In "Handbook of Approach and Avoidance Motivation", Andrew Elliot, Editor. 2008.
Mahwah, NJ: Lawrence Erlbaum Associates

Evolution of Evaluative Processes I

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

THE NEXT COGNITIVE REVOLUTION: THE ADAPTATIONIST INTEGRATION OF MOTIVATION AND COGNITION

The discovery by biologists and physicists that natural selection is the only antientropic force that builds functional machinery into organisms led to an important insight: Natural selection provides the underlying theories explaining why functional mechanisms in the species-typical architecture of the brain have the designs that they do (Tooby, Cosmides, & Barrett, 2003). This connects the evolutionary sciences to psychology and neuroscience directly. Models of selection pressures (adaptive problems) faced by a species provide the design criteria that a species' mechanisms evolved to solve. Mechanisms evolved their design features—their functional properties—as methods for solving these adaptive problems.

Evolutionary psychology as a framework emerged because of the scientific benefits of employing these facts explicitly in research (Buss, 2005; Tooby & Cosmides, 1992). It proceeds by (1) deriving models of adaptive problems from evolutionary biology and our knowledge of the structure of the ancestral world, and then (2) using these models to design critical empirical tests of competing theories about the architecture of the mechanisms (if any) that evolved to solve them.

An equally essential element of evolutionary psychology is its participation in the cognitive revolution. The brain's properties as a physical system were organized by natural selection so that they function as an information processing system or organ of computation. It takes information as input, performs operations on it, and uses the output to regulate behavior so that it solves adaptive problems more effectively than the organism could in the absence of those procedures.

The ability to describe the functional properties of psychological mechanisms in terms of their computational operations gives us the appropriate language for characterizing their designs in terms of their evolved functions—functions that are, by their nature, inherently computational and regulatory. In short, the brain contains, not metaphorically but actually, evolved programs designed by natural selection to compute the solutions to adaptive information-processing problems involving the regulation of behavior.

Because humans, like other organisms, were challenged over their evolution by a rich diversity of adaptive problems (e.g., disease avoidance, mate selection), successful behavior regulation favored the evolution of a multiplicity of programs to solve them (e.g., disgust, sexual attraction). As we will demonstrate with two main examples—kin detection and anger—the structure of an evolved program can be discovered to embody a computational problem-solving strategy whose circuit logic exploits the ancestral structure of the adaptive problem. For example, the structure of ancestral hunter-gatherer life provided stably informative cues to genetic relatedness that our kin detection system evolved to target (see below; Lieberman, Tooby, & Cosmides, 2007).

Although there is a great emphasis in the traditional cognitive sciences on how organisms perceive and understand the world, there is astonishingly little cognitive work mapping how motivation and valuation work to regulate action. Because cognitive science descended from philosophy, cognitive scientists often treat the mind as if it exists solely to discover truths (as with perception, learning, and reasoning) rather than to regulate action adaptively. Fodor, for example, expresses this view when he says that the function of cognition is "the fixation of true beliefs" (Fodor, 2000, p. 68). Of course, true beliefs may be one useful element in the adaptive regulation of behavior. But as Hume was the first to point out, true beliefs by themselves have no implications for how to behave—what to approach, what to avoid, what to value, how to feel, what to do (Tooby, Cosmides, & Barrett, 2005). Encyclopedias have no motivations. As Hume understood, value is not an objective property of the external world, there to be observed. A man may be sexually attractive to many women, but sexually repulsive to his sister—so which is he "really"? In reality, value information must be internally computed and, unlike true beliefs, may validly differ across individuals. Moreover, value information is an indispensable component of almost every decision about how to behave. We argue in this chapter that there is a large and often overlooked class of neurocomputational programs that evolved to compute adaptive valuations (and their inputs)—valuations that are incapable of being either true or false.

Fodor (2000) justifies cognitive scientists' neglect of so-called conative processes (processes governing preferences, approach, avoidance, motivation, and valuation) by arguing that cognitive and conative mechanisms are separate; therefore, cognitive science can neglect motivation without being deformed in the process. In contrast, we think the cognitive sciences have been impaired by this artificial division. As we explore below with two case studies—kin detection and anger—computational elements for fact and value are often inextricably joined within the same cognitive adaptations, and so must be studied together.

The purpose of this chapter is to sketch out a new framework for thinking about motivation that is not only

computational and grounded in evolutionary biology, but also shows how motivational elements articulate with the rest of the cognitive architecture as part of a single, coevolved functional system. Findings in the evolutionary sciences imply the existence of a large number of adaptive problems—including problems in social interaction—for which there exist no corresponding motivational theories. We will illustrate this computational approach to motivation with several of these adaptive problems, including incest avoidance, kin selection, power-based bargaining, and reciprocity.

In order to construct a theoretical framework capable of incorporating this new range of cases, we need to introduce a new class of computational elements that have no present counterpart in the cognitive sciences. We think serious analysis of how the human brain accomplishes certain tasks involving valuation and behavior-regulation forces us to posit such entities. Indeed, not only do we think they are theoretically mandated, but we are involved in a series of research programs to demonstrate that they are psychologically and neurally real. We call these computational elements "internal regulatory variables."

INTERNAL REGULATORY VARIABLES AND MOTIVATION

For both theoretical and empirical reasons, we expect that the architecture of the human mind is by design full of registers for evolved variables whose function is to store summary magnitudes (or parameters) that allow value computation to be integrated into behavior regulation (Kirkpatrick & Ellis, 2001; Lieberman et al., 2007). These internal regulatory variables are not traditional theoretical entities such as concepts, representations, goal states, beliefs, or desires. Instead, they are indices that acquire their meaning by the evolved behavior-controlling and motivation-generating procedures that access them. That is, each has a location embedded in the input–output relations of our evolved programs, and their function inheres in the role they play in the decision flow of these the programs. We have evolved specializations designed to compute them and to output them to critical junctures in our 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 hauling back to camp.

These mathematical models have implications for the computational architecture of the motivational systems that regulate approach and avoidance while foraging. That these models successfully predict behavior implies that the brain has programs that compute, for each food, 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. For !Kung foragers, the values that apply to mongongo nuts need to be stored in a separate mental file folder from those that apply to acacia beetles, Grewia berries, ivory palm, Tsama melons, hartebeest meat, and hundreds of other foods. Functionally, one would expect the evolution of a foraging-specialized data format consisting of (at least) four registers, each dedicated to indexing one of the four variables. When foraging, the values of these variables are accessed by a program that combines them, producing motivations expressed in choices. As a result, we observe foragers seeking foods with better joint combinations of package size, search time, calorie density, and handling time, over worse combinations, according to the algorithm in the motivational system that integrates them. Because foraging motivations are regulated by the magnitudes of these four variables, they are examples of internal regulatory variables.

Internal regulatory variables are not an exotic feature of human motivational systems; 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 simply one class of regulatory system. They differ from regulatory systems like the Krebs cycle primarily in that 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 designed by selection. 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 equally 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 those properties of persons, acts, and situations that are needed to compute adaptive motivations. Examples include 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 genetically 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 inputs needed by evolved decision-making programs in order to generate motivations to action. Internal regulatory variables have magnitudes or discrete parameter values. They encode value, provide formatted input to mechanisms that compute value, or provide parameter values to decision-making circuits.

Felt Experience and Internal Regulatory Variables

Because we are subjectively aware of a rich world of feeling involved in motivations, 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. 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. Before proceeding, we would like to explain how the intuitive clarity of felt experience neither contradicts nor pre-empts the need for a computational account of motivation.

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 cannot be consciously accessed; they are accessed only by mechanisms internal to 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 broadcasts, in a way that is accessible to many other mechanisms, the magnitude of certain internal regulatory variables (Tooby & Cosmides, 2008). In other cases, a felt experience may be the output of a motivational system, with its felt intensity regulated by the (nonconscious) 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 outputs of motivational systems 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). Imagined alternatives 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, 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.

Conscious and Nonconscious Access to Internal Regulatory Variables

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 is unaccompanied by any felt experience—the brain does not seem to have any design feature capable of consciously 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 CO_2/O_2 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 CO_2/O_2 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 caloric content. That these felt experiences can be generated by imagination alone suggests that values for an internal regulatory variable registering the caloric content of each were previously stored; imagining oneself, seeing, and eating them initiates a process that transforms their magnitudes into a data format of felt experience.

Tracking different properties of the world—caloric content versus handling time, for example—clearly requires distinct regulatory variables. But if felt experience is functional—allowing imagination-based planning, for example—then the data formats by which distinct variables are experienced need to be different from one another, and qualitatively different to the extent they need to encode different types of information. Different regulatory variables need to be associated with distinct types of *qualia*, to use the philosophers' term (Tye, 2003). So the output of different regulatory variables into consciousness feels qualitatively different. In order to make decisions, however, at some level in the architecture (conscious or nonconscious) these different data types need to be tagged with a kind of information that makes them comparable—payoff information.

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 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 estimate of a familiar other's degree of genetic relatedness to oneself (Lieberman et al., 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 kinship index produces feelings of disgust when accessed by the sexual motivation system at the possibility of sexual contact with the person, 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 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.

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

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. But from an evolutionary perspective, what is social interaction for? What problems of survival, reproduction, and fitness promotion do individuals face when they live socially, and what behavioral responses count as adaptive solutions to these problems? We cannot rely on intuition to answer these questions because the history of the behavioral and biological sciences shows that, until the 1970s, many of the most prominent behavioral theories were based on serious misunderstandings of how natural selection works (Williams, 1966; see also Tooby & Cosmides, 1992).

Fortunately, over the last 40 years, evolutionary researchers have carefully analyzed how natural selection shapes the social interactions of many species. As a result, they have developed formal theories defining a series of specific adaptive problems arising from social life—theories that also specify what behavioral patterns constitute adaptive solutions. These models have been validated using the behavior of thousands of species. For example, the theory of kin selection analyzes selection on altruism within the family. This theory specifies how human motivational adaptations should be designed to make decisions about, for example, 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 deleterious recessives and coevolving pathogens lead to predictions about motivational systems regulating inbreeding avoidance (Lieberman et al., 2007; Tooby, 1982). Theories of sexual selection define adaptive problems and solutions posed by courtship and mating (Buss & Schmitt, 1993; Daly & Wilson, 1983; Symons, 1979; Trivers, 1972; Williams, 1966). The asymmetric war of attrition is a game theoretic model of the selection

pressures shaping bargaining, aggression, dominance, and resource division (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) describe adaptive problems that friendships and within-group sharing solve. Theories of reciprocal altruism and social exchange illuminate selection pressures shaping two-person exchange (Axelrod & Hamilton, 1981; Boyd, 1988; Cosmides & Tooby, 1989; Trivers, 1971). 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).

To map certain components of the evolved psychological architecture of our species, we have found it useful to start with a task analysis of the adaptive problems defined by these models. This helps to specify what properties computational systems capable of solving them would need. In doing this, it rapidly became clear that the computational systems that produce social motivations would need internal regulatory variables. They are necessary in order to track those properties and actions of persons that are relevant to computing the adaptive solutions specified by these theories.

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 renders them sexually valueless to me, but not to you; on the other hand, if we are both sick and need care, that same sibling is likely to be more valuable to me than to you.

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 *i*'s value as a sexual partner to me, their genetic relatedness 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, and so on. We will illustrate this first with genetic relatedness, 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, lie at the opposite end of this spectrum. Hunter-gatherer children typically grow up in families with 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 are motivated to live with relatives nearby, if that is an option.

That close genetic relatives frequently interacted ancestrally 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). Genetic relatedness refers to the increased probability, compared to the population average, that two individuals will both carry the same randomly sampled gene, given information about common ancestors. The relatedness between two individuals is typically expressed by a measure, the degree of relatedness, r_{ij}, expressed as a probability. This is a continuous variable that for humans usually has an upper bound around five (with full siblings, parents and offspring) and a lower bound of zero (with nonrelatives). Two different social motivation systems require an internal regulatory variable that tracks genetic relatedness: one governing sexual attraction/aversion, the other governing altruism. We first describe the selection pressures that should have shaped these motivational programs, then turn to computational models of the motivational programs 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 every generation. Many of them disrupt the

functioning of our tightly engineered regulatory systems. A single mutation can, for example, prevent a gene from being transcribed (or from producing the right protein). Given that our chromosomes come in pairs (one from each parent), a nonfunctional mutation need not be a problem for the individual it appears in. If it is found on only one chromosome of the pair and is recessive, the other chromosome will produce the right protein and the individual may be healthy. But if the same mutation is found on both chromosomes, the necessary protein will not be produced by either. The inability of an organism to produce one of its proteins can impair its development or kill it.

Such genes, called "deleterious recessives," are not rare. They accumulate in populations precisely because they are not harmful when heterozygous—that is, when it is matched with an undamaged allele. Their harmful effects are expressed, however, when they are homozygous—that is, when the same impaired gene is supplied from both parents. Each human carries a large number of deleterious recessives, most of them unexpressed. When expressed, they range in harmfulness from mild impairment to lethality. A "lethal equivalent" is a set of genes whose aggregate effects, when homozygous, completely prevent the reproduction of the individual they are in (as when they kill the bearer before reproductive age). It is estimated that each of us has at least one to two lethal equivalents worth of deleterious recessives (Bittles & Neel, 1994; Lieberman, 2004). However, the deleterious recessives found in one person are usually different from those found in another.

These facts become socially important when natural selection evaluates the fitness consequences of mating with a nonrelative versus mating with a close genetic relative (for example, a parent or sibling). In reproduction, each parent places half of its genes into a gamete, which then meet and fuse to form the offspring. For parents who are genetically unrelated, the rate at which harmful recessives placed in the two gametes are likely to match and be expressed is a function of their frequency in the population. If (as is common) the frequency in the population of a given recessive is 1/1000, then the frequency with which it will meet itself (be homozygous) in an offspring is only 1 in 1,000,000.

In contrast, if the two parents are close genetic relatives, then the rate at which deleterious recessives are rendered homozygous is far higher. The degree of relatedness between full siblings, or parents and offspring is ½. Therefore, each of the deleterious recessives one sibling inherited from her parents has a 50% chance of being in her brother. Each sibling has a further 50% chance of placing any given gene into a gamete, which means that for any given deleterious recessive found in one sibling, there is a ⅛ chance that a brother and sister will pass two copies to their joint offspring (a ½ chance both siblings have it times a ½ chance the sister places it in the egg times a ½ chance the brother places it in the sperm). Therefore, incest between full siblings renders one-eighth of the loci homozygous in the resulting offspring, leading to a fitness reduction of 25% in a species carrying two lethal equivalents (two lethal equivalents per individual × ⅛ expression in the offspring = 25%). This is a large selection pressure—the equivalent of killing one quarter of one's children. 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).

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*. Incest is rare, but it sometimes happens, and studies of children produced by inbreeding versus outbreeding allow researchers to estimate the magnitude of inbreeding depression in humans. For example, 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).

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 introduce motivational design features that cost-effectively reduce the probability of incest will be strongly favored by natural selection. For species in which close genetic relatives who are reproductively mature are commonly exposed to each other, an effective way of reducing incest is to make cues of genetic relatedness reduce sexual attraction. Indeed, incest is a major fitness error, and so 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. From a selectionist perspective, 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 theory of kin selection showed that selection favors one organism weighting the welfare of another to some extent when the two are genetically related (Hamilton, 1964; Williams & Williams, 1957).

Making Welfare Trade-Offs

To capture this notion of a trade-off, let us define a variable: a "welfare trade-off ratio" or $WTR_{actor, j}$ (Tooby & Cosmides, 2008). 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 standard formulas in evolutionary biology, in which benefits and costs (welfare) are defined as increases and decreases in an individual's reproduction. (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, A, that benefits one's self while imposing a cost on individual j, take beneficial action 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, A, that benefits j at some cost to the self, take costly action 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 function 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 that is higher than to an unreliable partner (Trivers, 1971). 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 that damages you more than the resource gain would benefit you. This is the insight behind the *asymmetric war of attrition* (Hammerstein & Parker, 1982), a game theoretic model that explains why animals in many species engage in 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 natural selection should set your WTR toward j to be a function, at least in part, of your genetic relatedness to j (Hamilton, 1964; Williams & Williams, 1957). To make the insight clearer, let us leave aside factors such as reciprocation and the ability to cause injury, and consider two alternative motivational 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?

Biologists recognize that the second design is strongly favored by selection in species, such as humans, where close genetic relatives frequently interact. If you inherited this design from your ancestors, $r_{self, j}$ expresses the probability that your genetic relative also inherited that same mutation from the same ancestors. That 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 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 altruistic design 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 design spreads is a function of the probability that the same design 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 minute 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 large benefit to the reproduction of a relative at a minor cost to the self.

The selection pressure described by Hamilton's rule does not mean that $WTR_{self,j}$ (henceforth: WTR_j) should be never be higher than $r_{self,j}$—your full sib might also be a great reciprocation partner, or powerful enough to extort you into sacrificing your welfare for his. It means that the designs favored by selection should use genetic relatedness between self and j to place 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 ofextortion. It also means that selection should shape motivation so that the tendency to exploit is restrained by the detection of genetic relatedness (see Equation 15.1).

This analysis predicts that natural selection should have designed the human motivational architecture to embody programs determining how high one's welfare trade-off ratio toward other individuals should be set. These programs should take many variables into account, such as aggressive formidability or value as a cooperative partner. However, kin selection theory tells us that, all else equal, WTR 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 and 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_js 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 a felt experience. A high value of SV_j should be transformed into the felt experience of sexual attraction; a low

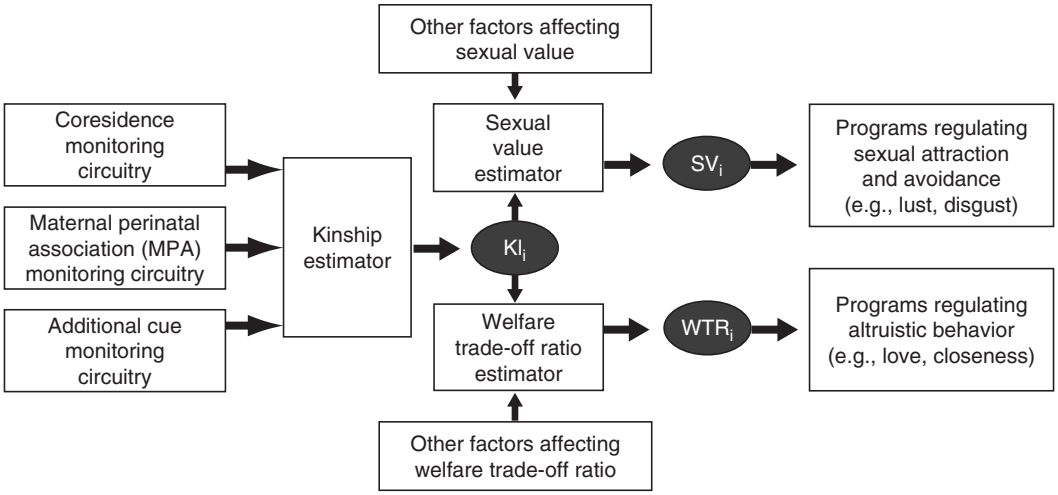

FIGURE 15.1 Model of the human kin detection system, and the internal regulatory variables (black ovals) it computes and regulates. Monitoring circuitry registers cues ancestrally correlated with genetic relatedness (e.g., coresidence duration, MPA). A "kinship estimator" transforms these inputs into 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.

value of SV_j should be transformed into the felt experience of sexual disgust. There does not seem to be a felt experience associated with KI_j per se, only with the variables it regulates.

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 much chocolate 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, the welfare trade-off ratio estimator, that takes into account a specific array of relevant 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 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.

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 co-ordinated 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 specific 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, therefore, we were able to infer which cues the kin detection system uses. 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 sibling. 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 also 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. Nor can the problem of detecting genetic relatives be solved by a domain-general learning mechanism that picks up local, transient cues to genetic relatedness: To deduce which cues predict relatedness locally, the mechanism would need to already know the genetic relatedness of others—the very information it lacks and needs to find. 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 that are relevant in computing 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. So, what cues does the monitoring circuitry register, and how does the kinship estimator transform these into a kinship index?

By considering the statistical 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.

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, the maternal perinatal association cue will not work, because you did not exist at the time your older sibling was born. So to detect older siblings, the mind defaults to a different but weaker cue: How long you coresided with this child during the period of parental investment, from your birth until late adolescence. 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 family lines, with children staying with parents (especially mothers). Under such conditions, the more time one child spends with another, the more closely related they are likely to be. (We found that duration of childhood coresidence is still highly correlated (r = ~.70) with relatedness (i.e., with a sibling being full, half, or unrelated step), even among the postindustrial subjects in our study.)

When the MPA cue is absent, our data show that levels of altruism and sexual aversion toward a particular sibling are set 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 being exposed to 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.

Our data indicate that the kinship estimator computes kinship indexes nonconsciously, and independently of consciously held beliefs about genetic relatedness. A striking example of this from our 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, this would 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. This is evidence for the existence of the kinship estimator program. 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 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

If internal regulatory variables are psychologically and neurally real, then selection could build adaptations whose function is recalibrate them advantageously. We have been testing the hypothesis that the adaptive function of certain emotion programs—anger, gratitude, and guilt, for example—is to recalibrate internal regulatory variables in one's own brain and in the brains of other people (Sell, 2005; Sell, Tooby, & Cosmides, in prep. a, b; Tooby & Cosmides, 2008; Sznycer, Price, Tooby & Cosmides, in prep.). Indeed, we think the WTR regulatory variable lies at the core of each of these emotion programs. We will use anger to illustrate the usefulness of framing emotions as programs that use and operate on regulatory variables.

Specifically, we propose 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. Its primary functional goal is to upregulate the WTR in the brain of the target of the anger, so that the target places more weight on the welfare of the angered individual. The anger program is designed to bargain for better treatment by deploying two negotiative tools: (1) in cooperative relationships, threats to withdraw benefits (or actually withdrawing them), and (2) in neutral or antagonistic relationships, threats to inflict costs (or actually inflicting them). The computational logic of anger orchestrates the advertisement of these contingencies through emotional display (e.g., anger face), verbal communication (e.g., threats), or action (e.g., striking, abandoning a relationship).

Before proceeding, it is important to recognize that the programs in an organism should be designed to trade-off its welfare differently when the organism is being observed than when it is not. When one's acts are being monitored by an individual whose welfare is affected, that individual can respond by retaliating or rewarding the actor. But when one's acts are private and will not be known to impacted individuals, selection should produce a system that weights their welfare only insofar as it is in the actor's intrinsic interest to do so. Hence, there should be algorithms that compute two parallel, independent WTRs for each social other: (1) an intrinsic WTR, which sets a lower boundary on how much weight the actor places on the other party's welfare even when the actor's choices are not being observed; and (2) the public or monitored WTR, which guides an individual's actions when the recipient (or relevant others) can observe them. The kinship index is one variable that sets intrinsic WTRs. Monitored WTRs are set by aggressive intimidation and reciprocity. Anger is designed to modify monitored WTRs.

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 (see 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

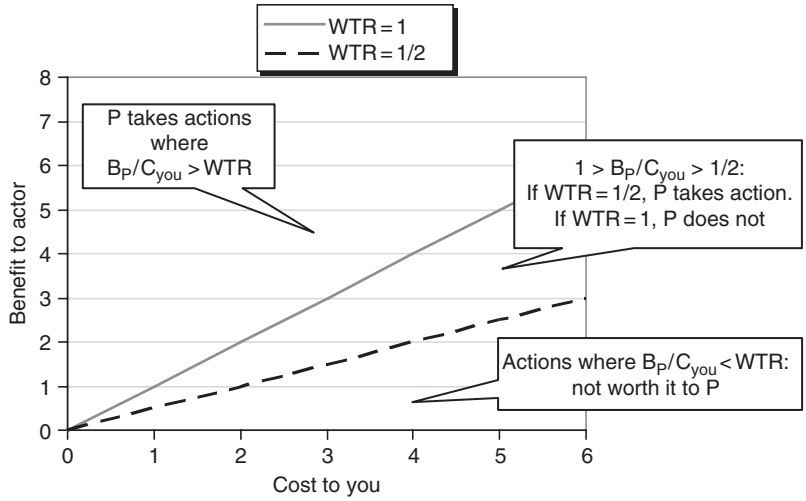

FIGURE 15.2 An actor's welfare trade-off ratio (WTR) toward you can be inferred by observing how large a cost that individual is willing to impose on you for how small a benefit gained. The gray line represents a WTR of 1, meaning that the actor values your welfare as heavily as his or her own. The black dashed line represents a WTR of ½, 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 ½, but not if it were 1. Raising an individual's WTR toward you allows you to avoid these costs.

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 ½, P values your welfare only half as much as his own; that is, P will take actions for which $B_P/C_{you} > ½$. This means there is a set of cost-imposing actions, ones for which $½ < B_P/C_{you} < 1$, that P will take when his $WTR_{you} = ½$, but not when his $WTR_{you} = 1$ (see Figure 15.2). You will be spared more of these costs to the extent there is some way of raising P's WTR toward you.

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 (Cosmides & Tooby, 2005; Gurven, 2004; Trivers, 1971), coalitional (group) cooperation (Tooby et al., 2006), food sharing as a form of risk pooling (Kaplan & Hill, 1985), and 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 inflict costs on him if he does not. Like most other species, humans sometimes use aggression to induce others to sacrifice their own welfare for the aggressor's. 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 fight for it (Hammerstein & Parker, 1982; Maynard & 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, more 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, genetic relatedness, or value as a reciprocator, respectively). 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, as well as the kinship index associated with that individual and a variable expressing that individual's formidability relative to one's own. Indeed, your welfare trade-off ratio estimator should be designed by selection 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 that 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 would be if you were to respond. 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, Tooby, & Cosmides, in prep. a, b).

Anger as a Negotiation over WTR Values

Social behavior publicly advertises WTRs. Given the ability to estimate the consequences of actions on welfare, the costs and benefits they impose on oneself and others, one can infer one person's WTR toward another from his or her actions. For example, assume that you observe a person named Aaron taking an action that inflicts a cost of 4 (notional) units on you to gain a benefit of 1 unit for himself. From this, you can infer that Aaron's $WTR_{you} \leq ¼$. ($B_{Aaron}/C_{you} = ¼$; Equation 15.1 means Aaron would take this action only if $B_{Aaron}/C_{you} \geq WTR_{you}$.)

Most theories of anger recognize that humans typically get angry when someone imposes a cost on them; and, all else equal, the larger the cost, the more angry the person becomes. But the recalibrational theory of anger further 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. (Thus, humans may become angry when they are benefited—but less than they feel entitled to.) 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.

Assume that your WTR estimator has computed, based on the nature of your relationship, that Aaron's equilibrium WTR toward you should be ½. You then see him ruin your expensive scarf, imposing a cost of 4 units on you. According to the recalibrational framework, whether you become angry should depend on how much Aaron benefited by using your scarf. 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 $WTR_{you} \leq ¼$. This is less than the equilibrium value of ½, 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 great—e.g., 24 units. This action should not trigger anger in you, despite the fact that it inflicts the same cost: In this case, Aaron's action is still consistent with a WTR_{you} of ½. 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. The anger system should not be activated under such circumstances, because the events do not reveal a WTR that needs to be recalibrated.

With these predictions in mind, we conducted experiments that held the cost imposed on the subject constant, while 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, Tooby, & Cosmides, in prep. a).

According to the recalibrational theory of anger, the program monitoring WTRs is activated when someone imposes a cost on you (or fails to provide an expected benefit). If the detection component inside the anger program 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 Anger Program Orchestrating Cooperation

Assume, for this example, that 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 asignal activating the anger program and its regulation of cooperation. This program structures arguments and other communicative acts according to a functional logic of anger, each of whose features is 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 anger program 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 anger program 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 anger program'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 anger program'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 anger program 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 p[illegible]n'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 program, returning the cooperation system to normal mode and deactivating the aggression system. (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 [illegible] 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 program 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 anger program 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: The anger program recalibrates 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, economizing on unrewarding social outlays.

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 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 Anger Program Orchestrating Aggression

Another way to negotiate WTRs is by threatening harm, so there are circumstances in which the anger program will regulate aggression. However, if aggression is used exploitatively inside a cooperative relationship, then the cooperative partner should avoid the exploiter (when possible), dissolving the relationship. Withdrawal of cooperation is a less expensive bargaining tool than aggression. In contrast, non-cooperators have no cooperation to threaten to withdraw. Hence, threats of aggression should be more common in noncooperative relationships, while threats of downregulating cooperation should be more common in cooperative ones.

Threatening harm is a more effective tactic the more capable the threatener is of inflicting harm at low relative cost. Therefore, anger should more easily trigger aggression as a negotiative tool in more formidable individuals than in weaker ones. This effect should be particularly pronounced in men, because in humans, males are stronger and tend to pre-empt force as a social tool. Although absolute levels of aggression vary between cultures, within cultures women are far less likely than men to resolve conflicts by using physical force (Campbell, 2002; Daly & Wilson, 1988).

Now, assume that circumstances force you and Aaron to interact, but you do not have a 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 activate the anger system in its aggressive mode. When this happens, the anger program should motivate specific actions and goals, each designed to solve a recalibrational problem. For example:

Problem 7: Aaron's estimate of your relative ability to inflict costs on him—his formidability index with respect to you (FI_{you})—is too low.
Solution: The anger program activates a specific goal: to recalibrate the FI_{you} regulatory variable in Aaron's brain. It should motivate actions that demonstrate your ability to harm him, displays such as chest thrusting, pushing, or breaking things.

If these 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 coalition-derived 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 signal how much you value a resource, or how large a cost Aaron's action imposed on you. That is, they can serve as communicative function as well, providing a solution to Problem 2 above (Aaron's mis-estimate 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 anger program 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 consider it a bluff. Therefore, this motivation is more likely to be activated when you actually are more formidable than Aaron, or when external constraints would prevent a fight from actually breaking out (e.g., friends or authorities are present who will hold you back).

Indeed, the logic of negotiation through the threat or actuality of inflicting costs is general, regardless of whether the costs are inflicted through violence, social manipulation, or other means. Different kinds of power have different effects, and so we expect them to be encoded by different regulatory variables (formidability being different from status, for example).

When the anger program is orchestrating aggression, it 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 9: 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 anger program 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 cause formidability-based WTR recalibration, 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. First, the anger program should be easier to trigger in people who are stronger (more formidable) because they can physically inflict more costs than weaker people can, enforcing 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 the angered individual expected from others (based on an implicit computation of a power- or reciprocity-based equilibrium). 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 ½.

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 another 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 (two to five 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 placed too little weight on your welfare—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 anger program orchestrates cooperation, the approach response is to exchange information, argue, and, if necessary, withdraw cooperation, or even terminate the relationship and avoid the individual. When the anger program orchestrates aggression, 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.

CONCLUSIONS

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 a satisfying level of abstraction can still be achieved while 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] for yourself! I've been so 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 (make X suffer; put more weight on Y's welfare) that get filled in with concrete content depending on the situation.

Just as psychophysics allowed the principled study of perception, this framework opens a principled gateway into the scientific study of feeling—a previously intractable topic. According to this approach, conscious focus on a situation feeds new information through the architecture that triggers procedures designed to register or recalibrate the array of regulatory variables the new information is relevant to (Tooby & Cosmides, 2008). Next, signals of the significant changes in (some of) these variables are fed back into conscious awareness—presumably as a method to broadcast them to other programs they are relevant to. This cycle often appears to lead to chain reactions (as with grief, anger, and betrayal), where downstream programs are set off in their turn by receipt of further recalibrational information, triggering them then to broadcast their own contributions into conscious awareness. That is, the tapestry of felt experience that is directly elicited by the objects of awareness are, we think, annotations and evaluations about those objects in terms of changes in the internal regulatory variables relevant to them (that person is stronger than I thought; my sister is dead; this person was surprisingly kind to me; acacia beetles taste better than I thought). The demand for feeling computation often exceeds available bandwidth. When this happens, the individual spends time engaging in a particular form of behavior designed to maximize feeling

computation, by suspending other activities that would distract from attention to the internal panorama of endogenous responses to new information. In short, feeling is a form of computation in which the values of regulatory variables are set, recalibrated, broadcast through the architecture, and output into awareness so that they can be fed into other programs designed to use them.

Finally, models provided by evolutionary biology can help identify internal regulatory variables whose computational role in our evolved motivational architecture we might not otherwise suspect. Indeed, they provide us with the experimental guidance necessary for constructing abstract yet fine-grained maps of the responses our motivational systems were evolutionarily designed 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.

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., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology*, *13*, 171–195.

Boyd, R. (1988). Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethology and Sociobiology*, *9*, 211–222.

Buss, D. (2005). *The handbook of evolutionary psychology*. New York: Wiley.

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 & Sociobiology*, *10*, 51–97.

Cosmides, L., & Tooby, J. (2000). Consider the source: The 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. (2000). 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, New Jersey: Erlbaum.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, New Jersey: Wiley.

Daly, M., & Wilson, M. (1983). *Sex, evolution, and behavior, 2nd edition*. Boston: Willard Grant.

Daly, M., & Wilson, M. (1988). *Homicide*. Chicago: Aldine.

Dawkins, R. (1979). Twelve misunderstandings of kin selection. *Zeitschrift fur Tierpsychologie*, *51*, 184–200.

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.

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.

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.

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.

Kirkpatrick, L. A., & Ellis, B. J. (2001). An evolutionary-psychological approach to self-esteem: multiple domains and multiple functions. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell handbook of social psychology: Interpersonal processes* (pp. 411–436). Oxford, UK: Blackwell Publishers.

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. (2004). Mapping the cognitive architecture of systems for kin detection and inbreeding avoidance: The Westermarck hypothesis and the development of sexual aversions between siblings. Lieberman, Debra Lyn; Dissertation Abstracts International: Section B. *The Sciences & Engineering*, *64*(8–B), 4110.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, *445*, 727–731.

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.

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 tradeoff 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.

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.

Sznycer, 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. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

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.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of emotions and their relationship to internal regulatory variables. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions, 3rd edition*. New York: Guilford.

Tooby, J., Cosmides, L., & Barrett, H. C. (2003). The second law of thermodynamics is the first law of psychology: Evolutionary developmental psychology and the theory of tandem, coordinated inheritances. *Psychological Bulletin*, *129*(6), 858–865.

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, New Jersey: 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.