

Part IV

Evaluative Processes

Tooby, J., Cosmides, L., Sell, A., Lieberman, D. & Sznycer, D. (2008).
Internal regulatory variables and the design of human motivation: A computational
and evolutionary approach. In Elliot, Andrew J. (2008). Handbook of approach and
avoidance motivation. (pp. 251-271). New York, NY, US: Psychology Press.



15 Internal Regulatory Variables and the Design of Human Motivation: A Computational and Evolutionary Approach

John Tooby, Leda Cosmides, Aaron Sell, Debra Lieberman, and Daniel Sznycer

CONTENTS

Where is the Value in Cognition?.....	249
Reasoning, Concepts, and Value	250
Internal Regulatory Variables and Motivation.....	250
Internal Regulatory Variables and Felt Experience	252
Discovering Internal Regulatory Variables: The Role of Theories of Adaptive Function	254
The Computational Architecture of Sibling Detection in Humans	255
Degree of Relatedness and Inbreeding Depression: Selection Pressures	255
Degree of Relatedness and Altruism: Selection Pressures	256
Making Welfare Trade-Offs	256
The Kinship Index as an Internal Regulatory Variable	258
Sexual Motivation System	258
Altruistic Motivation System.....	259
Triangulating the Kinship Index	259
Computing the Kinship Index for Siblings.....	259
Olders Detecting Younger Siblings	260
Youngers Detecting Older Sibs.....	260
Cue Integration by the Kinship Estimator.....	260
Anger as a Recalibrational Emotion	261
Raising Others' WTRs Toward You.....	261
Anger as a Negotiation over WTR Values	263
The Cooperation System in Anger Mode	263
The Aggression System in Anger Mode.....	264
Approach Motivations in Anger	266
Conclusion.....	266
Acknowledgments	267
References	267

WHERE IS THE VALUE IN COGNITION?

Cognitive psychology and evolutionary biology both offer compelling visions of what the brain does, but there is a radical disconnect between them. Evolutionary biology

tells us that we are beings whose brains evolved to pursue utilitarian goals, goals that promoted survival and reproduction in ancestral environments. Cognitive psychology tells us that our brains contain mechanisms of perception,

attention, memory, reasoning, and learning that, while not infallible, produce knowledge that tracks at least some aspects of reality. Fodor goes even further: he says that the function of cognition is “the fixation of true beliefs” (Fodor, 2000, p. 68). This sounds fine, until you realize the problem. True beliefs, by themselves, contain no information about how to behave—what to value, what to approach, what to avoid. Encyclopedias have no motivations. Value is not a property of the world, there to be discovered. Dung is not disgusting—it is disgusting to us, but attractive to dung flies. A man may be sexually attractive to many women, but sexually repulsive to his sister. Like beauty, value is in the adaptations of the beholder (Symons, 1995).

The question we have been addressing in our research is how to incorporate the study of value and motivation into the framework of the cognitive sciences. There have been two strands to these explorations. The first involves reasoning and concepts; the second is a new way of thinking about motivation that is both computational and grounded in evolutionary biology.

REASONING, CONCEPTS, AND VALUE

A common view in cognitive psychology, most forcefully articulated by Fodor (2000), is that reasoning and concepts are designed for acquiring true beliefs, so their study can be walled off from the study of motivation—from so-called conative mechanisms, governing preferences, approach, and avoidance. In direct contrast to this view, our earliest research on reasoning was motivated by results from evolutionary game theory, which models conative (i.e., behavior-producing) mechanisms. Agents endowed with different behavior-producing mechanisms are allowed to interact, and the payoffs of these interactions determine how many copies of each (mechanism-bearing) agent appear in the next generation. This simulates how natural selection works, allowing one to see which mechanism types, which designs, will be retained over time and which will be eliminated from the population. One can then test to see whether mechanisms with the expected design actually exist in the human mind.

One of the earliest applications of evolutionary game theory was to social exchange (reciprocation), modeled as a repeated prisoners’ dilemma (Trivers, 1971). These analyses showed that mechanisms causing social exchange cannot evolve in a species unless co-operators have mechanisms for detecting and avoiding cheaters—those who take benefits without ever providing them in return. Using standard reasoning tasks, we provided evidence for a functionally and neurally dissociable system that is

specialized for reasoning about social exchange, which includes a subroutine for detecting cheaters—a value-laden concept (for review, see Cosmides & Tooby, 2005).

AQ1

Note that the function of this system is not to discover something true about the world; instead, it is to produce adaptive behavior by feeding into systems that regulate approach and avoidance motivations. By monitoring information that can reveal who is a co-operator and who a cheater, it provides the information necessary for approaching co-operators and avoiding cheaters. It also structures the approach in a detailed way: The social exchange system has criteria defining the cost–benefit relations that should apply before initiating a social exchange relationship, and inferences regulating how it should proceed—who should do what for whom and when (Cosmides & Tooby, 1989, 2007).

The discovery of this system undercut the traditional idea of content-free reasoning mechanisms designed only for discovering properties of the world. Moreover, because the system operates over concepts like “benefit to agent,” “obligation,” “entitlement,” and “cheater,” it provided evidence against the view that the study of concepts can be walled off from the study of conative mechanisms. These are value-laden concepts, whose function is to regulate behavior. Brains evolved to produce adaptive behavior under ancestral conditions, so motivation, value, and cognition must have co-evolved. Consequently, we should expect the content and computational structure of many concepts to reflect the role they played in motivating behavior that promoted survival and reproduction among our hunter-gatherer ancestors (Tooby, Cosmides, & Barrett, 2005).

INTERNAL REGULATORY VARIABLES AND MOTIVATION

Systems that are domain-specialized for producing knowledge and inferences are necessary for motivating adaptive behavior, but they are not sufficient. Solving motivational problems requires computational elements that are not exactly concepts, beliefs, desires, preferences, or drives, but something else: internal regulatory variables and evolved specializations that compute them and deliver them to evolved decision-making systems.

To take a (seemingly) simple example, it is not enough to know that mongongo nuts belong to the category “food” and are therefore to be approached. Studies of the foraging behavior of living hunter-gatherers show that the decision to look for and pick up any given food resource is based on complex calculations that combine several variables (Smith & Winterhalder, 1992; Winterhalder &

Smith, 2000). These variables include (at minimum) the calories per gram of each food resource, its average package size (grams per unit caught or gathered), its average search time (how long it takes to find it), and its average handling time (how long it takes to capture it and convert it into edible form—cracking the nuts, butchering the animal, cooking it, and so on). Models using all four variables predict more variance in what foragers actually look for and take than ones based on caloric value alone. These models predict *foraging motivations*—which foods people actively search for when they go out foraging, which foods they do not bother with even when they come across them, and which they decide are worth the effort of capturing/extracting/gathering and schlepping back to camp.

These mathematical models have interesting implications for the computational architecture of the motivational systems that regulate approach and avoidance while foraging. That they successfully predict behavior implies that the brain has programs that compute, for each food resource, the value of these four variables (or of proxy variables correlated with them). Each computed value has a magnitude that represents, respectively, how calorie rich, how big, how difficult to find, and how difficult to obtain and prepare each food resource is. A different constellation of these four values will be computed for each food resource, and the constellation applying to a given animal or plant needs to be stored and retrieved in tandem when deciding whether to forage for it. This means the values that apply to mongongo nuts need to be stored in a separate mental file folder from those that apply to palm fruit, giraffe meat, and other foods, using a template or “data organizer” designed for this purpose (Boyer, 2001; Sperber, 1994), equipped with at least four “slots,” each specialized for registering and storing one of the four variables. When it is time to forage, the values of these internal regulatory variables are accessed by a computational system that combines them, producing motivations expressed in behavior. As a result, we observe foragers seeking calorie rich foods with low search and handling times more than calorie rich foods with high search and handling times, and so on for other combinations of the four variables, according to the algorithm by which the motivational system combines them. (Obesity in industrialized societies is a predictable consequence of this system operating outside of its ancestral context: it produces stronger motivations for seeking quick, easy, calorie dense foods than slow, laborious, low calorie ones.)

Because foraging motivations are regulated by the values of these four variables, they can be thought of as internal regulatory variables. Internal regulatory variables are not an exotic feature of the human mind or of

motivational systems per se; they are key features of every feedback-regulated process in multicellular organisms. Exquisitely designed regulatory systems permeate the human body, producing functional outcomes by entraining processes at all levels of organization, from gene activation and protein synthesis to organ function to behavior. Motivational systems are just one class of regulatory system. They differ from regulatory systems like the Krebs cycle primarily in that their goal—their adaptive function, the problem they were organized by natural selection to solve—is to regulate behavior rather than metabolism. Even this divide is not sharp—many metabolic regulatory systems require behavior-regulating motivational systems (e.g., glucose delivery and hunger, electrolyte balance and thirst), and many motivational systems cannot do their job without regulating metabolism as well as behavior (e.g., predator evasion and the flight–fight response).

Our working hypothesis is that motivational systems, like other regulatory systems, are interpenetrated by networks of internal regulatory variables. This is known to be true for the motivational systems regulating fluid balance (for thirst), energy reserves (for hunger), body temperature (for thermoregulation), and carbon dioxide levels (for breathing). We think it is also true for motivational systems regulating social interaction. Just as there are internal regulatory variables that register the caloric value of a food resource or the level of glucose in the blood, there should be internal regulatory variables that register the value to you of various social partners and their actions: how much a particular person is willing to sacrifice his or her own welfare for yours (a welfare trade-off ratio), how valuable a particular person would be to you as a sexual partner (a sexual value index), how much harm a person could inflict on you in a fight (a formidability index), how closely related a person is to you (a kinship index), and so on.

According to this view, internal regulatory variables evolved to track those narrow, targeted properties of the body, the social environment, and the physical environment whose computation provided the necessary inputs to evolved decision-making systems, generating motivations to action. Internal regulatory variables have magnitudes, and they either express value or provide input to mechanisms that compute value.

It may seem odd, even bloodless, to talk about a computational approach to motivation, where behavior is regulated by internal variables. After all, every one of us has felt the pushes and pulls of motivation—the impulse to help a friend, to yell at a bully, to discharge an obligation, to express gratitude for an unexpected act of kindness at

a time of need. We all have phenomenal experiences, and their existence raises many interesting and unsolved philosophical puzzles (Dennett, 1988; Tye, 2003). But the success of vision science shows that scientific progress can be made nevertheless, by investigating the computational processes that *generate* experiences.

In discussing the relationship between computation and conscious experience, Jackendoff (1987) points out that differences in perceived color, the experience of yellow versus blue, can be thought as a *data format* by which the mind represents differences in the reflectant properties of surfaces. The computed products of lower level visual processing are represented in data formats that can be accessed only by mechanisms within the visual system; in contrast, the data format we experience as color can be accessed by a wide variety of behavior-regulating systems. We suspect a similar view of felt experiences will emerge from a computational approach to motivation. Some felt experiences may be a data format by which the mind represents, in a broadly accessible way, the magnitude of certain internal regulatory variables. In other cases, a felt experience may be the output of a motivational system, with its felt intensity regulated by the magnitude of the internal regulatory variables it accesses while performing its computations. That is, differences in the magnitudes of these variables cause increases or decreases in your impulse to help or harm, your feelings of sexual attraction, disgust, gratitude, guilt, shame, obligation, pride, entitlement, and so on. Representing the conclusions of a motivational system in the broadly accessible data format of felt experience may be one key to the human ability to improvise novel solutions to adaptive problems (Cosmides & Tooby, 2000a, 2001). Rather than being limited by fixed response patterns, improvised responses to imagined situations can be evaluated by how they change the intensity of these felt experiences—an internal feedback system that steers behavioral responses toward adaptive outcomes.

Felt experience is so central to folk theories of motivation that it can blind us to the need for computational accounts of motivation, just as the immediacy of perceptual experience blinded vision scientists of the 1960s to the need for computational accounts of vision (Marr, 1982). So before turning to social motivation, we would like to pause, briefly, to consider the ways in which felt experience may be related to internal regulatory variables and computation.

INTERNAL REGULATORY VARIABLES AND FELT EXPERIENCE

Sometimes the operation of internal regulatory variables is entirely nonconscious. For example, the kidneys are

equipped with an internal regulatory variable that registers levels of oxygen in the blood. When blood oxygen falls below a certain threshold value, this stimulates the production of erythropoietin, a hormone that triggers maturation of red blood cells in the bone marrow. This goes on all the time, unaccompanied by any felt experience—the mind does not seem to have any design feature capable of representing levels of erythropoietin or blood oxygen. Blood oxygen level is not represented as a felt experience even when it is dangerously low: Only the *consequences* of hypoxia, as it damages organ systems, are felt, causing headache, nausea, breathlessness, and other aversive experiences.

In contrast, some motivational systems are designed to produce felt experiences as a result of having processed an internal regulatory variable, and those felt experiences guide behavior in a direct and adaptive fashion. The suffocation alarm system is a familiar example. There is an internal regulatory variable that registers carbon dioxide to oxygen levels in circulation. When this ratio increases too quickly, the suffocation alarm system is triggered. It downregulates motivations to pursue ongoing activities (e.g., we stop reading under the covers), upregulates motivations to change position, and produces the felt experience of suffocation. That felt experience guides our movements: We change position, sometimes frantically, following any experienced decline in the sense of suffocation until the awful felt experience ceases entirely—which happens when the regulatory variable reaches a normal level again. Ondine's curse, a disorder of the CO₂/O₂ regulatory variable and its ability to trigger the alarm system, is usually fatal: children born with this disorder suffocate in their sleep.

The felt experience of suffocation could be considered a readout of the magnitude of the CO₂/O₂ regulatory variable—a data format that allows movement programs to access changes in its value on a second-by-second basis, until its value falls below threshold again. That is, changes in the *intensity* of a given felt experience can be thought of as a special data format, one that makes changes in the *magnitude* of an internal regulatory variable accessible to a broad array of behavior-regulating mechanisms.

Differences between stimuli in key properties—fat content of foods, for example—should produce different values for the regulatory variable associated with each stimulus; the magnitude of these values can, in turn, be represented as different intensities of felt experience. A chocolate truffle generates a more intense felt experience of richness than a celery stick, whether you are eating them or just imagining eating them, and that intensity reflects their relative fat content. That these felt experiences

can be generated by imagination alone suggests that values for an internal regulatory variable registering the fat content of each were previously stored; imagining, seeing, and eating them initiates a process that transforms their magnitudes into a data format of felt experience.

Tracking different properties of the world—fat content versus handling time, for example—clearly requires distinct regulatory variables. But if felt experience is functional—allowing simulation-based planning, for example—then the data formats whereby distinct variables are experienced need to be qualitatively different from one another. Distinct variables need to be associated with distinct qualia, to use the philosophers' term (Tye, 2003). Accordingly, the felt experience of richness is qualitatively distinct from the felt experience of effort—or of anticipated effort, for that matter. Watching an ice cream commercial in the kitchen can activate the felt intensity of richness associated with ice cream, exerting a motivational pull. But this pull can be trumped by the (quite different) felt experience of anticipated effort that arises as you imagine trekking across the town to get it, especially when you are already tired. Algorithms in the foraging motivation system combine the magnitudes of both variables (caloric value and anticipated effort) and others as well; you experience the output of these algorithms as a motivation to action—either to go for the ice cream or just stay home.

An internal regulatory variable may have no associated felt experience, yet increase or decrease the felt experiences produced by various motivation systems. An example we will discuss later is the kinship index, a regulatory variable whose magnitude represents an individual's probable degree of genetic relatedness to oneself—roughly, whether an individual is a full, half, or step sibling (Lieberman, Cosmides, & Tooby, 2007). There does not seem to be a felt experience uniquely associated with its value. But the magnitude of the kinship index up- and downregulates distinct types of felt experiences. A high value produces feelings of disgust when accessed by the sexual motivation system, and impulses to help when accessed by the system regulating altruistic motivations.

Obviously the value of an internal variable can be stored without being transformed into a felt experience, just as episodes from one's life can be stored without being transformed into a remembered experience of the past—a transformation that requires the operation of particular computations at retrieval (Klein, German, Cosmides, & Gabriel, 2004). In many cases, especially those requiring fast action, the computational systems that produce motivations may be able to access the values of internal regulatory variables without their having first

been processed and reformatted as a felt experience. Indeed, there should be principles of good design determining when stored values and summary conclusions are accessed directly rather than being first transformed into felt experiences (Klein, Cosmides, Tooby, & Chance, 2002). For example, if foraging algorithms have repeatedly registered a particular food as calorie poor, hard to find, and difficult to prepare, and repeatedly performed calculations on those variables, the motivational implications for action—*don't bother with food X*—might simply be stored as a summary conclusion and quickly retrieved, without any accompanying affect.

Transforming the magnitude of regulatory variables into felt experience may be necessary, however, when we are faced with a choice but have no precomputed summary conclusion. It may also be necessary when the computations of two or more regulatory systems produce motivations to action that are in direct conflict with one another. Indeed, this last case may be when it is most important to make the information stored in regulatory variables available to a broad array of mechanisms through felt experience. Imagining stimuli and situations in a quasi-perceptual way can activate felt experiences, ones reflecting the magnitude of stored regulatory variables and ones reflecting the output of the motivational systems these variables feed (Cosmides & Tooby, 2000b; Tooby & Cosmides, 1990). But it does so in a way that is decoupled from action—a design feature that allows us to simulate how we would feel about the outcomes of actions, which is pivotal for choosing between alternative courses of action and planning for the future (Cosmides & Tooby, 2000a; Tooby & Cosmides, 2001). Seen in this way, the ability to transform the magnitudes of internal regulatory variables and their motivational outputs into felt experience is a crucial facet not just of improvisational intelligence, but of human foresight and choice, allowing us to not only simulate what would happen, but how we would feel about what would happen.

Our point is this: There should be principled relationships between internal regulatory variables and felt experience. The fact that we experience ourselves as motivated by feelings and impulses does not render a computational account of motivation unnecessary, any more than our experience of seeing the world renders a computational account of vision unnecessary. The existence of felt experiences linked to motivation tells us that our brain is doing something interesting. Given that the brain is an organ of computation—that is, an evolved system whose function is to process information—it is worth trying to explain that something in computational (i.e., information-processing) terms.

DISCOVERING INTERNAL REGULATORY VARIABLES: THE ROLE OF THEORIES OF ADAPTIVE FUNCTION

Discovering the internal variables that govern the operation of a regulatory system is difficult, but progress can be made when one's research is informed by a theory of what problem the system evolved to solve. This is not too difficult when the problem is metabolic, such as glucose delivery, oxygen delivery, or maintaining electrolyte balance, because even a naïve theory can help in these cases (though not infallibly; see Nesse & Williams, 1995). But what is social interaction for? What problems of survival and reproduction do individuals face when they live socially, and what behavioral responses count as adaptive solutions to these problems?

The answers to these questions are not intuitive. Indeed, as recently as 40 years ago many biologists were on the wrong track, defining an animal's behavior as adaptive if it was thought to promote the survival of that species (Williams, 1966). This "greater goodism" was applied even to antagonistic interactions: Lorenz famously, and wrongly, thought that ritualized dominance displays evolved because it was better for the group if weaker males stepped aside so that only the strongest males would breed (Lorenz, 1966). If we are to discover internal regulatory variables that govern social motivations, we need to properly understand the adaptive problems of social life that these variables evolved to solve.

Over the last 40 years, evolutionary biologists and behavioral ecologists have carefully analyzed how natural selection shapes the social interactions of many species, discovering how the evolved outcomes vary as a function of different ecological and life history parameters. As a result, they have developed formal models defining a series of specific adaptive problems arising from social life, and what a behavioral solution that is evolutionarily stable might look like. There are many examples. Kin selection theory analyzes resource division within the family, encompassing topics such as when to help siblings, and when siblings will be in conflict with their parents and each other over how parents allocate investments of time, effort, and resources among them (Hamilton, 1964; Trivers, 1974). Analyses of the selection pressures posed by pathogens and the existence of deleterious recessive genes lead to predictions about motivational systems regulating inbreeding avoidance (Lieberman et al., 2007; Tooby, 1982). Parental investment and sexual selection theory elucidates adaptive problems of courtship and mating (Buss & Schmidt, 1993; Daly & Wilson, 1983; Symons, 1979; Trivers,

1972). The "asymmetric war of attrition" is a game theoretic model of the selection pressures shaping dominance hierarchies and resource division—showing, *pace* Lorenz, that ritualized dominance displays provide individually beneficial information about when it is better to cede a resource than fight for it (Hammerstein & Parker, 1982; Huntingford & Turner, 1987). The banker's paradox model of deep engagement relationships (Tooby & Cosmides, 1996) and risk-buffering models of sharing (Gurven, Allen-Arave, Hill, & Hurtado, 2000) express selection pressures relevant to the nurturing of friendships and within-group sharing. Reciprocal altruism theory illuminates selection pressures shaping two-person exchange (Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971), and models of the evolution of n-person cooperation illuminate the problems that must be solved for coalitional alliances and group cooperation to be evolutionarily stable (Boyd & Richerson, 1992; Tooby, Cosmides, & Price, 2006).

As psychologists, we have found it useful to start with a task analysis of the adaptive problems defined by these models, to see what properties a computational system capable of solving them would need. When we did this, it became clear that the computational systems that produce social motivations would need internal regulatory variables that track the properties and actions of other people that make them valuable (or dangerous) as social interactants. But this poses an interesting problem for systems regulating approach and avoidance motivations. For certain stimuli, the value of an internal regulatory variable can be computed in a way that takes no account of the properties of the individual doing the computing: The number of calories per gram of mongongo nuts is the same, regardless of who will be eating them. In contrast, the value of a person as a social partner sensitively depends on the circumstances and properties of the valuer. For example, if you and I are both looking for a sexual partner, the fact that the attractive person walking by is my sibling is a deal breaker for me, but not for you; on the other hand, if we are both sick and need care, that same sibling is more valuable to me than to you as someone to spend time with.

In other words, a social partner cannot have an invariant value that makes them a stimulus eliciting approach or avoidance; their value depends on who they are interacting with and what type of interaction is at issue. For this reason, there should be programs that compute and represent the magnitude of each internal regulatory variable in a way that is indexed to the self: Person X's value as a sexual partner to me, their kinship to me, their aggressive formidability relative to mine, their status

relative to mine, their value as a cooperative partner to me, how much of their own welfare they are willing to sacrifice to enhance my welfare. We will illustrate how this works first with kin detection, and then with the motivational system that produces anger.

THE COMPUTATIONAL ARCHITECTURE OF SIBLING DETECTION IN HUMANS

Oysters never know their siblings. Their parents release millions of gametes into the sea, most of which are eaten. Only a few survive to adulthood, and these siblings are so dispersed that they are unlikely to ever meet, let alone interact. The ecology of many species causes siblings to disperse so widely that they never interact as adults, and siblings in species lacking parental care typically do not associate as juveniles either. Humans, however, could not be more different. Hunter-gatherer children grow up in a family of parents and siblings, and live in bands that often include grandparents, uncles, aunts, and cousins. The uncles, aunts, and cousins are there because human siblings also associate as adults—like most people in traditional societies, adult hunter-gatherers prefer to live with relatives nearby, if that is an option.

That close genetic relatives frequently interact is an important fact about our species. Some of the best established models in evolutionary biology show that genetic relatedness is an important factor in the social evolution of such species (Hamilton, 1964; Williams & Williams, 1957). But when evolutionary biologists talk about genetic relatedness, they are not talking about the total proportion of genes two individuals share in common—after all, we humans are genetically identical at 66% of our loci, as measured by protein coding genes (and at over 99% of loci, as measured by base pairs).

By genetic relatedness, they mean something much stranger. They are referring to the probability that two individuals, i and j , share a *particular gene* by inheriting it from the same parent, grandparent, or other recent common ancestor. This probability is called their *degree of relatedness*, and is expressed as a number, r_{ij} , that can range from zero to one. Assuming we are talking about a diploid species and a gene from the nucleus that segregates by Mendel's law, the degree of relatedness is $\frac{1}{2}$ for full siblings and for parents and offspring, $\frac{1}{4}$ for half siblings and for aunts/uncles and their nieces/nephews, $\frac{1}{8}$ for first cousins, and $\frac{1}{32}$ for second cousins.

It turns out that two different social motivation systems require an internal regulatory variable that tracks this strange statistic: one governing sexual attraction or

revulsion, the other governing altruism. We first explain the selection pressures, then turn to the computational model of social motivations that these selection pressures led us to propose and test.

DEGREE OF RELATEDNESS AND INBREEDING DEPRESSION: SELECTION PRESSURES

Animals are highly organized systems (hence “organisms”), whose functioning can easily be disordered by random changes. Mutations are random events, and they occur all the time. Many of them disrupt the functioning of our tightly engineered regulatory systems. A single mutation can, for example, prevent a gene from being read (or, if read, from producing the right protein). Given that our chromosomes come in pairs, a nonfunctional gene like this need not be a problem: If it is found on only one chromosome of the pair, the other chromosome will produce the right protein and all may be well. But if the same mutation is found on *both* chromosomes, the right protein will not be produced by either. This is a problem. Not producing the right protein can impair a developing organism's ability to function or even kill it.

Such genes, called *deleterious recessives*, are not rare; they can accumulate in genomes precisely because they are not harmful when heterozygous—that is, when you have only one copy. Their harmful effects are unmasked, however, when you have two copies—that is, when they are homozygous. A *lethal equivalent* is a gene whose effects, when unmasked by homozygosity, are lethal to its bearer before he or she reaches reproductive age (or a set of genes whose effects summed together are lethal when unmasked). It is estimated that each of us has 1–2 lethal equivalents worth of deleterious recessive genes (Bittles & Neel, 1994; some estimates are higher: see Lieberman, 2003 for review).

AQ2

Fortunately, the deleterious recessives found in one person are usually different from those found in another. This is because they are the products of *random* mutations. So mating with an unrelated person is very unlikely to result in an offspring with two copies of a particular deleterious recessive: The probability is only 1 per million for a recessive whose prevalence in the population is 1 per 1000.

Now consider a deleterious recessive gene in you. This mutation could have happened after you were born, or you could have inherited it from one of your parents. The probability that that particular recessive gene is also in your full sibling is half ($\frac{1}{2}$)—your degree of relatedness. If you were to mate with your sibling, each of you would have a $\frac{1}{2}$ chance of passing that particular gene to

the resulting offspring, resulting in a probability that your offspring will have it of $1/8$ ($[r = 1/2] \times [1/2 \text{ chance mother passed it on}] \times [1/2 \text{ chance father passed it on}]$). If you have two lethal recessives, then the probability that an offspring conceived by you and your full sibling will be homozygous for one (or more) of them and therefore die before reaching reproductive age, is $2 \times 1/8 = 25\%$ —a huge selection pressure. (The probabilities drop to 6.25% for first cousin marriage, and 1.6% for second cousin marriage.)

The decline in the fitness of offspring (in their viability and consequent reproductive rate) resulting from matings between close genetic relatives is called *inbreeding depression*. Inbreeding may result in zygotes that die shortly after conception, infants that die after birth, or children that suffer from congenital abnormalities. First degree relatives ($r = 1/2$) rarely mate because we have mechanisms that cause us to avoid inbreeding. But it sometimes happens, and in one study it was possible to compare children fathered by first degree relatives (brothers and fathers) to children of the same women who were fathered by unrelated men. The rate of death, severe mental handicap, and congenital disorders was 54% in the children of first degree relatives, compared to 8.7% in the children born of nonincestuous matings (Seemanova, 1971; see also Adams & Neel, 1967). Notice that even this is an underestimate of inbreeding depression, as it does not take into account offspring loss through miscarriage.

Because inbreeding makes children more similar to their parents, it also defeats the function of sexual reproduction, which is to produce genetic diversity that protects offspring against pathogens that have adapted to the parents' phenotype (Tooby, 1982). Both selection pressures, deleterious recessives and pathogen-driven selection for genetic diversity, have the same reproductive consequence: Individuals who avoid mating with close relatives will leave more descendants than those whose mating decisions are unaffected by relatedness. This means that mutations that tweak motivational systems in ways that cost-effectively reduce inbreeding depression will be favored by natural selection. In some species, this is accomplished by computational systems that motivate members of one sex or the other to disperse far from their natal group before they reach sexual maturity. But for species in which individuals can help one another, there are advantages to staying near your parents and siblings, even in adulthood (see below). When this is true, a good way to reduce the chances of inbreeding is by tweaking the sexual attraction system. People that others find sexually attractive should excite no sexual interest in you if they are your siblings or parents. Indeed, the prospect of

sex with a sibling or parent should elicit sexual disgust or revulsion—an avoidance motivation.

DEGREE OF RELATEDNESS AND ALTRUISM: SELECTION PRESSURES

In species that live socially, conflicts of interest are ubiquitous. If I use a resource, you cannot; if I see a predator and warn you, allowing you to escape, you will benefit but the predator's attention will be drawn to me; if you successfully court an attractive person, that person becomes unavailable to me. That is, situations frequently arise in which you can take an action that will benefit you, but impose a cost on me; equally, there will be situations in which you do something that will benefit me, but at some cost to yourself. To what extent should your decisions take my welfare into account, and vice versa? When should you trade off some of your welfare to enhance mine? The answer depends, in part, on our degree of relatedness.

Making Welfare Trade-Offs

To capture this notion of a trade-off, let us define a variable: a *welfare trade-off ratio* or $WTR_{actor,j}$. By hypothesis, this is an internal regulatory variable signifying how much weight an individual actor places on j 's welfare relative to the actor's own. What we want to know is how natural selection will set the value of this variable. Equations 15.1 and 15.2 express decision rules for situations in which one's interests conflict with those of individual j . They are generalizations of very standard formulas in evolutionary biology, in which benefits and costs (welfare) are defined as increases and decreases in an individual's reproductive rate. (Evolutionary models assume that humans, like other animals, have mechanisms for reckoning the benefits and costs of actions to self and others, and that these evolved because they reflect the average reproductive consequences of choices in our ancestral past.)

Given the possibility of taking an action, \underline{A} , that benefits one's self while imposing a cost on individual j , take beneficial action \underline{A} when Equation 15.1 is satisfied, but not otherwise:

$$B_{self} > (WTR_{self,j})(C_j), \text{ that is, when } B_{self}/C_j > WTR_{self,j}. \quad (15.1)$$

Given the possibility of taking an action, \underline{A} , that benefits j at some cost to the self, take costly action \underline{A} when Equation 15.2 is satisfied, but not otherwise:

$$C_{self} < (WTR_{self,j})(B_j), \text{ that is, when } C_{self}/B_j < WTR_{self,j}. \quad (15.2)$$

If $WTR_{self,j} = 0$, that means you place no weight on j 's welfare: Equation 15.1 means you will take self-beneficial actions no matter how large a cost they impose on j , and Equation 15.2 means you will never incur a cost to benefit j . If $WTR_{self,j} = 1$, that means you are as concerned with j 's welfare as your own: You will not take a beneficial action unless the cost it imposes on j is less than the benefit you gain (Equation 15.1), and you will help j whenever the cost to you is smaller than the benefit j gains (Equation 15.2).

So what WTR value will natural selection favor? That depends on many factors, some of which are important to our discussion of anger later in this chapter. For example, if j is a trustworthy cooperative partner who reciprocates favors often, then selection might favor a WTR toward j close to 1—that is one way of expressing the insights of reciprocal altruism theory (Trivers, 1972). If you have no cooperative relationship, then your WTR toward j may be set by your relative ability to harm one another: If you and j both value a resource equally, but j can easily injure you in a fight, then you will be better off ceding the resource to j than engaging in a fight you will surely lose. This is the insight behind the *asymmetric war of attrition* (Hammerstein & Parker, 1982), a game theoretic model that explains (contra Lorenz) why animals in many species engage in ritualized displays of their ability to harm one another, and why they settle on stable dominance hierarchies in which low ranking individuals cede resources to higher ranking ones without a fight (Huntingford & Turner, 1987). One way of expressing this is that your WTR toward j will be a function, at least in part, of your relative ability to injure one another—lower when you are the better fighter, higher when j is the better fighter.

The insight of kin selection theory is that your WTR toward j should be a function, at least in part, of your degree of relatedness to j (Hamilton, 1964; Williams & Williams, 1957). To make the insight clearer, let's leave aside factors such as reciprocation and the ability to cause injury, and consider two alternative designs. The first design sets $WTR_{self,j} = 0$, even when j is a genetic relative. The second design is a recent mutation in the population, which sets $WTR_{self,j} = r_{self,j}$, the self's degree of relatedness to j . Which WTR setting will spread by natural selection?

It turns out that the second design has a competitive advantage for a very interesting reason: If you inherited this new mutant design from your parents or grandparents, $r_{self,j}$ expresses the probability that your genetic relative also inherited that *same* mutation from the same common ancestors. This means the new design can promote *its own* reproduction by making trade-offs between your reproduction and the reproduction of your close

relatives—trade-offs reflecting the probability that your close relatives also have this new design.

When $WTR_{self,j} = r_{self,j}$, then Equation 15.2 reduces to Hamilton's famous rule: help j , but only when $C_{self} < (r_{self,j})(B_j)$, that is, when the costs to your own reproduction are outweighed by the benefits to j 's reproduction, discounted by the probability, $r_{self,j}$, that j has inherited the same mutant design from a recent common ancestor. The new mutation will also refrain from self-beneficial actions that are too costly to the reproduction of relatives: It will not take actions where $B_{self} < (r_{self,j})(C_j)$, Equation 15.1. These choices promote the replication of the design itself, by sometimes sacrificing your reproduction to enhance that of your genetic relatives. (As with deleterious recessives, you can see that whether this new mutation spreads is a function of the probability that that very same mutation is present in the genetic relative—not the total proportion of genes held in common).

In comparison, the design that sets $WTR_{self,j}$ equal to zero is at a competitive disadvantage. An actor equipped with a $WTR_{self,j} = 0$ design will take self-beneficial actions, even when the benefit to the actor's own reproduction is tiny and the cost to a relative's reproduction is huge. This means it indiscriminately imposes costs on the reproduction of relatives, who carry the same design with a probability equal to $r_{self,j}$. The design also loses opportunities to replicate itself by failing to take any action that is individually costly—even those that would provide a huge benefit to the reproduction of a relative at a tiny cost to the self.

Hamilton's rule does not mean that $WTR_{self,j}$ (henceforth: WTR_j) should never be higher than $r_{self,j}$ —your full sib might also be a great reciprocation partner, or strong enough to pound you into submission. It means that your degree of relatedness to j will be a lower boundary on WTR_j , causing you to help in accordance with Hamilton's rule even when there is no chance the favor will be reciprocated and no chance of violent reprisal. It also means that your sibling's willingness to pound you in submission will be tempered by your degree of relatedness: Equation 15.1—refrain unless $B_{sib}/C_{you} > 1/2$ —says that your full sibling will be unwilling to take self-beneficial actions at your expense unless he or she expects to realize a benefit at least twice as large as the cost imposed on you.

This analysis implies that the human mind should contain interesting mechanisms for assessing how high one's welfare trade-off ratio toward other individuals should be set, ones that take many variables into account, such as aggressive formidability or value as a cooperative partner. These might be converted to a WTR between self and j that is low or high. But whatever that magnitude

is, kin selection theory tells us that it should be upregulated for close genetic relatives, motivating us to help kin more and harm them less than we otherwise would.

THE KINSHIP INDEX AS AN INTERNAL REGULATORY VARIABLE

What might a computational approach to social motivation look like—what kind of internal regulatory variables are needed, and how they might regulate each other and behavior? The selection pressures just discussed suggested a number of hypotheses about the design of motivational systems. Our research has been testing the model shown in Figure 15.1. The key internal regulatory variables in this model are a sexual value index (SV_j), a welfare trade-off ratio (WTR_j) and, most importantly, a kinship index (KI_j).

The importance of degree of relatedness for inbreeding avoidance and altruism led us to expect that the human brain reliably develops a kin detection system. For each familiar individual j , this neurocomputational system would need to compute and update a continuous variable, the *kinship index*, KI_j . KI_j is an internal regulatory variable whose magnitude reflects the kin detection system's pairwise estimate of the degree of relatedness between self and j . The kinship index should serve as input to at least two different motivational systems: One regulating feelings of sexual attraction or revulsion, and

another regulating altruistic impulses. Each has its own proprietary regulatory variables.

Sexual Motivation System

Proprietary to the system-motivating sexual attraction is the sexual value index, SV_j . SV_j is a regulatory variable whose magnitude reflects j 's value as a sexual partner for the self (note that value as a sexual partner is not equivalent to value as a long-term mate). The sexual value estimator is a system designed to compute SV_j s based on many inputs, including cues that were correlated with fertility and health among our hunter-gatherer ancestors (for review, see Sugiyama, 2005). The kinship index associated with j is one of the variables that the sexual value estimator uses. When the magnitude of $KI_j = 0$, the magnitude of SV_j should be a function of all the other cues the sexual value estimator takes as input. But when the magnitude of KI_j is high, this should decrease the magnitude of SV_j dramatically. That is, the sexual value estimator's internal algorithms should be designed to weight a high KI_j more heavily than other inputs.

Cues, real or internally generated through imagination, signaling the possibility of sexual contact with j should activate the sexual motivation system. When this happens, the value of SV_j should be transformed into the felt experience of sexual attraction; a low value of SV_j should be transformed into the felt experience of sexual

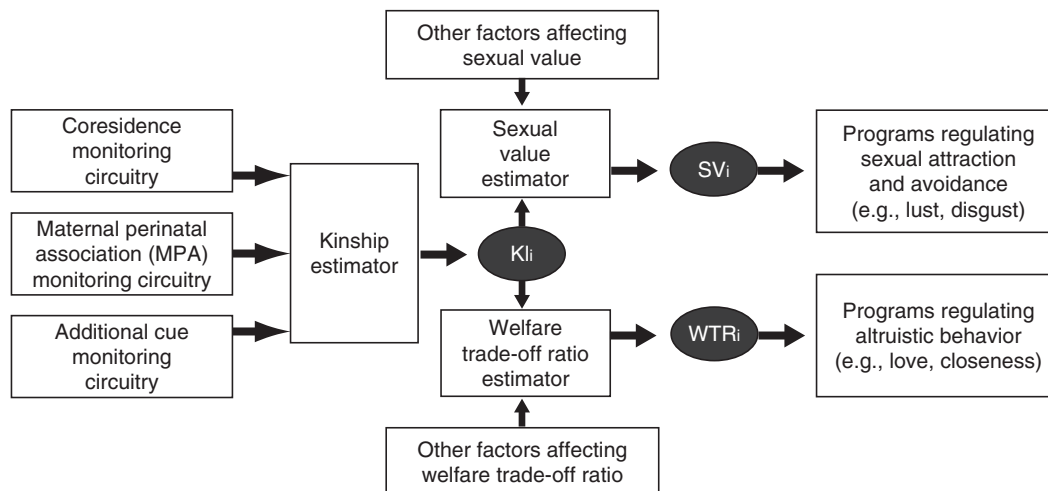


FIGURE 15.1 Model of the human kin detection system, and the internal regulatory variables (black ovals) it computes and regulates. Monitoring circuitry registers cues ancestrally correlated with genetic relatedness (e.g., coresidence duration, MPA). A *kinship estimator* transforms these inputs into a kinship index (KI_i) for each familiar individual i . The kinship index is used by downstream systems to compute two other regulatory variables: A sexual value index (SV_i) and a welfare trade-off ratio index (WTR_i). These serve as input to two motivational systems, one that regulates the allocation of mating effort and another that regulates altruism.

disgust. There does not seem to be a felt experience associated with KI_j per se, only with the variables it informs.

Altruistic Motivation System

According to the model in Figure 15.1, the welfare trade-off ratio, WTR_j , is an internal regulatory variable expressing how much you value j 's welfare relative to your own. Its value is nonconsciously expressed in many decisions you make throughout the day—how many cookies you leave for j , how loud to play your music when j is trying to work, whether to clean up the mess or leave it for j , whether to call home to let j know you will be late. It is computed by a system, let us call it the welfare trade-off ratio estimator, that takes into account many variables (cooperation, formidability, etc.), as discussed above. KI_j should be one of these variables: Higher magnitudes of KI_j should result in higher computed magnitudes for WTR_j . Conflicts of interest, both real and simulated through imagination, should activate decision rules that implement Equations 15.1 and 15.2 (above). The output of these decision rules can be represented in the data format of a felt experience—the impulse to help j (Equation 15.2) or to avoid harming j (Equation 15.1). When events trigger a recomputation of WTR_j , setting it at a higher or lower value, the newly recomputed value of WTR_j may itself be transformed, at least temporarily, into the data format of a felt experience: an increase or decrease in a feeling of warmth, love, or caring toward j . The felt experience makes the new WTR_j value broadly accessible, allowing many mechanisms to recalibrate the extent to which they take j 's welfare into account. (We discuss recalibrational emotions more in the next section.)

Triangulating the Kinship Index

That a kinship index should regulate two independent systems, altruism and sexual aversion, provides a method for determining which cues the kin detection system uses to compute the kinship index. If a computational element corresponding to KI_j exists, then any input to the kin detection system that increases the magnitude of KI_j should have two independent but coordinated effects: it should increase WTR_j and decrease SV_j . When asked to imagine the right activating situations, the magnitudes of these regulatory variables should be transformed into intensities of felt experience: A low SV_j should be represented as a high felt intensity of disgust at the thought of sex with j , and a high WTR_j should produce stronger impulses to help j than a lower WTR_j . This leads to a very surprising prediction: Inputs to the kin detection system that regulate feelings of altruism toward j should also regulate degree of sexual aversion toward j .

By triangulation, then, one can infer which cues serve as input to the kin detection system. People vary in their exposure to potential kinship cues, so variation in exposure to specific cues for a given sibling can be quantitatively matched to variation in the subject's feelings of sexual aversion and altruism toward that siblings. If a cue is used in computing the kinship index, then it should regulate sexual aversion and altruism toward j , and the pattern of cue use should be the same for both motivational systems. Using this logic, we were able to discover how the kinship estimator combines cues to compute a kinship index for siblings. Methods and details of the results we discuss below can be found in Lieberman et al. (2007).

COMPUTING THE KINSHIP INDEX FOR SIBLINGS

Detecting genetic relatedness is a major adaptive problem, but not an easy one to solve. Neither we nor our ancestors can see another person's DNA directly and compare it to our own, in order to determine genetic relatedness. So the best evolution can do is to design a kin detection system that uses cues that were reliably correlated with genetic relatedness in the ancestral past to compute the magnitude of a kinship index. This requires *monitoring circuitry*, which is designed to register cues to relatedness. It also requires a computational unit, the *kinship estimator*, whose procedures were tuned by a history of selection to take these registered inputs and transform them into a kinship index. The question is, what cues does the monitoring circuitry register, and how does the kinship estimator transform these into a kinship index?

Across cultures, people have various ways of signaling their relatedness to others—they wear clan tartans, create totems, use special kin terms, and so on. But a domain-general learning mechanism that picks up local, transient cues to genetic relatedness like these cannot solve the problem of detecting genetic relatives: To deduce which cues predict relatedness locally, the mechanism would need to already know the genetic relatedness of others—the very problem it needs to solve.

The adaptive function of the kin detection system is to learn which individuals are biological kin, so it is a type of learning mechanism. But it cannot do its job unless its procedures are specialized for this function: It needs to contain within its evolved design a specification of the core cues that it will use to determine relatedness (although it may elaborate these ontogenetically). Because this specification is the product of evolution, other things being equal, selection will have favored the use of cues that performed well broadly across the species range, and consistently enough across generations to allow selection to act.

By considering the information about genetic relatedness that was built into the structure of hunter-gatherer life, we predicted that the kin detection system would use two independent cues as the source of its information about relatedness of siblings: maternal perinatal association and duration of coresidence during the period of parental investment. (We do not rule out the possibility that other cues may be found.)

Olders Detecting Younger Siblings

As mammals, human mothers nurse and care for their newborn infants, so seeing your own mother care for a newborn is a reliable cue that this baby is your sibling. We call this the *maternal perinatal association cue*, or MPA. Our data show that levels of altruism and sexual aversion toward a particular younger sibling are high for subjects who have been exposed to the MPA cue—that is for subjects who are older than their siblings and were present in the home when their biological mother was caring for that new baby. This is true no matter how long the subject and younger sibling subsequently coreside in the same household.

Youngers Detecting Older Sibs

If you are younger, this MPA cue will not work, because your older sibling was born before you. So to detect older siblings, the mind defaults to a different cue: How long you coresided with this child during the period of parental investment, between your ages 0–18. Hunter-gatherer bands are composed of several nuclear and extended families; as conditions change, these bands fission into smaller groups and later fuse back together again. But when they fission, they do so along nuclear family lines, with children staying with parents (especially mothers). Under such conditions, the more time you log with a child while growing up, the more closely related to you that individual is likely to be. (We found that duration of childhood coresidence is still highly correlated ($r = \sim .70$) with degree of kinship (i.e., full, half, or step siblinghood), even among the subjects in our study.)

When the MPA cue is absent, levels of altruism and sexual aversion toward a particular sibling are predicted by duration of childhood coresidence. It takes 14–18 years of coresidence to produce levels of altruism and sexual aversion toward siblings that are as high as those produced by the MPA cue. The group of people who are not exposed to the MPA cue includes all youngers detecting older siblings, all subjects with step and adoptive siblings, and about 12% olders with younger siblings.

The kinship estimator computes kinship indexes non-consciously, and independent of your conscious beliefs.

A striking example of this from our own research involves siblings who are step or adoptive—that is, siblings who the subject *knows* are not genetically related. Duration of coresidence predicts altruism and sexual aversion toward step and adoptive siblings, just as it does for youngers detecting older siblings. This shows that when conscious beliefs conflict with the output of the kin detection system, the criteria used by the kin detection system prevail.

Cue Integration by the Kinship Estimator

If the effects of MPA and coresidence duration were additive, the results for MPA and coresidence duration would also be consistent with a model in which data from the monitoring circuitry were being fed directly into each of the two motivational systems (sexual and altruism), with no intervening regulatory variable—that is, with no kinship index. But their effects were not additive: There is an interaction between the two cues. When the MPA cue is present, levels of altruism and sexual aversion toward that sibling are high, *and long coresidence durations do not result in any increase in their levels*. Coresidence duration affects levels of altruism and sexual aversion only when the MPA cue is absent.

That is, the effects of coresidence duration are conditional on the presence or absence of the MPA cue. For cues to be combined in this nonadditive way, there needs to be a mechanism that does the combining: That mechanism is what we are calling the *kinship estimator*. The data showing conditional cue use indicate that in computing kinship indexes, the kinship estimator employs an algorithm that combines the two cues in a noncompensatory way (as in a decision tree).

Importantly, the pattern of conditional cue use is the same, whether the dependent measure assesses levels of altruism (number of favors done for sibling *j* in the last month; willingness to donate a kidney to sibling *j*), levels of disgust at the thought of sex with sibling *j*, or degree of moral opposition to third party sibling incest (an unobtrusive measure of sexual aversion, which can be used in assessments of subjects with only one opposite sex sibling). This is important converging evidence for the model in Figure 15.1: Sibling altruism, sibling sexual aversion, and moral opposition to third party sibling incest—wildly disparate kin-relevant behaviors—are all being regulated by the same developmental cues, MPA and coresidence duration, combined in precisely the same way. It is a surprising finding, predicted by no other theory. Yet it is precisely what one would expect if the same internal regulatory variable, a kinship index, serves as input to two different motivational systems.

ANGER AS A RECALIBRATIONAL EMOTION

The existence of internal regulatory variables raises a series of interesting computational problems. Once computed, when are they recomputed? Do they get recalibrated as circumstances change and if so, which circumstances trigger their recalibration? Is it possible to change an internal regulatory variable in another person's head? We have been exploring the idea that the adaptive function of certain emotions—anger, gratitude, and guilt, for example—is to activate responses that recalibrate internal regulatory variables in one's own head and in the heads of other people (Sell, 2005; Sell, Tooby & Cosmides, in prep. a, b; Tooby & Cosmides, in press; Sznycer, Price, Tooby & Cosmides, in prep.).

AQ3

The welfare trade-off ratio should have a special relationship to social interaction and social motivations, because it governs the circumstances under which you will be willing to benefit yourself at another's expense or to help another person at your expense. For each individual j with whom one interacts, the mind should compute an intrinsic WTR_j and a monitored WTR_j . The intrinsic WTR_j governs how much you trade off your own welfare to enhance that of j regardless of any response from j (e.g., when j will not know what you did). For example, the kinship index should set a lower boundary on your intrinsic WTR toward a sibling, inclining you to help your sibling—within the constraints of Equation 15.2—even when there is no chance that your sibling will know

about or return the favor. The monitored WTR_j should govern the actions you take when j might observe or learn of the welfare trade-off ratio expressed in your actions and respond, either with anger or gratitude. Both, we propose, are recalibrational emotions. We will illustrate the idea of a recalibrational emotion with anger.

Our working hypothesis is that anger is the expression of a neurocomputational system that evolved to adaptively regulate behavior in the context of resolving conflicts of interest in favor of the angry individual. It evolved as an instrument of social negotiation, and was designed to operate in cooperative relationships as well as in antagonistic ones.

RAISING OTHERS' WTRs TOWARD YOU

Equations 15.1 and 15.2 express decision rules that should guide behavior when there is a conflict of interest. An implication of these equations is that any person, P , will treat you better when P 's welfare trade-off ratio toward you is higher (Figure 15.2). For example, Equation 15.1 says that if person P 's WTR toward you is 1, P values your welfare as much as his (or her) own; accordingly, P will refrain from taking any action that imposes a cost on you (C_{you}) that is greater than the benefit it provides to P (B_p)—that is, P will refrain when $B_p/C_{you} < 1$. But if P 's WTR toward you is $1/2$, P values your welfare only half as much as his own; that is, P will take actions for which $B_p/C_{you} > 1/2$. This means there is a set of cost-imposing

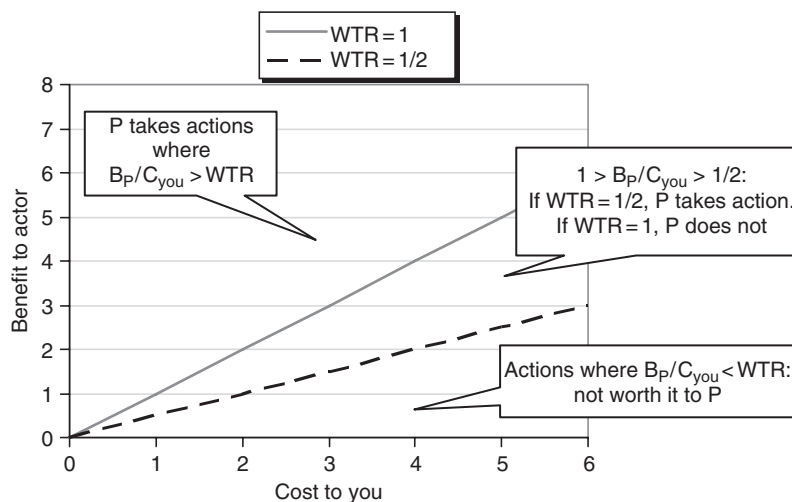


FIGURE 15.2 An actor's welfare trade-off ratio (WTR) toward you can be inferred by observing how large a cost that individual is willing to impose on you for how small a benefit gained. The gray line represents a WTR of 1, meaning that the actor values your welfare as heavily as his or her own. The black dashed line represents a WTR of $1/2$, meaning the actor values your welfare only half as much as his or her own. The area between these two lines represents the set of cost-imposing actions an actor would take if his or her WTR toward you were $1/2$, but not if it were 1. Raising an individual's WTR toward you allows you to avoid these costs.

actions, ones for which $\frac{1}{2} < B_p/C_{you} < 1$, that P will take when his $WTR_{you} = \frac{1}{2}$, but not when his $WTR_{you} = 1$ (Figure 15.2). You will be spared these costs if there is some way of raising P's WTR toward you from $\frac{1}{2}$ to 1.

But why should P raise his WTR toward you, when this reduces the set of self-beneficial actions that he will be willing to take? Humans, unlike most species, engage in many forms of cooperation: dyadic reciprocation (Gurven, 2004; Trivers, 1971), coalitional (group) cooperation (Tooby et al., 2006), food sharing as a form of risk pooling (Kaplan & Hill, 1985), deep engagement relationships (Tooby & Cosmides, 1996). If P does not raise his WTR toward you—that is, if he does not treat you better—then he may lose you as a cooperative partner.

If you are a good and reliable reciprocator, for example, then P benefits from having you as a cooperative partner. If your motivational system is designed to make your level of cooperation contingent on how well P treats you, then P might be able to increase your level of cooperation by treating you better, by raising P's own WTR_{you} . But P pays a price by increasing his WTR_{you} : A higher WTR_{you} means P will be sacrificing his own welfare more often for you, and refraining from a larger set of self-beneficial actions. So what price, in the form of a higher WTR_{you} , should P be willing to pay to maintain or increase your cooperation toward him?

There is an *equilibrium* WTR value, at which the marginal increase in price P would pay, in the form of a higher WTR_{you} , is exactly offset by the marginal increase in benefits P would gain by doing so, through increased cooperation from you. If P's WTR toward you is below this equilibrium value, the marginal decrease in your cooperation that this elicits will make P worse off than he could be. When this is true, there is the possibility of raising P's WTR toward you. By threatening to lower your level of cooperation with P—or even withdraw it by switching to a partner who values your welfare more highly (i.e., whose WTR toward you is higher)—it should be possible to raise P's WTR_{you} to a value closer to P's equilibrium point.

Another reason P might raise his WTR toward you is that you will hurt him if he does not. Like most other species, humans sometimes use aggression to force others to sacrifice for their welfare. Using variables such as the relative value of a resource to two contestants and their relative fighting ability, game theoretic models such as the Asymmetric War of Attrition (AWA) specify conditions under which a contestant should cede a resource or fight for it (Hammerstein & Parker, 1982; Maynard Smith & Parker, 1976). The AWA predicts that, if Y does not relinquish a resource, X will fight Y when $v(X)/v(Y) > k(X)/k(Y)$,

that is, when the relative value (v) of the resource to X exceeds the relative costs (k) that X will incur by fighting Y. More specifically, $k(X)$ is the rate at which X will incur injuries if a fight between X and Y ensues, which is a function of their relative fighting ability. Behavior consistent with the AWA requires programs that compute one's formidability relative to others, and use this information to adaptively regulate responses to resource conflict. For example, if you and person P value a resource equally and both of you know that P is more aggressively formidable than you are, the AWA predicts that P will try to take the resource and you will relinquish it rather than risk injury in a fight.

This means that, all else equal, very formidable individuals will be more willing to initiate resource contests than less formidable ones, and less formidable individuals will defer to these demands. If cooperation (and so forth), is not an issue, then there is an equilibrium WTR value toward you, based on your formidability relative to P, where the benefits to P of getting or keeping a resource of value V_p are exactly offset by the costs P will suffer by fighting you for it.

If P's current WTR toward you is below this equilibrium value, there is the possibility of raising it by threatening to aggress against P. Dominance hierarchies in species lacking cooperation are the result of such negotiations. In the absence of any contested resource, individual animals aggressively display toward one another, assessing who can hurt whom. Having determined this, injurious fights become unnecessary: Weaker individuals cede resources to stronger ones, whenever the relative value of the resource to the weaker one is less than the value of a regulatory variable expressing their relative formidability.

The AWA, Hamilton's rule, and reciprocal altruism theory each express how selection should shape an equilibrium WTR based on a single factor (formidability or kinship or value as a reciprocator). But humans engage in cooperation as well as aggression, and we live in the presence of kin as well as nonkin. This should select for a welfare trade-off ratio estimator equipped with algorithms that compute equilibrium WTRs based on the values of several different regulatory variables: Ones expressing an individual's value as a reciprocator, coalition mate, sexual partner, and friend in need, as well as the kinship index associated with that individual and a variable expressing that individual's formidability relative to one's own. Indeed, the welfare trade-off ratio estimator in your head needs to be able to compute two sets of WTRs: the WTRs that should regulate your behavior toward others, and the equilibrium WTRs that others should express toward you.

If P knows you will not respond by threatening to withdraw benefits or inflict costs, then P can benefit by having a WTR toward you that is lower than the equilibrium value. What can raise P's WTR_{you} nearer to the equilibrium value is your ability to monitor P's actions to see what WTR_{you} they express, and respond. Anger, we propose, is the activation of a response system designed to negotiate the value of the offending person's WTR toward you. We call this proposal the *recalibrational theory of anger* (Sell, 2005; Sell et al., in prep. a, b).

ANGER AS A NEGOTIATION OVER WTR VALUES

Given the ability to estimate the consequences of actions, the costs and benefits they impose on oneself and others, you can infer a person's WTR_{you} from his or her actions. Let us say you frequently associate with Aaron. If you observe Aaron taking an action that inflicts a cost of 4 units on you to gain a benefit of 1 unit for himself, you can infer that Aaron's $WTR_{you} \leq 1/4$. ($B_{Aaron}/C_{you} = 1/4$; Equation 15.1 means Aaron would take this action only if $B_{Aaron}/C_{you} \geq WTR_{you}$.)

It is common to get angry when someone imposes a cost on you and, all else equal, the larger the cost, the more angry you become—many theories predict that. But the recalibrational theory of anger predicts that being harmed will not be sufficient to trigger anger. If anger is the expression of a system designed to negotiate WTRs, then it should be triggered when the offending person's action expresses a WTR_{you} that is too low—below what you feel entitled to or, more specifically, below what your WTR estimator has computed as the appropriate equilibrium value. This leads to a counterintuitive prediction: Holding the cost imposed constant, more anger will be triggered when the offending person imposed that cost to gain a small benefit than to gain a large one.

Let us say that your WTR estimator has computed, based on the nature of your relationship, that Aaron's equilibrium WTR toward you should be $1/2$. You then see him ruin your expensive scarf, imposing a cost of 4 units on you. Whether you become angry should depend on how much he benefited by doing so. If the benefit he got was only 1 unit—let us say Aaron ruined your scarf by using it to wipe ketchup off his face—then this action expresses a $WTR_{you} \leq 1/4$. This is less than the equilibrium value of $1/2$, and so should trigger anger. But if Aaron ruined your scarf while using it to make a tourniquet to stop blood spurting from his child's arm, then the benefit he got was huge—let us say it was 24 units. This action should not trigger anger in you: Aaron's action is consistent with a WTR_{you} of $1/2$. Indeed, the

benefit to Aaron relative to the cost to you is consistent with Aaron having a WTR toward you as high as six ($B_{Aaron}/C_{you} = 24/4 = 6$). This means that Aaron would have taken this action even if his WTR_{you} was very high—even if he valued your welfare almost six times as much as his own. We have conducted experiments that held the cost imposed on the subject constant, varying the size of the benefit the offending individual expected to gain by imposing it. Learning that the offending action was taken to procure a large monetary benefit made subjects less angry; learning that it was taken to procure a small one made them more angry (Sell, 2005; Sell et al., in prep. a).

According to the recalibrational theory of anger, an *anger detection system* is activated when someone imposes a cost on you (or fails to provide an agreed upon or expected benefit—which will be seen as having imposed a cost). If the detection system infers that this person's monitored WTR_{you} is below an estimate of the appropriate equilibrium value, then the anger system is triggered. The detection system sends an "anger signal" that regulates two downstream motivational systems as negotiative tools—one regulating cooperation, the other regulating aggression.

The Cooperation System in Anger Mode

Let us say Aaron is a cooperative partner of yours, a friend or colleague, and you observe him taking an action that imposes a large cost on you for a small benefit. Your detection system infers that this action expresses a WTR_{you} of B_{Aaron}/C_{you} . This value is lower than the equilibrium value your welfare trade-off ratio estimator had computed as reasonable based on the benefits Aaron gains by your association, so your detection system sends an anger signal to the motivational system regulating cooperation. This flips the cooperation system into its anger mode, thereby motivating very specific arguments and forms of information exchange, each designed to solve a different recalibrational problem.

Problem 1: Aaron may not realize that his action imposed a cost on anyone; alternatively, he may realize his action very likely imposed a cost on someone, but the fact that it imposed a cost on you may be something he did not realize or intend. *Solution:* The cooperation system activates two specific motivational goals: To tell Aaron that the offending action imposed a cost on you, and to find out if Aaron realized his action would have this consequence before taking it. (If he could not have known his action would impose a cost on you, it does not imply his WTR_{you} is too low; discovering this should deactivate your anger system).

Problem 2: Aaron may have misestimated the magnitudes of the cost imposed for benefit gained. *Solution:* The cooperation system activates the goal of recalibrating those estimates, motivating you to argue that the cost imposed on you was higher and the benefit Aaron gained was lower than he thinks.

Problem 3: Aaron has underestimated your WTR_{Aaron} , resulting in an equilibrium WTR_{you} that is too low. (All else equal, Aaron—like everyone else—is better off associating with individuals whose WTR toward him is high rather than low, because such individuals will impose fewer costs on him and provide more benefits to him.) *Solution:* The cooperation system's search engines scour episodic memory for examples of times when you sacrificed your welfare for his (i.e., incurred high costs to provide even small benefits), as these imply that your WTR_{Aaron} is high. Retrieval of these episodes will be accompanied by an intense desire to remind Aaron of these acts.

Problem 4: Aaron has underestimated how much he benefits from having you as a cooperative partner, resulting in an equilibrium WTR_{you} that is too low. (This is different from Problem 3: Even if your WTR toward Aaron is low, you could be in a position to help and support him (at low cost to yourself), by virtue of your status, connections, or special skills.) *Solution a:* The cooperation system's search engines scour your episodic memory for examples of times you helped Aaron, providing important benefits to him. Such episodes should be easily retrieved, and accompanied by an intense desire to remind Aaron of these acts. *Solution b:* The cooperation system activates a specific motivation: To threaten to withdraw cooperation, accompanied by the desire to vividly describe how this will cause Aaron to suffer. Aaron's equilibrium WTR_{you} should increase if either response convinces him that the future benefits he will obtain from your association are high; Solution b adds the threat that he will be losing these future benefits if he does not treat you better.

Trying to solve problems 1–4 will elicit an information exchange. Aaron might come to agree with you and apologize. On the recalibrational theory of anger, a sincere apology expresses the offending person's willingness to place more weight on your welfare in the future, by recalibrating his WTR_{you} upwards and by recalibrating his misestimates of costs and benefits to self and to you. A sincere apology is a signal that the anger system's recalibrational function has been accomplished, so it should deactivate the anger system, flipping the cooperation system (and the aggression system) from anger mode to normal mode. (In normal mode, the cooperation system motivates goals consistent with social exchange, providing help, and soliciting help; Cosmides & Tooby, 2005; Tooby & Cosmides, 1996).

Alternatively, Aaron might respond that your variables need recalibrating: that you are exaggerating the cost he imposed, underestimating the benefits he gained, attributing bad intentions when he had none, exaggerating how much you have helped him in the past (overestimating your value to him) and at what personal cost (overestimating your WTR_{Aaron}), and forgetting how often he has come through for you and at what personal cost (i.e., your WTR_{Aaron} is lower than he deserves, justifying his lower WTR_{you}). If you come to agree with his points, this too should deactivate your anger system because you will no longer see his action as expressing a WTR_{you} that is too low. A complete meeting of the minds on all points is unnecessary to dispel your anger: Adjustment of variables sufficient to indicate that Aaron's WTR_{you} is not too low should be enough. But what if this does not happen?

Problem 5: Aaron's estimates of the costs and benefits associated with his action agree with yours, and so does his estimate of the appropriate equilibrium value for his WTR_{you} . But he believes you will not respond when his actions express a WTR_{you} below equilibrium. *Solution:* The cooperation system activates a specific motivation: to threaten to withdraw cooperation from Aaron. Demonstrating that you are monitoring his WTR_{you} and are willing to respond by downregulating your cooperation is a way of increasing his *monitored* WTR_{you} to nearer his equilibrium value.

Problem 6: After all this, Aaron does not apologize; indeed, he indicates that he has no intention of raising his WTR_{you} . *Solution:* Recalibrate the value of your equilibrium WTR_{Aaron} , lowering it to reflect the fact that he places less weight on your welfare than you had expected. The functional product of this will be to downregulate your levels of cooperation toward Aaron.

In cooperative relationships, lowering—or threatening to lower—your WTR toward someone has functional consequences: Threatening to lower it motivates reform in insufficient reciprocators; actually lowering it cuts losses with cheaters.

Research testing for these specific anger responses as solutions to problems 1–7 is still in progress, but we have already tested for and confirmed a number of them, using vignette experiments and naturally occurring arguments collected from subjects. These experiments and results are reported in Sell (2005), Sell et al. (in prep. a), and Sznycer et al. (in prep.).

The Aggression System in Anger Mode

Another way to negotiate WTR s is by threatening harm, so the detection system's anger signal should sometimes activate the motivational system regulating aggression.

This should be more common in antagonistic relationships than in cooperative ones—although even cooperative relationships can explode into threats of aggression when the target of anger refuses to acknowledge the costs imposed on you, your value as a cooperative partner, or otherwise expresses unwillingness to raise his or her WTR_{you} . Whether the relationship is cooperative or not, threatening harm is a bad tactic unless you can back it up: That is, unless you are capable of inflicting harm at low cost to yourself. Therefore, the aggression system should flip into anger mode more easily in aggressively formidable people, especially formidable men (absolute levels of aggression vary between cultures, but within cultures women are far less likely than men to resolve conflicts by using physical force; Campbell, 2002; Daly & Wilson, 1988).

Let us say that circumstances force you and Aaron to interact, but you do not have a close cooperative relationship. Moreover, Aaron's WTR_{you} is low because he has a low estimate of your formidability relative to his. He communicates this to you and others through insults: comments impugning your willingness to fight, disparaging your strength, advertising a flippant disregard for your distress, and other forms of “disrespect”—claims or demonstrations that he can treat you badly without fear of harm from you. If his estimate of your formidability is correct, you may need to accept a low WTR from Aaron. If it is not correct, insults and actions expressing a low WTR_{you} should flip your aggression system into anger mode (as opposed, for example, to a mode appropriate for hunting or for defending against predators). When this happens, the aggression system should motivate specific actions and goals, each designed to solve a recalibrational problem. For example:

AQ4

Problem 7: Aaron's formidability index for you (FI_{you})—the regulatory variable in his head that expresses your formidability relative to his own—is too low. *Solution:* The aggression system activates a specific goal: to recalibrate FI_{you} in Aaron's head. The aggression system's search engines should scour your episodic memory for times when you were able to injure Aaron, and retrieval of such episodes should be accompanied by an intense desire to remind Aaron of them. It should also motivate actions that demonstrate your ability to harm him: towering over him, breaking something that is difficult to break, confronting him with your friends present (demonstrating formidability through coalitional aggression), and so on. Provocative actions—taking or breaking something Aaron values while he and others are watching—can serve the same function: They are challenges one would not issue without the ability to back them up.

If these reminders and demonstrations are successful, they should raise Aaron's FI_{you} and his WTR_{you} , because his WTR should be based, at least in part, on his assessment of your formidability (see discussion of the AWA, above) and that of your coalitional allies. (Your coalitional formidability should be registered by a distinct regulatory variable in Aaron's motivational architecture, not merely by FI_{you} , which indexes your individual formidability).

Note that these displays can also serve a parallel function: to communicate how much you value a resource or how large a cost Aaron's action imposed on you. That is, they can serve an expressive function as well, providing a solution to Problem 2 above (Aaron's misestimate of costs imposed or benefits gained). In nonverbal animals, escalating displays and an unwillingness to back down are means used to signal how much one values a contested resource (Austad, 1983; Enquist & Leimar, 1987).

Problem 8: Despite your displays, Aaron does not adjust his FI_{you} (and WTR_{you}) or his estimates of the costs imposed for benefits gained: He refuses to signal deference, submission, or “respect.” Indeed, he makes clear his belief that you will not respond with aggression when his actions express a monitored WTR below equilibrium. *Solution:* The aggression system should activate a specific motivation: to threaten to harm Aaron. The harm can be physical or social.

Threatening physical harm carries the risk that Aaron will call your bluff. Therefore, this motivation is more likely to be activated (1) when you actually are more formidable than Aaron, (2) when Aaron cannot harm you by withdrawing benefits, or (3) when social constraints would prevent a fight from actually breaking out (e.g., friends or authorities are present who will “hold you back”). If you are not more aggressively formidable than Aaron, the aggression system should search for means by which you or an ally can harm Aaron by preventing him from gaining important benefits (e.g., ruining his reputation in the eyes of his girlfriend or employer). If such means exist, then the aggression system should motivate you to make a social threat: to block benefits from flowing to Aaron, either personally or by inducing an ally to do so. Being in the position to do so might be conceived as a form of social status or social formidability, registered by a different variable than the one that registers aggressive formidability.

In anger mode, the aggression system should activate the motivation to escalate the displays and threats until one of you backs down. But what if neither of you backs down?

Problem 10: Despite your threats, Aaron does not back down: The threats do not cause him to recalibrate

his FI_{you} (and WTR_{you}) upwards. *Solution:* The aggression system activates the goal of actually harming Aaron. This may lead to a fight, which will end when its informational function has been accomplished—that is, when it becomes clear that one of you can, in fact, inflict more injury on the other. The function of this escalation, from insults to threats to aggression, is to establish formidability, not to kill, but on rare occasions people die from injuries incurred during this “negotiation.” Of the homicides that do occur, a large number result from the escalation of what police call a “trivial altercation”—a public confrontation between two men over “face” or “respect” (Daly & Wilson, 1988).

Note two implications of this analysis of the role of aggression in negotiating WTRs. The first we mentioned above: given a situation that triggers anger, the aggression system should more easily flip into anger mode in people who are stronger (more formidable), because they can more safely use aggression to enforce a higher WTR toward themselves. Second, because they can inflict more injury at lower cost to themselves, aggressively formidable people should expect a higher equilibrium WTR from others, one where the benefits of not being harmed by the formidable person are exactly offset by the price of the higher WTR. All else equal, stronger, more formidable individuals should feel more entitled to deference and respect, more entitled to having other people’s actions take their interests into account.

According to the recalibrational theory, anger is triggered by actions expressing a WTR below the equilibrium value you expect from others. This means that those who expect a higher WTR will be provoked by a larger set of actions than those who expect a lower WTR. For example, the set of actions between the two curves in Figure 15.2 should trigger anger in someone expecting a WTR of 1, but not in someone expecting a WTR of $\frac{1}{2}$. If more aggressively formidable people expect a higher equilibrium WTR from others, then there is a set of cost-imposing actions that will trigger anger in them, but would not trigger anger in someone expecting a lower WTR. This leads to a surprising prediction that we have confirmed (Sell, 2005; Sell et al., in prep. b): Men who are physically stronger (as measured by lifting strength at the gym) are more prone to anger, feel more entitled to having their way, and have greater success resolving conflicts of interest in their favor. They have also been in more fights and believe more in the efficacy of aggression to settle conflicts. Interestingly, this belief in the efficacy of aggression reflects more than a rational assessment of their ability to win fights: It extends to international conflicts, where their personal strength could not possibly make a difference. We had

predicted this in advance, on the grounds that modern humans think about conflicts between nation states with a mind designed for the ancestral world of hunter-gatherers. In that smaller world, a man’s personal strength would be an important factor contributing to the formidability of the small coalitions (2–5 individuals) in which he takes part (Tooby et al., 2006).

Approach Motivations in Anger

A common way of conceptualizing approach–avoidance motivation is to view positive stimuli as eliciting approach and negative stimuli as eliciting avoidance (Elliot, 2006). But in anger, a very negative stimulus—someone who has done you wrong—elicits approach, not avoidance. Indeed, the motivation for “approach” when you are angry can be overwhelming—so much so that when circumstances prevent you from expressing your feelings to the person you are angry with, the sense of frustration can be intense.

Nor is there a single way of characterizing approach in anger. When the cooperation system operates in anger mode, the approach response is to exchange information, argue, and, if necessary, withdraw or threaten to withdraw cooperation. When the aggression system operates in anger mode, the approach response is to demonstrate formidability, threaten harm, and, if necessary, actually injure the antagonist. Approach is a very rough way of characterizing behavioral responses. Like anger, foraging, courtship, and helping all involve approaching stimuli, yet the motivational systems regulating these activities have little in common with one another, and the approach behaviors they produce are unrecognizably different.

CONCLUSION

We can only move toward or away from things, so approach and avoidance capture a lot of what we do in life. The great appeal of describing responses in this way is that it characterizes behavior at an abstract level, allowing generalizations that apply across many different concrete situations. What we have been trying to show, however, is that this satisfying level of abstraction can be achieved while still providing fine-grained descriptions of behavioral responses. The recalibrational theory of anger, for example, contains a fine-grained description of the *specific content* of arguments, yet these are described at an abstract level that applies to countless concrete situations (“You inflicted a large cost on me!” “You did it on purpose!” “You did it for a trivial benefit to yourself!” “I’ve been good to you!”, “I’ve sacrificed for you!” “If you’re going to continue to treat me this way, I won’t treat you so well in the future!”).

The key to achieving abstract yet detailed characterizations of social motivations lies in taking an evolutionary and computational approach to motivation. Internal regulatory variables are by their nature abstract: They may use concrete situations as input—acts of sacrifice for welfare trade-off ratios, duration of coresidence, and observations of one's own mother caring for an infant for kinship indexes—but they use these concrete situations to compute the magnitude of a variable, abstracted from those situations. These values are used by motivational systems, which activate abstract goals (“tell him how much he hurt me by doing X”) that get filled in with concrete content depending on the situation.

From afar, evolutionary approaches to motivation can sound like a cartoon—avoid the lion, find the mate. But the knowledge developed in evolutionary biology and behavioral ecology has become quite sophisticated. In presenting a computational approach to motivation, we described the relevant selection pressures for each adaptive problem in detail, including some of the game theoretic analyses that have been done. We did this to communicate how far evolutionary analyses have come: Models of optimal foraging, kin selection, reciprocal altruism, the asymmetric war of attrition, the population genetics of deleterious recessives (and many others) offer important insights to anyone interested in studying motivation, social, or otherwise. These models tell us what problems of survival and reproduction our hunter-gatherer ancestors faced. More importantly, they tell us what responses would have counted as reproduction-promoting solutions to these problems in the ancestral environments that selected for the architecture of our motivational systems.

Scientists from different fields provide gifts to one another. We psychologists have developed a framework for talking about the regulation of behavior by information, which has been transforming the study of animal behavior, from vague talk about dispositions to computational, information-processing models. Evolutionary biologists and behavioral ecologists have provided us with a gift too, and we should open it and take a look. Their models can prompt us to search for internal regulatory variables whose existence we might not otherwise suspect, and provide us with insights necessary for constructing abstract yet fine-grained descriptions of the responses our motivational systems were designed by natural selection to produce.

ACKNOWLEDGMENTS

We thank Howard Waldow and the NIH Director's Pioneer Award (LC) for making this research possible.

REFERENCES

- Adams, M. S., & Neel, J. V. (1967). Children of incest. *Pediatrics*, *40*, 55–62.
- Austad, S. (1983). A game theoretical interpretation of male combat in the bowl and doily spider. *Animal Behaviour*, *31*, 59–73.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press. AQ5
- Bittles, A., & Neel, J. (1994). The costs of human inbreeding and their implications for variation at the DNA level. *Nature Genetics*, *8*, 117–121.
- Boyd, R. (1988). Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethology and Sociobiology*, *9*, 211–222.
- Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology*, *13*, 171–195.
- Boyer, P. (2001). *Religion explained: The evolutionary roots of religious thought*. New York: Basic Books.
- Boyer, P., & Lienard, P. (2006). Why ritualized behavior? Pre-AQ6 caution systems and action parsing in developmental, pathological, and cultural rituals. *Behavioral and Brain Sciences*, *12*(6), 595–613.
- Buss, D. (2005). *The murderer next door*. New York: The Pen-AQ7 guin Press.
- Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, *100*, 204–232.
- Campbell, A. (2002). *A mind of her own: The evolutionary psychology of women*. London: Oxford University Press.
- Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II Case study: A computational theory of social exchange. *Ethology and Sociobiology*, *10*, 51–97.
- Cosmides, L., & Tooby, J. (2000a). Consider the source: The AQ8 evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 53–115). Vancouver Studies in Cognitive Science. New York: Oxford University Press.
- Cosmides, L., & Tooby, J. (2000b). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions, 2nd Edition*. (pp. 91–115). New York: Guilford.
- Cosmides, L., & Tooby, J. (2001). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Sternberg & J. C. Kaufman (Eds.), *The evolution of intelligence* (pp. 145–198). Hillsdale, NJ: Erlbaum.
- Cosmides, L., & Tooby, J. (2005a). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: Wiley.
- Cosmides, L., & Tooby, J. (2005b). Social exchange: The evolutionary design of a neurocognitive system. In M. S. Gazzaniga, (Ed.), *The new cognitive neurosciences, III* (pp. 1295–1308). Cambridge, MA: MIT press.

- AQ9 Cosmides, L., & Tooby, J. (2007). Can a general deontic logic capture the facts of human moral reasoning? How the mind interprets social exchange rules and detects cheaters. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 1: The evolution of morality*. Cambridge, MA: MIT Press.
- Daly, M., & Wilson, M. (1983). *Sex, evolution, and behavior, 2nd edition*. Boston: Willard Grant.
- Daly, M., & Wilson, M. (1988). *Homicide*. Chicago: Aldine.
- AQ10 Dawkins, R. (1979). Twelve misunderstandings of kin selection. *Zeitschrift für Tierpsychologie, 51*, 184–200.
- AQ11 Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford.
- AQ12 Dennett, D. (1988). Quining qualia. In A. Marcel & E. Bisiach, (Eds.), *Consciousness in modern science*. New York: Oxford University Press.
- Elliot, A. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion, 30*, 111–116.
- Enquist, M., & Leimar, O. (1987). Evolution of fighting behaviour: The effect of variation in resource value. *Journal of Theoretical Biology, 127*, 187–205.
- AQ13 Ermer, E., Guerin, S., Cosmides, L., Tooby, J., & Miller, M. (2006). Theory of mind broad and narrow: Reasoning about social exchange engages ToM areas, precautionary reasoning does not. *Social Neuroscience, 1*(3–4), 196–219.
- Fodor, J. (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.
- Gurven, M. (2004). To give or not to give: An evolutionary ecology of human food transfers. *Behavioral and Brain Sciences, 27*, 543–583.
- Gurven, M., Allen-Arave, W., Hill, K., & Hurtado, M. (2000). It's a wonderful life: Signaling generosity among the Ache of Paraguay. *Evolution and Human Behavior, 21*, 263–282.
- AQ14 Haig, D. (1993). Genetic conflicts in human pregnancy. *The Quarterly Review of Biology, 68*(4), 495–532.
- Hamilton, W. D. (1964). The genetical evolution of social behavior. I and II. *Journal of Theoretical Biology, 7*, 1–52.
- Hammerstein, P., & Parker, G. A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology, 96*, 647–682.
- AQ15 Hart, J., Berndt, R. S., & Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction. *Nature, 316*, 439–440.
- Huntingford, F. A., & Turner, A. K. (1987). *Animal conflict*. New York: Chapman & Hall.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Kaplan, H., & Hill, K. (1985). Food sharing among ache foragers: Tests of explanatory hypotheses. *Current Anthropology, 26*, 223–239.
- Klein, S., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109*, 306–329.
- Klein, S., German, T., Cosmides, L., & Gabriel, R. (2004). A theory of autobiographical memory: Necessary components and disorders resulting from their loss. *Social Cognition, 22*(5), 460–490.
- Lieberman, D., Cosmides, L., & Tooby, J. (2007). The architecture of human kin detection. *Nature, 445*, 727–731.
- Lorenz, K. (1966). *On aggression*. New York: Harcourt, Brace and World.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Maynard Smith, J., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Behavior, 24*, 159–175.
- Nesse, R. M., & Williams, G. C. (1995). *Why we get sick*. New York: Times Books.
- Oberman, L., & Ramachandran, V. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin, 133*(2), 310–327. AQ16
- Seemanova, E. (1971). A study of children of incestuous matings. *Human Heredity, 21*, 108–128.
- Sell, A. (2005). Regulating welfare trade-off ratios: Three tests of an evolutionary-computational model of human anger. Doctoral dissertation.
- Sell, A., Tooby, J., & Cosmides, L. (in prep.) Anger and welfare trade-off ratios: Mapping the computational architecture of a recalibrational emotion system.
- Sell, A., Tooby, J., & Cosmides, L. (in prep.) The logic of anger: Men, formidability and conflict.
- Smith, E. A., & Winterhalder, B. (1992). *Evolutionary ecology and human behavior*. New York: Walter de Gruyter.
- Sperber, D. (1994). The modularity of mind and the epidemiology of representations. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain-specificity in cognition and culture*. New York: Cambridge University Press.
- Sugiyama, L. S. (2004). Illness, injury, and disability among Shiwi forager-horticulturalists: Implications of health-risk buffering for the evolution of human life history. *American Journal of Physical Anthropology, 123*, 371–389. AQ17
- Sugiyama, L. S. (2005). Physical attractiveness in adaptationist perspective. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 292–343). New York: Wiley.
- Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford.
- Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In P. R. Abramson & S. D. Pinkerton (Eds.), *Sexual nature, sexual culture* (pp. 80–118). Chicago: University of Chicago Press.
- Szyner, D., Price, J. G., Tooby, J., & Cosmides, L. (in prep.). Recalibrational emotions and welfare trade-off ratios: Cooperation in anger, guilt, gratitude, pride, and shame.
- Tooby, J. (1982). Pathogens, polymorphism, and the evolution of sex. *Journal of Theoretical Biology, 97*, 557–576.
- Tooby, J., & Cosmides, L. (1989). Kin selection, genic selection, and information-dependent strategies. *Behavioral and Brain Sciences, 12*, 542–544.
- Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11*, 375–424.

- Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Eds.), *Evolution of social behaviour patterns in primates and man. Proceedings of the British Academy*, 88, 119–143.
- Tooby, J., & Cosmides, L. (2001). Does beauty build adapted minds? Toward an evolutionary theory of aesthetics, fiction and the arts. *Substance, Issue 94/95*, 30(1), 6–27.
- AQ18 Tooby, J., & Cosmides, L. (2006). The evolved architecture of hazard management: Risk detection reasoning and the motivational computation of threat magnitudes. Comment on Boyer & Lienard. *Behavioral and Brain Sciences*, 29(6), 631–633.
- Tooby, J., & Cosmides, L. (in press). The evolutionary psychology of emotions and their relationship to internal regulatory variables. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions* (3rd ed.). New York: Guilford.
- Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and content* (pp. 305–337). New York: Oxford University Press.
- Tooby, J., Cosmides, L., & Price, M. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics*, 27, 103–129.
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man: 1871–1971* (pp. 136–179). Chicago: Aldine.
- Trivers, R. (1974). Parent-offspring conflict. *American Zoologist*, 14, 249–264.
- Tye, M. (2003). Qualia. *Stanford Encyclopedia of Philosophy*. www.plato.stanford.edu/entries/qualia.
- Williams, G. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.
- Williams, G. C., & Williams, D. C. (1957). Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution*, 17, 249–253.
- Winterhalder, B., & Smith, E. A. (2000). Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evolutionary Anthropology*, 9, 51–72.

AUTHOR QUERIES

- [AQ1] Please fix “a and b” for the reference “Cosmides & Tooby, 2005” in both citations and list.
- [AQ2] The reference “Lieberman, 2003” is not provided in the list. Please provide.
- [AQ3] Kindly update the details for the references “Sell et al., in prep. a,b; Tooby & Cosmides, in press; Sznycer et al., in prep.” in both citation and list.
- [AQ4] Please check “For example:” Sentence is incomplete.
- [AQ5] The reference “Baron-Cohen, 1995” is not cited in the text. Please check.
- [AQ6] The reference “Boyer & Lienard, 2006” is not cited in the text. Please check.
- [AQ7] The reference “Buss, 2005” is not cited in the text. Please check.
- [AQ8] Please confirm if the insertion of “a” and “b” in the year “2000” is correct in the references “Cosmides & Tooby, 2000.”
- [AQ9] Please provide the Chapter page numbers for the reference “Cosmides & Tooby, 2007.”
- [AQ10] The reference “Dawkins, 1979” is not cited in the text. Please check.
- [AQ11] The reference “Dehaene, 1997” is not cited in the text. Please check.
- [AQ12] Please provide the Chapter page numbers for the reference “Dennett, 1988.”
- [AQ13] The reference “Ermer, Guerin, Cosmides, Tooby, & Miller, 2006” is not cited in the text. Please check.
- [AQ14] The reference “Haig, 1993” is not cited in the text. Please check.
- [AQ15] The reference “Hart, Berndt, & Caramazza, 1985” is not cited in the text. Please check.
- [AQ16] The reference “Oberman & Ramachandran, 2007” is not cited in the text. Please check.
- [AQ17] The reference “Sugiyama, 2004” is not cited in the text. Please check.
- [AQ18] The reference “Tooby & Cosmides, 2006” is not cited in the text. Please check.

