably rejected and they can not expect any support. Under these circumstances, trust would be highly risky.

Cooperative breeding increases a child’s risks for underinvestment

In the hominine line, it was ultimately the evermore costly reproductive effort which the children demanded from conception to social independence which set a significant fitness limit for mothers. With the evolutionary implementation of cooperative breeding, and thus the possibility of more or less delegating the reproductive effort to alloparents, the mothers’ burden would be decreased in this respect. The resources which thus became available were utilized to increase maternal fertility which on average also resulted in an increase in the mothers’ lifetime reproductive success, and therefore, their genetic fitness. Even for grandmothers, especially for maternal grandmothers, the opportunities for maximizing fitness increase with cooperative breeding. The post-generative lifespan is filled with kin support strategies instead of being reproductively wasted, whereby their share of indirect fitness in total fitness can be increased.

The parental and the grandparental generations very obviously derive substantial gains from cooperative breeding (Voland et al. 2005). What about the outcome for the children? They also derive benefits from the implementation of cooperative breeding, because having multiple caregivers around means some kinds of risks are minimized. Fluctuating supply situations can be compensated for by kin support, and if dire need were to arise because the mother was no longer available as an investor having alloparents could save lives. However, it should not be overlooked that these indisputable benefits are associated with costs. Due to the reduction of the interbirth intervals, sibling rivalry occurs even during the infantile or juvenile phase. Sibling rivalry – typically within a litter or a clutch – is widespread in the animal kingdom and has more or less harmful, in extreme cases, even lethal consequences for some sibs (Mock 2004; Roulin and Dreiss 2012). By reducing litter size in the evolution of primates, the sibling conflict was first eased, but due to cooperative breeding it was re-ignited, because children and juveniles live together in families and have to share the same resources. This leads to scrambling competition, which can be more or less strong depending on the resources available. Primarily under precarious resource situations, and in particular if this leads to resource depletion, sibling rivalry can coincide with disadvantageous consequences for fitness. Thus Hagen et al. (2006) were able to show for the Shuar (horticulturists and hunters in Ecuador) that the more children that lived in one household, the more retarded these children were with regard to their height and their weight. Lawson and Mace (2008) reported similar findings on the basis of a large-scale longitudinal study on the development of British children. Under otherwise equal conditions, ten-year-olds with four or more siblings were more than 3 cm shorter – on average – than single children of the same age. Sibs clearly are both rivals and resources for one another (Nitsch et al. 2012; Pollet
and Hoben 2011), and on the intrafa-
milial battleground, a cascade of com-
plex dynamics decides what trade-off the
evolutionary ambivalence of sib relation-
ships will bring.

Costs not only occur through the in-
crease in sibling rivalry but also due to
the fact that mothers, grandmothers and
all other alloparents, as a rule, have in-
vestment alternatives (Hawkes 2012).
Due to the cooperative networking on
the one hand and the more or less man-
ifest in-law conflicts within the cooper-
ative-breeding community on the other,
alloparents gain in strategic opportuni-
ties for differential investment. They can
decide where they want to invest and to
whom they wish to delegate reproductive
effort. Consequently, this leads to alar-
er variance among the offspring with re-
gard to the question of how much invest-
ment they can expect. Some may receive
above-average care, because they stand in
the center of the reproductive interests
of several caregivers, others may tend to
receive not enough care, because they do
not have features that potential investors
might be interested in.

For the offspring, this means that
they must offer themselves in special
ways as worthwhile objects for invest-
ment by their kin group. More than un-
der the conventional regime of solitary
breeding, they should be able to win over
caregivers for themselves. They will be
even more dependent on the approval of
their social environment than is already
the case with solitary breeding. Parental
nervous systems can be influenced in
one’s own interests, and consequently,
natural selection has shaped offspring
so that they will make use of this op-
tion. The evolution of begging behavior,
which shows honest or dishonest need-
iness (Qvarnström et al. 2007; Royle et
al. 2002); the temper tantrums of human
and chimpanzee children; regression
after the birth of a sibling (Dunn and
Kenrick 1982); postnatally opened, but
unseeing, eyes; non-social smiles (Hrdy
1999); and finally, infant cries (Soltis
2004) – all of these behaviors can be in-
terpreted as childish attempts to exert
beneficial influence on parental and al-
oparental investment decisions.

According to sociobiology, this be-
havior is evolved, as it puts newborns
in a situation where they can win over
their mothers for themselves; because
from an evolutionary point of view, there
must have been adaptive consequences
for newborns as a result of the fact that
post-natal infanticide was a pragmatic re-
productive strategy option for parents in
all phases of human history. All histori-
cal societies (and those who see abortion
as a form of infanticide, may confident-
ly count modern societies among them)
made use of this option in a socially ac-
cepted manner – partly with double-digit
rates of infanticide (Hrdy 1999).

Approval by the mothers and the
kin group as a whole can be achieved in
particular through the takeover of help-
er functions, because it is easy to see
that under otherwise equal conditions,
parents will prefer those children who
allow the recognition of their suitabili-
ty as helpers. In the net balance sheet,
the rearing of these children is cheaper,
because they will repay some of the in-
vestments made in them through their
kin support. Indeed, there are studies
on differential postnatal parental invest-
ment which allow the interpretation to
be made that humans take into account
a possible later repayment by their off-
spring when allocating their reproduc-
tive investment (e.g., Bereczkei and
Dunbar 2002 for gypsies in Hungary and
Das Gupta 1987, for a population in rural Punjab, see above).

In view of the parent-offspring conflict, offspring should, in their own interests, correspond to parental expectations of their helper role. Through subordination, or to put it differently, through an “evolutionary capitulation” in the parent-offspring conflict, children increase their survival chances – but for significant price, namely the price of obedience to altruistic demands made by their parents.

**Conclusion and summary**

Early childhood cognitive constraints and evolved parental reproductive interests meet in the evolution of conscience. The morality generating nature of this constellation only becomes evolutionarily manifest under the conditions of a cooperative breeding system in which the genetic parent-offspring conflict escalates to eventually lifelong helper conflicts over the production and consumption of altruism. This is the core of the helper theory of the evolution of conscience. It closely links the concept of morality with parental exploitation and manipulation.

Now that the arguments for the helper theory of the evolution of conscience have been unfolded, the question arises of which of the competing theories, the helper theory or the navigator theory, fit the nature of conscience better. To find an answer to this question, we now resume the characteristics of consciences as listed in section 2.1.

Asymmetric Evaluation: A punishing conscience is not a guide, but is a hindrance. However, one would expect that a navigator system would also emotionally reward the right decisions. A conscience does not do this and also does not need to do this, because this task is assumed by the strategic emotions (Nesse 1990). Furthermore, the navigator theory can not explain why the perception of guilt coincides with feelings of inferiority. If the perception of guilt is intended to serve the mere avoidance of errors, it does not require any reduction in self-esteem. On the other hand, a reduced self-esteem tends to be advantageous for the helper role, because it contributes to the subordination of one’s own needs and interests to those of others. To put it bluntly, under the helper theory it should be expected that a strong conscience will tend to correlate with depression; while under the navigator theory, it tends to correlate with the healthy self-confidence of someone who is taking the correct actions. There is empirical evidence for the former (Zahn-Waxler and Van Hulle 2012).

Second perspective: Does not initially make a distinction between these two theories, even if both make varying assumptions concerning the origin of the “second perspective”; namely of “family” vs. “society”. Psychological research documents the early influence of social proximity on the formation of a conscience, whereas the influence of society, if any, only appears much later ontogenetically. Therefore, this criterion fits the helper theory better.

Domain unspecificity: Both theories make different predictions here. Whereas the navigator theory only expects moral judgments in cooperation issues, the helper theory is broader, by expecting moral judgments in all domains affecting parental concerns, such as the allocation of the resources acquired by the family, or dealing with family authority and loyalty. Both are aspects of moral behavior which do not generate long-term gains
in cooperation within society. Thus, this criterion tends to speak in favor of the helper theory.

Nonconformity: Sharply contradicts the navigator theory. A conscience does not unconditionally motivate persons to social conformity. On the contrary, it can evoke social rebellion. There must, therefore, be a surplus function which goes beyond securing social conformity. This surplus function is reflected by the helper theory with its primacy of parental interests over social functioning.

Motivational pressure: Even from the standpoint of the navigator theory, motivational pressure by the conscience is to be expected. However, a navigator system should habituate to the circumstances. After all, life goes on. According to the helper theory, it is precisely the behavior motivated by a conscience that counts – not the lifetime perspective. The neutralization of motivational pressure through habituation would undermine the effectiveness of a conscience according to the helper theory but not to the navigator theory. Thus, this aspect speaks in favor of the helper theory.

Non-cognitivism: Does not discriminate between the two theories under discussion.

Non-consequentialism: A navigator system is as good as the results that it produces and will be selected accordingly. Non-consequentialism is, therefore, fatal for a theory in which genetic self-interest is the explanatory ultima ratio. On the other hand, a consequentialistic conscience would be fatal for the helper theory.

Egocentricity: This aspect is a point in favor of the helper theory, which attempts to explain the altruistic demands made by single individuals of themselves; and not just social competency, as the navigator theory does. If merely social competency were required, then it should not make any big difference whether moral judgements aim at actions and omission by oneself or by others.

I am not going to contest the fact that the comparison undertaken here definitely still requires differentiating empirical research; nevertheless, a rather clear result emerges on the basis of the findings compiled here. All of the eight examined features of conscience are compatible with the helper theory but only one is unconstrainedly compatible with the navigator theory. On the other hand, seven features balk at a straightforward interpretation by the navigator theory, whereby a few features generate definite contradictions, the theoretical resolution of which is likely to prove difficult in line with the navigator theory. Therefore, I see the helper theory of the evolution of conscience as providing a more powerful explanation than the navigator theory does. The non-consequentialistic judgment of conscience, and thus human morality in a narrower sense, occurred evolutionarily not during the course of an increase in social complexity, but during the transition of hominines to cooperative breeding and the resulting helper conflict, which was new in hominine evolution. This is why Kurland and Gaulin (2005:453) could be right when they point out (with a nod to Sigmund Freud) that “some humanists have found in our peculiarly intensive family ecology the source of all neurosis, psychosis, and the world’s troubles”. Anyway, “there are many implications of Trivers’ (1974) family-conflict model for psychiatry, socialization theory, and the politics of the family” and, I would like to add, for the evolution of morality and conscientiousness as well.
Conflict of interest

The author has declared that no competing interest exists.

Acknowledgments

This paper was written while the author was a senior fellow at Alfried-Krupp-Wissenschaftskolleg (Institute for Advanced Studies), Greifswald (Germany).

Corresponding author

Eckart Voland, Institut für Philosophie, Universität Giessen, Rathenaustrasse 8, D-35394 Giessen, Germany
e-mail address: eckart.voland@phil.uni-giessen.de

References


Nitsch A, Faurie C, Lummaa V. 2013. Are elder siblings helpers or competitors? Antagonistic fitness effects of sibling in-
The evolution of conscience

271

Parker GA, Royle NJ, Hartley IR. 2002. Intra-
Pollet TV, Hoben AD. 2011. An evolution-
Royle NJ, Hartley IR, Parker GA. 2002. Beg-
Schlomer GL, Del Giudice M, Ellis BJ. 2011. Parent-offspring conflict theory: An evo-
Simon HA. 1990. A mechanism for social sele-
Tyson P, Tyson RL. 1993. Psychoanalytic the-
ories of development – An Integration. New Haven CT: Yale University Press.
Voland E, Chasiotis A, Schiefenhövel W editors. 2005. Grandmotherhood – the evolu-
tionary significance of the second half of female life. New Brunswick NJ and Lon-
don Rutgers University Press.
Zahn-Waxler C, Van Hulle, C. 2012. Empa-