

On intrapersonal reciprocity

David Haig*

*Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street,
Cambridge, MA 02138, USA*

Received 26 June 2003; received in revised form 5 August 2003

Abstract

The existence of conflicts between different sets of genes within the genome is now widely accepted. But where there is conflict, there are also benefits to be gained from cooperation between the contending parties to reduce conflict costs. The potential for reciprocal altruism (Trivers, 1971) within an individual organism has hitherto attracted little attention but raises the possibility of complex interactions within the self.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Internal conflict; Genomic imprinting; Reciprocity; Prisoner's Dilemma

1. Introduction

Introspection suggests that I often attempt to modify my own behavior by an internally voiced mixture of exhortation, bribes, and threats. Such intrapersonal cajolement has limited effectiveness, in part, because threats and promises to myself lack credibility. If I renege on a contract with myself, who will enforce sanctions? If I offer myself a reward now—for later good behavior—why am I bound to fulfil the second half of the bargain? Is it credible that I would knowingly do harm to myself to punish a past transgression?

The existence of intrapersonal persuasion poses a philosophical conundrum. I know what I know, and I know what I want. So, why do I need to persuade myself of anything? One might argue that internal bargaining is simply a reuse, or misapplication, of tools that are effective in the control of others' behavior to self-control. I modify my behavior in response to the threats

* Tel.: +1-617-496-5125; fax: +1-617-495-5667.

E-mail address: dhaig@oeb.harvard.edu (D. Haig).

and bribes of others and, in turn, use threats and bribes to modify others' behavior. Why not use these same methods—of proven effectiveness—on myself? Or, one might argue that intrapersonal bargaining resembles interpersonal bargaining because the self contains multiple agents with different, sometimes conflicting, agendas. It is this possibility—of an internally complex self in which there is potential for strategic negotiation between contending parties—that I will explore in this note. In particular, I will address the possibility of intrapersonal reciprocity.

2. An internal conflict

Evolutionary theory has revealed the possibility of conflicts within organisms that arise when different replicators have different rules of transmission (Hurst, Atlan, & Bengtsson, 1996). Could these conflicts underlie problems of self-control? The potentially competing agendas within the self, both genetic and memetic, are complex. Therefore, to focus discussion it will be useful to consider a simple exemplar of an internal genetic conflict. For this purpose I will consider a conflict between the genes that an individual (call him Bob, in honor of Robert Trivers) inherits from his mother and his father.

Suppose that Bob's parents divorce and remarry, and that his mother and father each have a child in their new marriages. Bob thus has a maternal half-sib (Maddy) and a paternal half-sib (Paddy). What cost (C) would Bob's genes be prepared to accept in exchange for a benefit (B) to Maddy? Bob's maternally derived (henceforth madumnal) genes have one chance in two of being transmitted to Maddy. Therefore, madumnal alleles would favor Maddy receiving the benefit, as long as its value exceeded twice the cost to Bob ($B > 2C$). However, Bob's paternally derived (henceforth padumnal) genes are absent from Maddy. For these genes, no benefit—no matter how great—to Maddy could justify any cost—no matter how small—to Bob (i.e., $C < 0$ for the exchange to be profitable). Bob's madumnal and padumnal genomes are in conflict whenever $B > 2C > 0$. A similar conflict exists over benefits that Bob could confer on Paddy, but in this case the roles of madumnal and padumnal genomes are reversed.

3. Conflict resolution

In what ways could Bob's internal conflict be resolved? The traditional answer invokes a "veil of ignorance" (Rawls, 1971). If genes lack information about their parental origin, they are constrained to behave in the same manner when inherited from either parent. If one of Bob's genes is faced with a choice whether to help Maddy, the gene is equally likely to be padumnal, and necessarily absent from Maddy, or madumnal, with one chance in two of being present in Maddy (presence or absence, of course, referring to identity by recent common descent). Therefore, an uninformed gene has one chance in four of being present in Maddy and would favor transfer of the benefit if $B > 4C$.

What theoretical predictions can be made about how the conflict would be resolved if the relevant genes are imprinted and thus possess information about their parental origin (Goos &

Silverman, 2001; Haig, 2002)? The simple answer is little: genes of both parental origins would favor withholding the benefit from Maddy if $B < 2C$ and both sets of genes would favor conferring the benefit if this directly benefited Bob ($C < 0$). Specific prediction of outcomes within the zone of conflict defined by $B > 2C > 0$ require assumptions about the mode of gene action and the relative power of the contending parties to influence the decision.

One possible resolution of a conflict is that one party has the power to dictate the outcome. If Bob's padummal genome had the dictatorial role, the benefit would only be conferred if $C \leq 0$, whereas, if Bob's madummal genome had the dictatorial role, the benefit would be conferred whenever $B > 2C$. In the first case, Bob's padummal genome can be considered to have veto power as to whether the benefit is conferred, whereas, in the second case, Bob's madummal genome can be considered to be able to present his padummal genome with a *fait accompli* (Haig, 1992). Whether a particular dictatorial decision is interpreted as a veto or *fait accompli* may simply depend on how the question is framed. Models of gene expression at an imprinted locus result in a dictatorial outcome when madummal and padummal alleles have conflicting interests over the level of gene expression. At evolutionary equilibrium, whichever allele favors the higher amount produces that amount (the "loudest-voice-prevails" principle; Haig, 1997a).

Stalemates can result if different genes have dictatorial power over different aspects of a decision-making process. Wilkins and Haig (2001) provide an example of this process in a two-locus model of the coevolution of a padummally expressed demand enhancer and a madummally expressed demand inhibitor. Increased padummal production of the demand enhancer favors increased madummal production of the demand inhibitor which favors increased padummal production of the demand enhancer, and so on, until the evolutionary spiral comes to a halt because any further escalation would be prohibitively expensive for both parties. At evolutionary equilibrium in this model, the loudest voice prevails at both loci and there are substantial costs of conflict.

A possibility that has received little attention is that madummal and padummal genes could reach a deal that is a compromise between their competing interests, without having to bear the full costs of evolutionary escalation. Suppose that separate opportunities arise for Bob to confer benefits on Maddy and Paddy. If the two opportunities were considered separately, the inclusive fitness of Bob's padummal genes would suffer if the benefit were conferred on Maddy, whereas Bob's madummal genes would suffer if the benefit were conferred on Paddy. Therefore, if Bob's madummal and padummal genes each had an effective veto on decisions whether to confer a costly benefit on half-sibs, neither benefit would be conferred. However, if the two decisions were considered jointly, both sets of genes might be able to increase their fitness if they could strike a deal. In effect, Bob's madummal genes could offer his padummal genes a bargain, "We will let you help Maddy if you let us help Paddy." The value of the deal for both parties is $B/2 - 2C$. That is, each of Bob's genes experiences the costs of conferring the benefit on both Maddy and Paddy, in return for an expected half-share in the benefit to Maddy, but no share in the benefit to Paddy (or vice versa). If $B > 4C$, both sets of genes would receive a net inclusive fitness benefit relative to what they would achieve if they exercised their respective vetoes. If the deal could be trusted, it should be accepted.

In the above example, genes that knew their parental origin settle on the same decision rule ($B > 4C$) as that which would be used by genes that were uninformed about their parental

origin. This is a consequence of artificial symmetries in the model and the limitation of bargaining to a single take-it-or-leave-it deal. Suppose instead that Bob never interacts with Paddy—he lives with Bob’s father in another city—and that Bob’s only opportunities to benefit a half-sib involve Maddy. Bob’s madumnnal alleles have no opportunities to exercise their veto and, therefore, have nothing to offer in exchange for Bob’s padumnnal alleles withholding their veto. In this scenario, uninformed genes would use the decision rule $B > 4C$, but Bob’s padumnnal alleles, informed of their parental origin, have dictatorial power and should exercise their veto whenever $C > 0$.

Now consider an intermediate case in which there are two opportunities to benefit Maddy (combined benefit to Maddy = $2B$; combined cost to Bob = $2C$) for each opportunity to benefit Paddy (benefit to Paddy = B ; cost to Bob = C). In one possible deal (“two-for-one”), Bob’s padumnnal alleles would withhold their veto on both benefits to Maddy in exchange for Bob’s madumnnal alleles withholding their veto on the single benefit to Paddy. The value of this deal to Bob’s madumnnal alleles is $B - 3C$ whereas its value to Bob’s padumnnal alleles is $B/2 - 3C$. Bob’s madumnnal alleles would accept the deal if $B > 3C$ whereas his padumnnal alleles would accept the deal if $B > 6C$, but reject the deal if $B < 6C$. However, a deal in which one benefit to Maddy is exchanged for the benefit to Paddy (“one-for-one”) is also possible. As shown above, this deal would benefit both parties (relative to no deal) if $B > 4C$. If $B < 4C$, neither deal is acceptable to Bob’s padumnnal alleles who should exercise their veto. If $6C > B > 4C$, the “one-for-one” deal is acceptable to both parties, but the “two-for-one” deal is unacceptable to Bob’s padumnnal alleles. “One-for-one” should be accepted by both parties. If $B > 6C$, either deal is superior to no deal for both parties, but Bob’s padumnnal alleles would prefer “one-for-one” whereas his madumnnal alleles would prefer “two-for-one.” Theory provides no simple prediction as to which deal, if any, will be chosen. The complications of multiple possible contracts are exacerbated if costs and benefits are continuous rather than discrete.

4. Repeated interactions

Genes of course are not subject to binding contracts. In the absence of enforceable contracts, the proposed deal-trading benefits to Maddy and Paddy has the nature of a Prisoner’s Dilemma. If the effects of genes are fixed, and not conditional on the behavior of the other genes with which they are temporarily associated, the deal is a one-shot Prisoner’s Dilemma, and evolutionarily unenforceable.

Although mutual defection is the evolutionarily stable strategy (ESS) if a Prisoner’s Dilemma is played once with a particular partner, cooperation can sometimes succeed if the game is played repeatedly with the same partner and there is uncertainty about when the relationship will end (Axelrod & Hamilton, 1981; Bendor & Swistak, 1997). If this criterion of long-term association of uncertain duration were the only consideration, strategic cooperation among genes within an organism should be possible. The possibility was recognized by Axelrod and Hamilton (1981) in their discussion of female meiotic segregation as an iterated Prisoner’s Dilemma (IPD). Most successful cooperative strategies in IPDs require memory of past interactions to allow a player’s actions to be conditional on the past

behavior of its partner. If genes were to implement a simple strategy like Tit For Tat (“do what the other player did on the previous move”), they would need to perceive, and remember, a partner’s previous move and use this information to guide current behavior.

Analytical investigations and computer simulations of the IPD have revealed a wealth of possibilities for complex behavior. In general, no strategy is an ESS and multiple strategies may coexist in a population. Polymorphism is possible because two strategies may be indistinguishable when played against each other, so that their relative performance will be determined by their interactions with third strategies, including the degenerate versions of themselves produced by mutation (Bendor & Swistak, 1997). In the context of intrapersonal reciprocity, genetic polymorphism would raise the intriguing possibility that different combinations of strategies in the internal IPD could contribute to personality differences: sometimes one interest gets its way; sometimes the different interests come to a relatively harmonious compromise; sometimes the different interests contest every point of contention.

It is tempting to simply transfer results obtained from theoretical investigations of the IPD to discussions of reciprocity within the genome. I resist the temptation because most simulations of the IPD have competed strategies in a stereotypical tournament (Bendor & Swistak, 1997). In each round of the tournament, pairs of strategies play multiple rounds of the Prisoner’s Dilemma, with “fitness” determined by the sum of pay-offs. In the next round, a strategy increases or decreases in frequency determined by its performance relative to the average fitness of the population. Significantly, a strategy is inherited as a whole and does not undergo recombination. This kind of tournament models asexual or single-locus inheritance of strategies.

Consider however an alternative tournament in which teams play against each other, and in which the members of a team have different tasks in the collective implementation of their team’s strategy. In this tournament, a team’s relative performance in one round determines the probability that its members will participate in the next round, but the teams themselves are ephemeral. At the end of each round, members of successful teams (or their clonal progeny) are reorganized into new teams, which are once again competed in pairs. This kind of tournament models sexual or multilocus inheritance of strategies, with the units of inheritance determining components of strategies rather than entire coherent strategies. A successful component must do well, on average, across all the combinations in which it is tested. The sexual tournament would appear to select for a team of champions rather than a champion team (Haig, 1997b). It is possible that the simpler asexual tournaments provide a good guide to equilibrium behavior in the more complex sexual tournaments, but this is something to be demonstrated rather than simply asserted.

5. Are genes that smart?

The implementation of a conditional strategy in an IPD requires a degree of sophistication that some may be wary of ascribing to genes, even though genes are known to possess forms of memory and to exhibit complex conditional expression. There is a danger, however, in underestimating the options available to genes. Imprinted genes, for example, remember the

sex of the parent from whom they were inherited and use this information to guide strategic behavior. The *GNAS* locus, to take a particularly complex example, produces transcripts from multiple promoters in a complex tissue-specific pattern, with further diversity generated by alternative splicing. Some transcripts are expressed from both parental alleles, some are expressed only from the maternal allele, some only from the paternal allele, and some show imprinted or biallelic expression depending on cell type. Some transcripts are translated, whereas others function solely as RNAs. One transcript is antisense to an oppositely imprinted transcript. Another encodes two interacting proteins in overlapping reading frames (Hayward, Moran, Strain, & Bonthron, 1998; Holmes et al., 2003; Klemke, Kehlenbach, & Huttner, 2001; Wroe et al., 2000).

There is also a danger of overestimating the difficulties of genes implementing a reciprocal strategy, if one's implicit model is of a group of organisms with complex nervous systems. Green-beard effects have similarly been considered implausible, but Haig (1996) has shown how single genes, or pairs of closely linked genes, can encode the green-beard triad of feature, recognition, and altruistic response (for a concrete example, see Queller, Ponte, Bozzaro, & Strassmann, 2003). A gene such as *GNAS* (see above) has a much more versatile behavioral repertoire than the successful Tit For Tat algorithm in Axelrod's IPD tournament (Axelrod & Hamilton, 1981).

Finally, cooperation between organisms is the result of complex patterns of gene expression that result from the interaction of incoming information with evolved genetic algorithms. Why should reciprocity be more difficult to evolve for interactions between genes within organisms than for interactions between genes in different organisms?

6. Prospects

Bob's dilemma over helping his half-sibs was chosen as a simple, and symmetric, example with which to illustrate some underlying issues in the resolution of internal conflicts. It was not chosen for its realism, nor did I intend to restrict future discussions of intrapersonal reciprocity to interactions between genetic agents. Some of the most interesting conflicts within the human psyche occur between genetic and memetic replicators, or between different memetic replicators, and there may also be opportunities in these interactions for mutually beneficial reductions in the costs of conflict.

Internal conflicts may often be played out over predispositions to behave in particular ways—between generosity and selfishness or between impulsiveness and rational calculation—rather than over specific decisions whether to help particular individuals of specified relatedness. For an individual in a social group with male-biased dispersal, other group members may be close matrilineal relatives, but patrilineal nonrelatives. If so, an individual's maternal genes may favor relatively more group-beneficial behaviors whereas her paternal genes may favor relatively more self-beneficial behaviors (Haig, 2000). In a dilemma whether to plant a seed or consume it now, some genes within an individual may be more likely than others to still be around to reap the harvest.

Speculation about intrapersonal reciprocity prompts the question how one would ever know that it is occurring. How does one distinguish strategic cooperation from veil-of-ignorance cooperation on the one hand and from escalated noncooperative stalemates on the other? The problem of testability has bedevilled progress in the study of internal conflicts by behavioral biologists, because there may be little that one can observe in external behavior that allows one to test hypotheses about internal conflicts. Progress in this field is likely to come from understanding the details of genetic mechanisms, just as an appreciation of internal conflicts between maternal and paternal genes has in large part depended on increased knowledge about the molecular functions of imprinted genes (Wilkins & Haig, 2003). One will need to demonstrate conditionality of gene expression, not just expression that is conditional on parental origin, but also expression that is conditional on the behavior of other genes. This is a challenge, but surely not one that is insurmountable.

I began this essay with a discussion of the subjective experience of internal conflicts, but then shifted ground to a discussion of metaphorical conflicts between genes. What, if anything, does subjective conflict have to do with conflicting evolutionary forces acting on genes? Clearly, subjective experience is in some manner dependent on neural activity, and hence on gene expression, but beyond that, all is mystery.

Acknowledgments

The manuscript benefited from the comments of Martin Daly, Daniel Dennett, Robert Trivers, and an anonymous reviewer.

References

- Axelrod, R., & Hamilton, R. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.
- Bendor, J., & Swistak, P. (1997). The evolutionary stability of cooperation. *American Political Science Review*, *91*, 290–307.
- Goos, L. M., & Silverman, I. (2001). The influence of genomic imprinting on brain development and behavior. *Evolution and Human Behavior*, *22*, 385–407.
- Haig, D. (1992). Intragenomic conflict and the evolution of eusociality. *Journal of Theoretical Biology*, *156*, 401–403.
- Haig, D. (1996). Gestational drive and the green-bearded placenta. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 6547–6551.
- Haig, D. (1997a). Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society of London B*, *264*, 1657–1662.
- Haig, D. (1997b). The social gene. In: J. R. Krebs, & N. B. Davies (Eds.), *Behavioural Ecology* (4th ed., pp. 284–304). Oxford: Blackwell Scientific.
- Haig, D. (2000). Genomic imprinting, sex-biased dispersal, and social behavior. *Annals of the New York Academy of Sciences*, *907*, 149–163.
- Haig, D. (2002). *Genomic imprinting and kinship*. New Brunswick, NJ: Rutgers University Press.
- Hayward, B. E., Moran, V., Strain, L., & Bonthron, D. T. (1998). Bidirectional imprinting of a single gene: *GNAS1* encodes maternally, paternally, and biallelically derived proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 15475–15480.

- Holmes, R., Williamson, C., Peters, J., Denny, P., RIKEN GER Group and GSL Members, & Wells, C. (2003). A comprehensive transcript map of the mouse *Gnas* imprinted complex. *Genome Research*, 13, 1410–1415.
- Hurst, L. D., Atlan, A., & Bengtsson, B. O. (1996). Genetic conflicts. *Quarterly Review of Biology*, 71, 317–364.
- Klemke, M., Kehlenbach, R. H., & Huttner, W. B. (2001). Two overlapping reading frames in a single exon encode interacting proteins: a novel way of gene usage. *EMBO Journal*, 20, 3849–3860.
- Queller, D. C., Ponte, E., Bozzaro, S., & Strassmann, J. E. (2003). Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science*, 299, 105–106.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Wilkins, J. F., & Haig, D. (2001). Genomic imprinting of two antagonistic loci. *Proceedings of the Royal Society of London B*, 268, 1861–1867.
- Wilkins, J. F., & Haig, D. (2003). What good is genomic imprinting: the function of parent-specific gene expression. *Nature Reviews Genetics*, 4, 359–368.
- Wroe, S. F., Kelsey, G., Skinner, J. A., Bodle, D., Ball, S. T., Beechey, C. V., Peters, J., & Williamson, C. M. (2000). An imprinted transcript, antisense to *Nesp*, adds complexity to the cluster of imprinted genes at the mouse *Gnas* locus. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 3342–3346.