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**

Edited by **Steven W Gangestad** and **Joshua M Tybur**

For a complete overview see the Issue and the Editorial

Available online 4th July 2015

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:

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

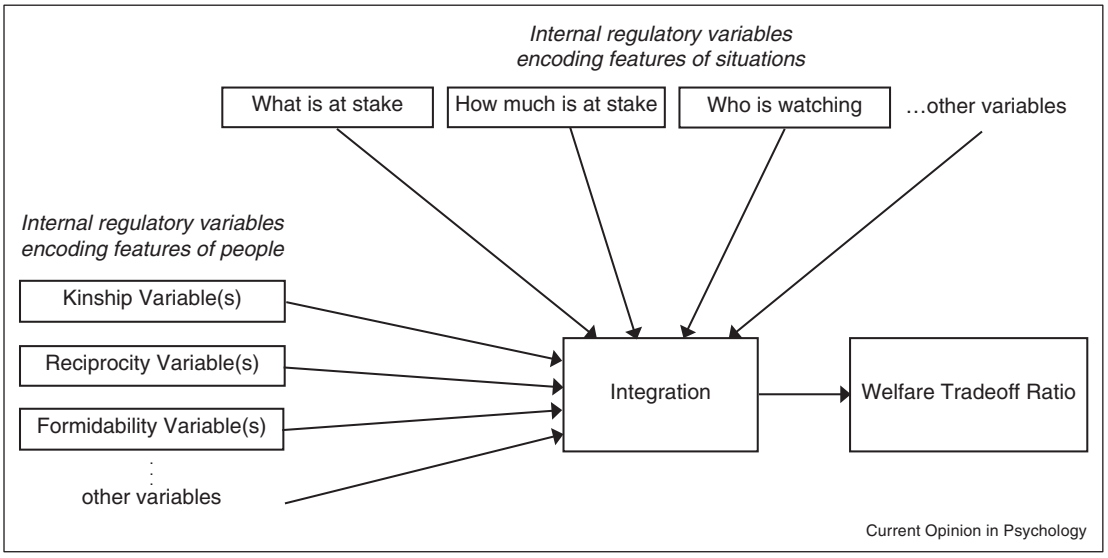
The greater the welfare tradeoff ratio, the more weight you place on the other person's welfare and the more likely you are to benefit them. Of course, the actual computations are likely to be more complicated, with future research yielding a more complex equation.

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.