Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

How the mind makes welfare tradeoffs: evolution, computation, and emotion

Andrew W Delton[1,2,3] and Theresa E Robertson[2]

It feels easy and intuitive to make decisions about *welfare tradeoffs* — decisions pitting personal welfare against the welfare of someone else. Just because something feels easy, however, does not mean the computations that give rise to it are simple. We review evidence that natural selection has designed a series of *internal regulatory variables* that encode features of the other person (e.g., kinship, formidability, cooperative value) and the situation (e.g., the magnitude of the welfare at stake). These variables combine into a final variable, a *welfare tradeoff ratio*, which determines welfare tradeoffs. Moreover, some emotions, such as anger and forgiveness, function to update welfare tradeoff ratios in your mind and the minds of others. Conscious simplicity hides complex evolved design.

Addresses
[1] Department of Political Science, Stony Brook University, United States
[2] College of Business, Stony Brook University, United States
[3] Center for Behavioral Political Economy, Stony Brook University, United States

Corresponding author: Delton, Andrew W (andrew.delton@stonybrook.edu)

Current Opinion in Psychology 2015, **7**:12–16

This review comes from a themed issue on **Evolutionary psychology 2016**

Edited by **Steven Gangestad** and **Josh Tybur**

http://dx.doi.org/10.1016/j.copsyc.2015.06.006

Making mundane decisions like whether to drive a friend to the airport ('sure, I've got time') or to leave the dishes for your spouse while you watch television ('those *Friends* reruns won't watch themselves') is almost always easy and intuitive. Surprisingly, this is even true in life and death situations, where some deliberation might seem warranted: people who take heroic risks, like jumping into a river to save a drowning child, describe their decision as being quick, intuitive, and made without conscious thought [1]. Whether serious or mundane, in *welfare tradeoffs* like these we must decide whether to trade off our personal welfare to enhance the welfare of someone else [2,3••].

Although these decisions seem easy to make, the history of cognitive science suggests that just because something feels consciously easy does not mean the computations underlying it are simple. Although vision feels simple — we open our eyes and there the world is — its computations are extraordinarily complex [4]. As we show, the computations that enable welfare tradeoffs are also anything but simple. Instead, making welfare tradeoffs involves an integrated network of *internal regulatory variables* [2,3••]. Internal regulatory variables are quantitative representations that encode features of the self and others (e.g., relatedness, value as a cooperation partner) and are used in decision making. Although making a welfare tradeoff feels easy or simple, it is actually guided by this complex network of variables.

The evolutionary biology of welfare tradeoffs

Reverse engineering a cognitive system requires a theory of what should be computed. In the case of vision, cognitive scientists can draw from theories in physics. But when it comes to social behavior, physics only goes so far. Fortunately, evolutionary biology has produced formal theories of when and how organisms should trade off their own welfare in favor of others. These theories can guide us in developing hypotheses about internal regulatory variables.

Inclusive fitness theory, for example, describes how organisms should trade off their welfare in favor of genetic kin [5]. This theory predicts that a focal organism ('you') should trade off its welfare in favor of another when the following condition obtains:

$$r \times b_{\text{other}} > c_{\text{you}}$$

This theory requires that that the benefits the other receives, b_{other}, are greater than costs you incur providing those benefits, c_{you} — but only after discounting those benefits by r, an index of genetic relatedness. As relatedness increases, the discounted benefits are more likely to exceed the costs.

Reciprocity theory describes how organisms should exchange benefits back and forth over time [6,7]. Although its domain differs from inclusive fitness theory, the equation is nearly identical:

$$w \times b_{\text{other}} > c_{\text{you}}$$

Instead of discounting the benefits by relatedness, however, this formula discounts the benefits by w, an index of how long the relationship will probably last. Longer relationships make exchange more likely.

Similar theories exist to describe not just cooperation and generosity, but also aggressive contests. According to one theory of conflicts [8], an animal should cede a resource to another according the following rule:

$$f \times b_{\text{other}} > c_{\text{you}}$$

Here, f indexes how much more formidable the other animal is than you. The greater the disparity in formidability, the more likely you are to cede the resource (even when the contest is over a fixed resource b_{other} and c_{you} are not necessarily equal because the two animals may value the resource differently).

These theories and others describe variables that determine welfare tradeoffs [9]. But there is a hidden problem: you cannot simultaneously give a resource to someone because they are your full sibling and withhold it because they are a terrible reciprocity partner. This leads to the hypothesis that the mind computes a summary variable that integrates internal regulatory variables about features of people (e.g., kinship or formidability) and about situations (e.g., the nature or quantity of the resource). This integration produces a final variable used for making welfare tradeoffs, a *welfare tradeoff ratio* (WTR) [2,3••,10,11]. The mind should trade off personal welfare when the following condition is satisfied:

$$\text{WTR} \times b_{\text{other}} > c_{\text{you}}$$

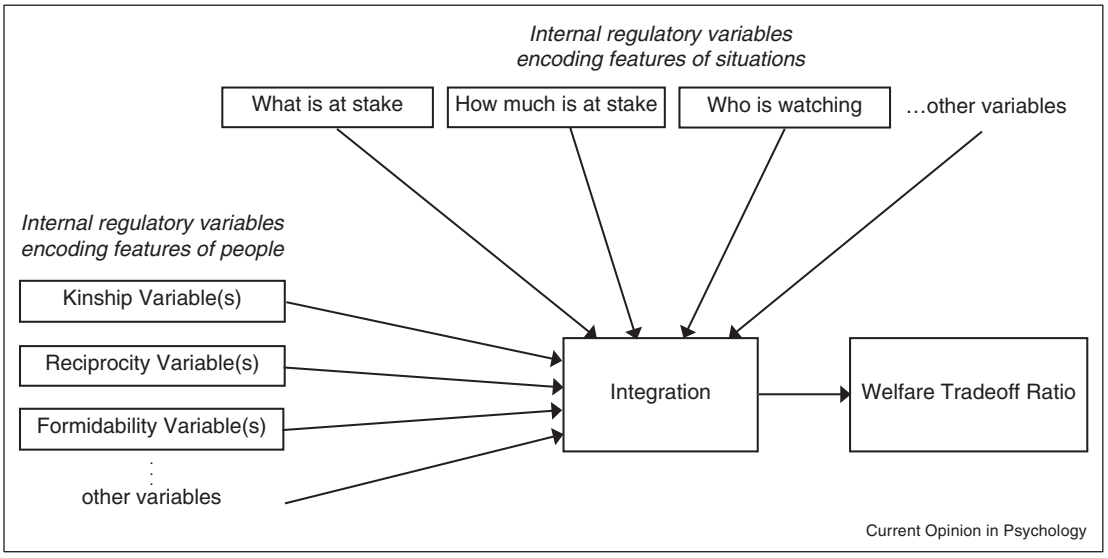
The greater the welfare tradeoff ratio, the more weight you place on the other person's welfare and the more likely you are to benefit them. Of course, the actual computations are likely to be more complicated, with future research yielding a more complex equation.

Below we review evidence that (a) the mind computes internal regulatory variables encoding kinship, reciprocity, and formidability, (b) the mind computes welfare tradeoff ratios by combining these other internal regulatory variables with situational information, and (c) some emotions are designed to change welfare tradeoff ratios in your mind and in the minds of others. Figure 1 summarizes key parts of the model.

The mind uses multiple cues to compute internal regulatory variables

Many internal regulatory variables are computed using multiple cues. For example, the variable encoding kinship between siblings is determined by at least two cues [10]: time living together growing up [10,12] and seeing your mother care for an infant (e.g., breastfeeding) [10,13]. However, the cues are not additive — they are non-compensatory, meaning the most predictive cue takes precedence [14]. Seeing your mother care for an infant is the better cue, because intense, prolonged neonatal care is almost always directed at a mother's own child. When this is present, living together has little effect on kinship variables. However, this cue is only ever available to older siblings. When it is

Figure 1



Model of the welfare tradeoff system. The model assumes that welfare tradeoff ratios are computed based on two types of internal regulatory variables. One set, shown in the column on the left, encodes features of people, such as relatedness or value as a reciprocity partner. (Not shown is that each of these variables is itself computed based on multiple cues. For instance, formidability is based on both personal physical strength and one's allies.) The other set, shown in the row on the top, encodes features of the situation, which would generally be more transient than features of people, such as who is watching or what types of resources are at stake. The final welfare tradeoff ratio is then used in determining behavior.

unavailable, the mind instead uses co-residence duration to predict sibship.

Variables encoding formidability and reciprocity value are also computed using multiple cues. Formidability can be assessed from the face, voice, and body [15–19]. Other cues to formidability include the presence of a weapon [20] or allies [21]. Formidability cues are also non-compensatory: size is a good, though imperfect, cue to someone's strength. Although it is used to infer formidability when it is the only cue available, it is deprecated when direct evidence of strength is available [22•]. Besides cues of how long a relationship will last [23–25], reciprocity value is also predicted by a person's ability to provide benefits and their willingness to cooperate [26–28].

The mind computes welfare tradeoff ratios

Cognitive psychology has a long history of using decision-making tasks to study internal regulatory variables, like those governing time preference ($5 now or $50 in 3 months?) and risk preference ($5 for sure or a 5% chance of $50?) [29–31]. Welfare tradeoff ratios have been studied using similar tasks. For example, an experimental subject might complete a series of monetary choices that affect themselves and another person [32]. Do you get $5 or a friend $10? You $10 or the friend $10? You $15 or the friend $10? By systematically varying the amounts at stake, researchers can estimate a subject's welfare tradeoff ratio toward their friend by looking for 'switch points,' the choices where subjects stop favoring their friend and start favoring themselves [33•,34]. For instance, a welfare tradeoff ratio of .25 toward a friend implies you would pass up anything less than $2.50 to give them $10 (e.g., $.25 \times \$10 > \$2.25$). People find this task easy and intuitive [33•,34], make similar decisions whether choices are real or hypothetical [35], and show similar decision-making processes across societies [36–38]. Moreover, this task recruits brain systems that encode valuation [39].

Two related approaches use this technique. One approach, the welfare tradeoff approach we take, is primarily inspired by evolutionary psychology. A typical experiment asks subjects about specific others (e.g., specific kin or friends) whose relationship with the subject can be directly measured (e.g., [33•,37•]). The other approach to welfare tradeoffs, called social discounting, is primarily inspired by cognitive psychology theories of time and risk discounting. Researchers in this approach usually ask about unspecified others who subjects are asked to imagine differ in their social closeness to the subject; welfare of less close others should be discounted most heavily (e.g., [32]). These researchers have shown that making welfare tradeoffs is dissociable from time and risk preferences [40], even while it obeys the same complex mathematical patterns these other internal regulatory variables do [32,41]. Ultimately, both literatures are addressing the same question: How does the mind make tradeoffs between personal welfare and the welfare of others?

Using these tasks, researchers have shown that welfare tradeoff ratios predict behavior in other contexts. Greater welfare tradeoff ratios toward others in general predict greater cooperation in laboratory games that involve tension between personal welfare and the welfare of others (the prisoners' dilemma and public goods games) [40,42•,43]. Greater welfare tradeoff ratios correlate with greater agreeableness and fluid intelligence [33•,44]. Women who smoke are more likely to quit when they become pregnant if they generally place more weight on others' welfare [45]. Boys with externalizing behavior problems place less weight on others' welfare [46]. Welfare tradeoffs can even be increased by experimental exposure to MDMA, a psychoactive drug known to have prosocial effects [33•].

Multiple internal regulatory variables are combined to compute welfare tradeoff ratios

We approach this hypothesis in two parts. First, do variables encoding kinship or reciprocity — taken one at a time — determine welfare tradeoffs? Yes: people are more likely to trade off their welfare in favor of close kin's welfare. This is true whether measured with rating scales [10,12] or with tasks with money at stake [47,48]. In the domain of reciprocity and friendship, people have higher welfare tradeoff ratios toward friends than strangers [33•] and when there is less social distance, whether measured as emotional closeness [32,49] or expectations of reciprocity [47,49].

Second, are multiple factors integrated to set welfare tradeoff ratios? Yes: for instance, kinship and reciprocity variables and situational cues about the size of the benefits are integrated in non-additive ways: the theory of inclusive fitness is a theory of unilateral giving, whereas reciprocity requires that benefits given eventually be returned. Thus, reciprocity presents an investment risk and this risk increases with stakes — a friend will have many chances to return small favors but few, if any, to return heroic sacrifices. As these considerations predict, when giving to kin people's willingness to make welfare tradeoffs remains relatively stable or even increases as the stakes increase (e.g., picking up a check versus donating a kidney); for non-kin, however, increasing stakes cause marked declines in welfare tradeoffs [50–52]. Moreover, differences between kin and non-kin cannot simply be reduced to a general variable of 'emotional closeness.' Even when holding emotional closeness constant, people have higher welfare tradeoffs for kin [37•,48,53], suggesting that multiple variables, some encoding kinship and others features like reciprocity value, are integrated in welfare tradeoff decisions.

Welfare tradeoff ratios are updated by emotions

Welfare tradeoff ratios are hypothesized to be dynamic — updated in light of new information. Some emotions

may have been designed, in part, to achieve this updating, either changing welfare tradeoff ratios in one's own mind or in the minds of others [2,3••]. Anger, for instance, appears designed to raise other people's welfare tradeoff ratios toward the self [18,54••]. To do this, the mind must estimate another person's welfare tradeoff ratio toward the self [11] and compare it to an expectation of what it should be. If the actual welfare tradeoff ratio is too low, then anger is activated, which causes bargaining for better treatment, such as by threatening withdrawal of cooperation (e.g., lowering your own welfare tradeoff ratio toward them).

Forgiveness may be another emotion designed to update welfare tradeoff ratios [54••]. In particular, it signals the possibility of reconciliation after bad behavior. If you treat me poorly, I might temporarily lower my welfare tradeoff ratio toward you in anger, to bargain for better treatment. Forgiveness signals an end to hostilities: I will return my welfare tradeoff ratio to normal levels if you will raise yours as well [54••]. It is not surprising, then, that forgiveness is more likely to be offered if the transgressor is judged to be valuable, likely to raise their welfare tradeoff ratio in return, or making conciliatory gestures [27,28].

Conclusion

The mind computes and dynamically updates welfare tradeoff ratios and other internal regulatory variables. But more work remains: What are other cues that determine variables, such as those indexing kinship, reciprocity, or formidability? What other situational cues moderate the effects of these variables on welfare tradeoff ratios? Do other emotions, like gratitude or pride, use or update welfare tradeoff ratios? We suspect that further work will continue to emphasize that even though making welfare tradeoffs feels easy and simple, they are enabled by a complex and subtle cognitive foundation.

Conflict of interest statement

There are no conflicts of interest to the best of the author's knowledge.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Rand DG, Epstein ZG: **Risking your life without a second thought: intuitive decision-making and extreme altruism**. *PLoS ONE* 2014, **9**:e109687.

2. Tooby J, Cosmides L, Sell A, Lieberman D, Sznycer D: **Internal regulatory variables and the design of human motivation: a computational and evolutionary approach**. In *Handbook of Approach and Avoidance Motivation.* Edited by Elliot AJ. Lawrence Erlbaum Associates; 2008:251-271.

3. •• Cosmides L, Tooby J: **Evolutionary psychology: new perspectives on cognition and motivation**. *Annu Rev Psychol* 2013, **64**:201-229.

This recent review describes internal regulatory variables in more detail and connects this approach to other recent advances in evolutionary psychology (e.g., concepts, visual attention).

4. Marr D: *Vision*. W.H. Freeman & Company; 1982.

5. Hamilton WD: **The genetical evolution of social behaviour**. *J Theor Biol* 1964, **7**:1-52.

6. Axelrod R, Hamilton WD: **The evolution of cooperation**. *Science* 1981, **211**:1390-1396.

7. Trivers RL: **Evolution of reciprocal altruism**. *Q Rev Biol* 1971, **46**:35-57.

8. Hammerstein P, Parker GA: **The asymmetric war of attrition**. *J Theor Biol* 1982, **96**:647-682.

9. Nowak MA: **Five rules for the evolution of cooperation**. *Science* 2006, **314**:1560-1563.

10. Lieberman D, Tooby J, Cosmides L: **The architecture of human kin detection**. *Nature* 2007, **44**:727-731.

11. Delton AW, Robertson TE: **The social cognition of social foraging**. *Evol Hum Behav* 2012, **33**:715-725.

12. Lieberman D, Lobel T: **Kinship on the Kibbutz: coresidence duration predicts altruism, personal sexual aversions and moral attitudes among communally reared peers**. *Evol Hum Behav* 2012, **33**:26-34.

13. Lieberman D: **Rethinking the Taiwanese minor marriage data: evidence the mind uses multiple kinship cues to regulate inbreeding avoidance**. *Evol Hum Behav* 2009, **30**:153-160.

14. Gigerenzer G, Todd PM, the ABC Research Group: *Simple Heuristics That Make Us Smart*. Oxford University Press; 1999.

15. Sell A, Bryant GA, Cosmides L, Tooby J, Sznycer D, von Rueden C, Krauss A, Gurven M: **Adaptations in humans for assessing physical strength from the voice**. *Proc R Soc B* 2010, **277**:3509-3518.

16. Sell A, Cosmides L, Tooby J, Sznycer D, von Rueden R, Christopher, Gurven M: **Human adaptations for the visual assessment of strength and fighting ability from the body and face**. *Proc R Soc Lond B Biol Sci* 2009, **276**:575-584.

17. Sell A, Cosmides L, Tooby J: **The human anger face evolved to enhance cues of strength**. *Evol Hum Behav* 2014, **35**:425-429.

18. Sell A, Tooby J, Cosmides L: **Formidability and the logic of human anger**. *Proc Natl Acad Sci U S A* 2009, **106**:15073-15078.

19. Delton AW, Sell A: **The co-evolution of concepts and motivation**. *Curr Dir Psychol Sci* 2014, **23**:115-120.

20. Fessler DMT, Holbrook C, Snyder JK: **Weapons make the man (larger): formidability is represented as size and strength in humans**. *PLoS ONE* 2012, **7**:e32751.

21. Fessler DMT, Holbrook C: **Friends shrink foes: the presences of comrades decreases the envisioned physical formidabiity of an opponent**. *Psychol Sci* 2013, **24**:797-802.

22. • Pietraszewski D, Shaw A: **Not by strength alone: children's conflict expectations follow the logic of the asymmetric war of attrition**. *Hum Nat* 2015, **26**:44-72.

In a careful series of studies, the authors show what cues the mind uses to determine who wins in a resource contest. Cues includes size, strength, ownership, and valuation of the resource. This is the first paper to investigate how multiple cues of formidability are combined.

23. Delton AW, Krasnow MM, Cosmides L, Tooby J: **The evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters**. *Proc Natl Acad Sci U S A* 2011, **108**:13335-13340.

24. Krasnow MM, Delton AW, Tooby J, Cosmides L: **Meeting now suggests we will meet again: implications for debates on the evolution of cooperation**. *Nat Sci Rep* 2013, **3**:1747.

25. Delton AW, Krasnow MM: **Adaptationist approaches to moral psychology**. In *The Moral Brain.* Edited by Decety J, Wheatley T. MIT Press; 2015:19-34.

26. Petersen MB, Sell A, Tooby J, Cosmides L: **To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice**. *Evol Hum Behav* 2012, **33**:682-695.

27. Burnette JL, McCullough ME, Tongeren DRV, Davis DE: **Forgiveness results from integrating information about relationship value and exploitation risk**. *Pers Soc Psychol Bull* 2012, **38**:345-356.

28. McCullough ME, Pedersen EJ, Tabak BA, Carter EC: **Conciliatory gestures promote forgiveness and reduce anger in humans**. *Proc Natl Acad Sci U S A* 2014, **111**:11211-11216.

29. Green L, Myerson J: **A discounting framework for choice with delayed and probabilistic rewards**. *Psychol Bull* 2004, **130**:769-792.

30. Griskevicius V, Tybur JM, Ackerman JM, Delton AW, Robertson TE, White AE: **The financial consequences of too many men: sex ratio effects on saving, borrowing, and spending**. *J Pers Soc Psychol* 2012, **102**:69-80.

31. Griskevicius V, Ackerman JM, Cantú SM, Delton AW, Robertson TE, Simpson JA, Thompson ME, Tybur JM: **When the economy falters, do people spend or save? Responses to resource scarcity depend on childhood environments**. *Psychol Sci* 2013, **24**:197-205.

32. Jones BA, Rachlin H: **Social discounting**. *Psychol Sci* 2006, **17**:283-286.

33. • Kirkpatrick M, Delton AW, Robertson TE, de Wit H: **Prosocial effects of MDMA: a measure of generosity**. *J Psychopharmacol* 2015, **29**:661-668.
This is the first paper to explicitly use a welfare tradeoff approach (rather than a social discounting approach) to map human generosity. It demonstrates the construct validity of welfare tradeoff ratios by showing how they are correlated with a predicted personality variable (agreeableness), how they differ across relationship type, and even how they are modulated by a psychoactive drug (MDMA).

34. Delton AW: *A Psychological Calculus for Welfare Tradeoffs*. Santa Barbara: University of California; 2010:. (doctoral dissertation).

35. Locey ML, Jones BA, Rachlin H: **Real and hypothetical rewards in self-control and social discounting**. *Judgm Decis Mak* 2011, **6**:552-564.

36. Boyer P, Lienard P, Xu J: **Cultural differences in investing in others and in the future: why measuring trust is not enough**. *PLoS ONE* 2012, **7**:e40750.

37. • Hackman J, Danvers A, Hruschka DJ: **Closeness is enough for friends, but not mates or kin: mate and kinship premiums in India and U.S.**. *Evol Hum Behav* 2015, **36**:137-145.
This paper uses a monetary decision making task to study how multiple variables, related to kinship and the size of the resource at stake, are combined to determine welfare tradeoff ratios.

38. Strombach T, Jin J, Weber B, Kenning P, Shen Q, Ma Q, Kalenscher T: **Charity begins at home: cultural differences in social discounting and generosity**. *J Behav Decis Making* 2014, **27**:235-245.

39. Strombach T, Weber B, Hangebrauk Z, Kenning P, Karipidis II, Tobler PN, Kalenscher T: **Social discounting involves modulation of neural value signals by temporoparietal junction**. *PNAS* 2015, **112**:1619-1624.

40. Jones BA, Rachlin H: **Delay, probability, and social discounting in a public goods game**. *J Exp Anal Behav* 2009, **91**:61-73.

41. Rachlin H, Jones BA: **Social discounting and delay discounting**. *J Behav Decis Making* 2008, **21**:29-43.

42. • Safin V, Locey ML, Rachlin H: **Valuing rewards to others in a prisoner's dilemma game**. *Behav Processes* 2013, **99**:145-149.
An excellent recent example of the social discounting approach. This paper shows how social discounting can be used to explain the choices people make in one of the most well-studied experimental games, the prisoners' dilemma.

43. Locey ML, Safin V, Rachlin H: **Social discounting and the prisoner's dilemma game**. *J Exp Anal Behav* 2013, **99**:85-97.

44. Osiński JT, Ostaszewski P, Karbowski A: **Social discounting rate is negatively correlated with fluid intelligence**. *Pers Indiv Differ* 2014, **59**:44-49.

45. Bradstreet MP, Higgins ST, Heil SH, Badger GJ, Skelly JM, Lynch ME, Trayah MC: **Social discounting and cigarette smoking during pregnancy**. *J Behav Decis Making* 2012, **25**:502-511.

46. Sharp C, Barr G, Ross D, Bhimani R, Ha C, Vuchinich R: **Social discounting and externalizing behavior problems in boys**. *J Behav Decis Making* 2012, **25**:239-247.

47. Osiński J: **Kin altruism, reciprocal altruism and social discounting**. *Pers Indiv Differ* 2009, **47**:374-378.

48. Rachlin H, Jones BA: **Altruism among relatives and non-relatives**. *Behav Processes* 2008, **79**:120-123.

49. Osiński J: **Social discounting: the effect of outcome uncertainty**. *Behav Processes* 2010, **85**:24-27.

50. Xue M: **Altruism and reciprocity among friends and kin in a Tibetan village**. *Evol Hum Behav* 2013, **34**:323-329.

51. Stewart-Williams S: **Altruism among kin vs. nonkin: effects of cost of help and reciprocal exchange**. *Evol Hum Behav* 2007, **28**:193-198.

52. Ostaszewski P, Osiński JT: **Social discounting of monetary rewards: the effects of amount and social relationship**. *Eur Psychol* 2011, **16**:220-226.

53. Curry O, Roberts SGB, Dunbar RIM: **Altruism in social networks: evidence for a ''kinship premium.''**. *Br J Psychol* 2013, **104**:283-295.

54. •• McCullough ME, Kurzban R, Tabak BA: **Cognitive systems for revenge and forgiveness**. *Behav Brain Sci* 2013, **36**:1-58.
This paper gives a detailed analysis of anger and forgiveness within a welfare tradeoff framework. It reveals the utility of viewing emotions as computational systems.