Two Levels of Selection

Some unpleasant emotions, like fear and disgust, appear straightforwardly susceptible to evolutionary explanation on account of the benefits they seem to provide to individuals. But guilt is more puzzling in this respect. Like other unpleasant emotions, guilt is often associated with a host of negative effects on the individual, such as psychological suffering and social withdrawal (Harder 1995; Luyten et al. 2002). Moreover, many guilt-induced behaviors, such as revealing one’s offenses and placing oneself before the mercy of others, could levy a cost to individuals that is not outweighed by guilt’s benefits. Supposing there is an evolutionary story to tell about the origins of guilt, the question is how such negative effects were sufficiently outweighed by the potential fitness payoffs that guilt might have yielded to individuals. In this article, we consider which forms of evolution could have resulted in guilt, and whether current evidence can tell us which form of evolution most likely occurred.

There are two forms of evolutionary selection on which we might draw when explaining guilt, individual selection and group selection. Individual selection occurs when individuals in a population vary from one another with respect to some trait, and some of this variation is fitness-
affecting and heritable. The traits in this case are individual-level traits, and the selection occurs within groups or populations of individuals, leading to increased frequency of some traits and a decrease of others within the populations. Much of evolution occurs in this way.

But there is also a form of evolution that operates at a group level. The idea here is that groups of individuals exhibit traits and can fare better or worse than other groups because of these traits. Although groups and their individual members can sometimes exhibit the same property (e.g., a group of wildebeest might be fast when all or most of its members are fast), groups frequently exhibit traits no individual organism can bear, such as sex ratio (the proportion of males and females). Things get interesting when a trait is selected against at the individual level, but selected for at the group level. Such a trait benefits the group and helps it prosper, even though it is not in the evolutionary interest of individuals to bear it.

One such trait with these characteristics may be biologically altruistic behavior. In this case, it might not on balance be biologically advantageous for individual organisms to exhibit biologically altruistic behavior, but it might be biologically advantageous for the group to be composed of altruistic individuals. If the group-level selection pressure is strong enough, the altruistic trait can spread throughout the species in spite of its individually-maladaptive character. This is the conclusion reached by Darwin in his (1871) *Descent of Man* when he argued that brave warriors could evolve because warring groups with brave warriors will conquer groups without—or with fewer—brave warriors, even if there is an individual-level fitness cost to being brave. Contemporary work on altruism draws similar conclusions with respect to strong group-level selection (Sober and Wilson 1998).

Group selection, however, has a fraught history, with some challenging the tenability of selection at a level higher than that of genes or individuals (see Okasha (2006) for a general overview of this debate). But the consensus position today among biologists and philosophers of biology is that group selection does occur; the main debate concerns how strong a force it is in nature. For our purposes here, we need only assume that group selection explanations can be tenable.
Returning to the subject of guilt, our main question here is whether guilt is likely to have been selected for at the individual level, at the group level, or both. Might it be individually advantageous to be guilt prone? Or is guilt proneness similar to a biologically altruistic trait in that it might be biologically disadvantageous for individuals, but good for the group?

One way to advance an answer to this question is through the use of evolutionary game theory. On the evolutionary game theoretical approach, the conditions and strategies for interaction among players are pre-specified as elements of a mathematical model. On the basis of this model, determinate conclusions can be reached with respect to the evolvability of a particular behavioral trait among the players. O’Connor (2016) and Rosenstock and O’Connor (this issue of Emotion Researcher) utilize this approach to determine under what specifiable conditions guilt could evolve at the individual level.

Our explanatory approach here, however, diverges from the game theoretical approach. Rather than specify the conditions under which guilt could evolve at the individual or group level, we begin by querying the clinical, empirical, and legal literatures on guilt in order to arrive an adequate understanding of guilt’s motivational profile and the role it plays in human interpersonal contexts. We then consider what can be inferred about the evolutionary function of guilt and whether guilt was favored at the individual level or the group level. Our approach should not be seen standing in opposition to the game theoretical approach; rather, each approach should be viewed as a potential complement to the other.

**The Nature of Guilt**

Subjects of clinical and empirical studies frequently identify the object of their guilt experiences as a specifiable set of past actions that deviated from or violated accepted social or moral standards, and for which they take responsibility (Barrett, 1995; Harmon-Jones, Amodio, & Zinner, 2007; Lindsay-Hartz, de Rivera, & Mascolo, 1995). The link between guilt experiences and the judgment that one is responsible for such transgressions underwrites the conception of guilt as primarily *action-focused*, that is to say, the object or focus of guilt tends to be particular actions or behaviors that an individual identifies as her/his own. This contrasts with the *self-focused* character of other painful emotions, such as shame and embarrassment, where the
emotion is directed toward the whole self or a negatively perceived aspect of the self (Barrett, 1995; Baumeister, Stillwell, & Heatherton, 1994; Tangney, Miller, Flicker, & Barrow, 1996).

Guilt thus centers on actions – manifesting as an unpleasant feeling associated with past wrongful actions or the contemplation of anticipated future wrongdoing. Because we wish to know whether guilt was favored by individual-level or group-level selection, we need to understand what the effects of guilt-induced behavior are on the individual who performs them. We can consider these effects within contemporary human social contexts, which could help us to understand why guilt evolved in humans.

Of course, the current adaptive effects of guilt will be imperfect indicators of guilt’s origin, but they can suggest hypotheses about past evolutionary pressures. It is also helpful to know whether guilt is more phylogenetically widespread, as this might provide us with some clues about how evolutionarily ancient guilt is. As we shall see, this may help us decide between individual and group level models of selection. Let us begin by considering some recent examples of individual-level selection models.

**Individual-Level Selection and Guilt**

Two approaches to developing an individual-level account of guilt have been especially popular. One account involves focusing on the way guilt functions as an emotionally painful check on motivations to violate normative standards. Joyce (2006) and James (2011) take this approach, developing what might be called ‘self-recrimination’ models of guilt. On their view, the experience of guilt functions as a sort of self-punishment for individuals who transgress social norms, and its anticipation can thus dampen motivations to defect or cheat on cooperative ventures governed by such norms. Guilt, then, reinforces cooperative tendencies in individuals, which, on Joyce’s and James’s view, explains why it was favored by selection at the individual level.

A second common approach to explaining guilt explores how guilt might have solved particular problems individuals faced in organizing and sustaining cooperative arrangements. Frank’s (1988) commitment model of the social emotions is perhaps the most familiar and influential of
such accounts. On Frank’s model, individual selection favored social emotions, including guilt, that enabled individuals to make credible commitments to one another in the organization and execution of cooperative ventures.

He suggests two pathways by which the social emotions would have evolved. First, those who experienced social emotions like guilt would have been disposed to maintain their commitments to cooperative ventures, and would thereby earn positive reputations for having such dispositions and for being reliable partners. Second, guilt and other social emotions are associated with hard-to-fake, largely involuntary facial and physiological expressions. These expressions serve as reliable indicators that an individual experiences social and moral emotions, from which others can infer that the individual is disposed to keep to the terms of cooperative agreements. Along either pathway, the community’s preference for individuals who are reliable cooperators and the community’s ability to discern who those cooperators are would have increased the selective pressure on the emotions that disposed individuals to cooperate.

Self-recrimination and commitment models of guilt highlight important roles guilt plays in the social life of individuals. However, these models leave unexplained important aspects of guilt and guilt-induced behavior. First, guilt is associated clinically and empirically with a number of psychologically maladaptive effects (Bybee & Quiles, 1998; Harder, 1995; Lindsay-Hartz, de Rivera, & Mascolo, 1995; Luyten, Fontaine, & Corveleyn, 2002), and it induces a number of behaviors that could be costly to the individual, including confession to violations of social norms, acceptance of external punishment, and self-penance (Katchadourian 2010; Radzik 2009). Moreover, there is evidence that psychopaths, who are thought either to have an attenuated capacity for guilt or to lack that capacity altogether, are more successful at obtaining conditional release from prison sentences (Porter et al., 2009), despite psychopathy being a reliable predictor of recidivism (Leistico, Salekin, DeCoste, & Rogers, 2008). This is likely due to the fact that they can be more strategic in their appeals because they do not have guilty feelings encroaching on their behavior (Porter et al., 2009). Thus, the putative cost of guilt on individuals may not be limited to the degree of emotional pain one experiences, but may also include behaviors that guilt prompts.
Second, the focus on guilt’s role as a painful psychological counterweight to motivations to defect on or cheat cooperative ventures does not explain why selective pressures targeted guilt proneness. Recent empirical research in developmental psychology and neuroscience suggest that guilt is costly in terms of the psychological hardware that underwrites it. Developmentally, guilt emerges late in children, concurrent with or after the emergence of capacities to take responsibility for particular actions, to evaluate those actions by normative standards, and to experience empathic concern for others (Ferguson, Stegge, & Damhuis 1997; Harris 1989; Kochanska et al. 2002; Lagattuta & Thompson 2007). Neuroimaging has shown that experiences of guilt are associated with reduced asymmetry between left- and right-cortical activity, which suggests that guilt experiences involve both withdrawal and approach orientations, motivating complex behavioral sequences involving self-regulation and reparation (Amodio, Devine, & Harmon-Jones 2007). This asymmetry does not appear to obtain during reported experiences of other negative, other-directed emotions (ibid.). While much more can be said with respect to the developmental and cognitive complexity of guilt – for further discussion, see Deem & Ramsey (2016), Section 3 – enough has been said to see that the self-recrimination and commitment models must answer the question of why selection might favor such complexity just to reinforce or augment other, more phylogenetically ancient emotions that already undergird prosocial behavior. Neither model provides a satisfactory answer to this question, leaving the evolutionary puzzle of guilt unsolved.

Guilt in the Legal Arena

Another area of research in which we can find clues for solving the evolutionary puzzle of guilt is the study of guilt in the legal arena. Although the modern legal system is, from an evolutionary time scale, a recent invention, norms and norm violations are much older. By seeing how guilt operates within the current legal milieu, we can cautiously extrapolate to what the consequences of guilt may have been during its evolutionary origin. This will serve as another source of evidence for whether we should take guilt to be good for individuals or bad at the individual level, but good for groups.

The term guilty has a dual meaning; it can refer to the fact of having committed a crime (one can be “guilty of a crime” without any negative feelings about it), or it can refer to the emotion (one
can feel guilty about a state of affairs). It is thus important not to conflate these two senses. To be as clear as possible, we will only use the word ‘guilt’ to refer to the emotion and will therefore use phrases like ‘committed the crime’ instead of ‘guilty of the crime’.

With mock trials, it is clear that displays of guilt help convince jurors that a defendant committed the crime (Jehle, Miller, & Kemmelmeier, 2009), especially if guilt is expressed shortly after the crime (Bornstein, Rung, & Miller, 2002). This shows the dangers of displaying guilt after suspected involvement in a crime.

Although displays of guilt can increase the chance of conviction, it generally has an opposite effect on sentencing (Gold & Weiner, 2000; Robinson, Smith-Lovin, & Tsoudis, 1994). Thus, once someone is convicted, if they have not displayed signs of guilt, they will generally receive a harsher punishment (Garvey, 1998). Such individuals are described as “cold-blooded” and are despised for their lack of emotion. This therefore poses a dilemma for those convicted: displaying guilt makes one more likely to be convicted, but if convicted without displaying guilt, a more severe punishment is likely.

Whether or not guilt is good for the convicted individual in such situations will therefore depend on the degree to which a display of guilt increases conviction probabilities and mitigates punishment. Furthermore, the weight of evidence will play a role in whether it is good or bad to exhibit guilt. If the defendant was caught red-handed, then guilt displays will do little to increase their chance of conviction and will generally be a good thing. But if there is little evidence, guilt will have a stronger negative effect on the accused.

In studies of actual court cases, displays of guilt are often linked to mitigated punishments. This is especially true of first-time offenders and those who committed less severe crimes (Harrel, 1981; Eisenberg, Garvey, & Wells, 1998). But why, we should ask, are we inclined to be lenient toward those who display guilt? One possible reason is that experiencing guilt is perceived by others as form of self-punishment, and the more that guilt is experienced, the less externally imposed punishment seems warranted. Another reason may be that those who experience intense guilt will not want to repeat the experience and will thus be less likely to recidivate. While it is
clear that guilt can serve as a form of punishment in its own right, is it the case that those who are more guilt prone are less likely to recidivate? Some studies show that genuine guilt feelings do indeed predict lower recidivation rates (Hosser, Windzio, & Greve, 2008).

More generally, guilt-prone individuals exhibit fewer offenses and less delinquent behavior (Cohen, Wolf, Panter, & Insko, 2011; Cohen, Panter, & Turan, 2012). They tend to be more prosocial, compliant, altruistic, and are more apt to perform reparative behaviors (Malti & Krettenauer, 2013; Silfver, 2007; Regan, 1971; Carlsmitth & Gross, 1969). Guilt-prone individuals also tend to be recognized by themselves and others as better leaders – they more frequently are chosen or volunteer for leadership positions (Schaumberg & Flynn, 2012).

**Evolutionary Scenarios for the Emergence of Guilt: Too Early To Tell?**

Does the evidence presented above point us toward an individual or group selection account of guilt? Intuitively, evidence that guilt-prone individuals exhibit less delinquent and more cooperative behavior, and that they are more apt to offer themselves for – and to be selected for – leadership positions supports both individual and group-level accounts. Being less counterproductive will tend to be good for individual actors and will tend to make for efficient groups. Assuming that being a leader is advantageous to the individual, and that groups tend to benefit from having guilt-prone leaders, this aspect of guilt proneness is clearly adaptive for individuals and groups.

Other consequences of guilt proneness, on the other hand, have fairly obvious benefits for the groups to which the individuals belong, but it is less certain that they are advantageous for the individual. For example, we have seen that manifesting guilt in legal trials can both help and hinder defendants. We also noted that the psychological discomfort of guilt experiences and the reparative and confessional behaviors that guilt motivates increase the chance that one’s transgressions will be discovered.

Finally, while it is good for group harmony to have individuals deterred from committing offenses, it may be better for individuals to lack a general deterrence from committing offenses so that they can strategically commit them when there will be no or little repercussion. While
these situations do not directly point to a group selection scenario for guilt, they present difficulties for individual-level accounts of guilt. At the very least, this suggests we should remain open to group selection models.

To sum up, the main accounts of the origin of guilt, like those of Joyce (2006), James (2011), and Frank (1988), focus on showing how guilt can be good for the individual. Although some aspects of guilt proneness seem to support an individual-level selection account of guilt, we’ve seen that some key features of guilt proneness may well be maladaptive for individuals. What we would like to conclude here is that we should take the group selection scenario seriously. We should not simply assume that guilt is individually adaptive, and that the sole task of giving an evolutionary account is to show how this is possible.

If we are right that we should take the group-level account seriously, how, then, might we decide between them? First of all, we should not think that the evolutionary story has to be one of either group or individual selection. It is quite possible that both levels of selection operated on guilt, and both promoted its evolution.

Evidence on the timing of the emergence of guilt can help us decide which evolutionary scenario is the most probable. The reason for this is that if guilt emerged preceding the origin of complex communication and cognitive abilities like reputation tracking, then bad reputations would not have been a negative factor for individuals. Individuals could take advantage of others without fear of eroding their reputation, and thus guilt would be a hindrance for individuals. In the absence of social traits like reputation, however, guilt proneness remains good at the group level: groups filled with individuals taking advantage of one another will not flourish and produce more groups than groups lacking such individuals. This is nothing more than Darwin’s (1871) insight about brave warriors discussed above.

A key source of information about the timing of the origins of guilt is the literature on whether nonhuman primates or other mammals exhibit guilt. If guilt is widespread, then this points to an early origin. Some primatologists and anthropologists contend that there is evidence that nonhuman primates are capable of experiencing at least a proto-form of guilt. For example, de
Waal (1996) suggests that some nonhuman primates are capable of internalizing social rules, and that the submission behaviors that subordinates perform before dominant conspecifics after violating those rules may be evidence of the phylogenetic starting point of guilt. A well-known example of such behaviors is exhibited in Coe and Rosenblum’s (1984) study of mating behaviors of male macaques. Subordinate males exhibited mating behaviors toward females in the absence of the dominant male, but would later perform submission behaviors when the latter returned. Similarly, Fessler and Gervais (2010) suggest that a guilt-like mechanism might best explain the reconciliation behaviors exhibited among nonhuman primates after intragroup conflicts.

Other researchers, however, are not confident that such submission and reconciliation behaviors constitute evidence of guilt in other primates. Drawing from his extensive fieldwork with chimpanzees in Gombe, Boehm (2012) contends that these behaviors are better explained in terms of fear of discovery and punishment subsequent to violation of social rules.

The current primatological data, then, appear inconclusive with respect to determining whether guilt or proto-guilt is present in nonhuman primates. Thus, in contrast to Joyce (2006), who maintains that there is no evidence of guilt in nonhuman primates, we keep open the possibility that guilt has a deeper phylogenetic origin than the hominin lineage.

This leads us to offer a provisional conclusion on the individual-level or group-level origins of guilt. We can frame this in terms of a pair of conditionals: If it turns out that guilt has a deep evolutionary history, then it likely evolved via group selection. If, on the other hand, guilt emerged relatively recently within our lineage, then the individual-selection account is tenable. This conclusion points to the central importance of data on the timing of the origin of guilt, and suggests how these data should lead us to favor one evolutionary scenario over another. Let’s now consider how we can move beyond these conditionals, and get closer to finding a solution to the evolutionary puzzle of guilt.
Future Directions in Research on the Evolution of Guilt

We have seen that the timing of the origins of guilt are of central importance to resolving the question of how guilt evolved. One vital area of research will involve work on how to identify guilt in nonhuman animals. How can we move beyond behaviors associated with guilt – such as the avoidance of future punishment – to identify the emotion of guilt in particular? This might involve reinterpreting existing behavioral data, generating new data, and further developing criteria for distinguishing genuine expressions of guilt.

In addition to further work on the identification of guilt, more work is needed in fleshing out and modeling evolutionary scenarios. For example, how strong could the group-selection pressure have been? If it is proposed as the primary factor underlying the emergence of the emotion, could it have been a sufficient cause of the emergence and maintenance of the emotion? What models of individual selection are capable of selecting for guilt proneness? And how much cognitive sophistication is really needed for guilt proneness to be individually adaptive?

Raising these important questions here helps to show what is needed in order to establish a clear evolutionary picture of guilt. The difficult task of answering them, we suspect, will require more collaboration and fruitful interchange between philosophers, biologists, and psychologists.

Acknowledgments

We thank Andrea Scarantino for helpful comments on an earlier draft of this article. This article was completed while one of us (Deem) was on a Claire Giannini Fund-supported fellowship at Children’s Mercy Hospitals and Clinics, and the other (Ramsey) was on a National Endowment for the Humanities-supported fellowship at the National Humanities Center. We thank the Claire Giannini Fund, CMH, NEH, and NHC for their support. Any views, findings, conclusions, or recommendations expressed in this article do not necessarily reflect those of the National Endowment for the Humanities.
References


