Principles of allostasis: optimal design, predictive regulation, pathophysiology and rational therapeutics.

Peter Sterling

IN: Allostasis, Homeostasis, and the Costs of Adaptation

J. Schulkin

Cambridge University Press
2004

Principles of allostasis: optimal design, predictive regulation, pathophysiology and rational therapeutics.[1,2]

Peter Sterling

Introduction.

This chapter compares two alternative models of physiological regulation. The first model, *homeostasis* ("stability through constancy"), has dominated physiology and medicine since Claude Bernard declared, "All the vital mechanisms...have only one object – to preserve constant the conditions of ... the internal environment". His dictum has been interpreted literally to mean that the purpose of physiological regulation is to clamp each internal parameter at a "setpoint" by sensing errors and correcting them with negative feedback (Cannon, 1935: Figure 1). Based on this model physicians reason that when a parameter deviates from its setpoint value, some internal mechanism must be broken. Consequently they design therapies to restore the "inappropriate" value to "normal".

The homeostasis model has contributed immeasurably to the theory and practice of scientific medicine, so to criticize it might almost seem absurd. Yet, all scientific models eventually encounter new facts that do not fit, and this is now the case for homeostasis. In physiology, evidence accumulates that parameters are *not* constant. And their variations, rather than signifying error, are apparently designed to *reduce* error. In medicine, major diseases now rise in prevalence, such as essential hypertension and type 2 diabetes, whose causes the homeostasis model cannot explain. For in contrast to the hypertension caused by a constricted renal artery and the diabetes caused by immune destruction of insulin-secreting cells, these newer disorders present no obviously defective mechanism. And treating these diseases with drugs to fix low-level mechanisms that are not broken turns out not to work particularly well. The chapter will expand upon each of these points.

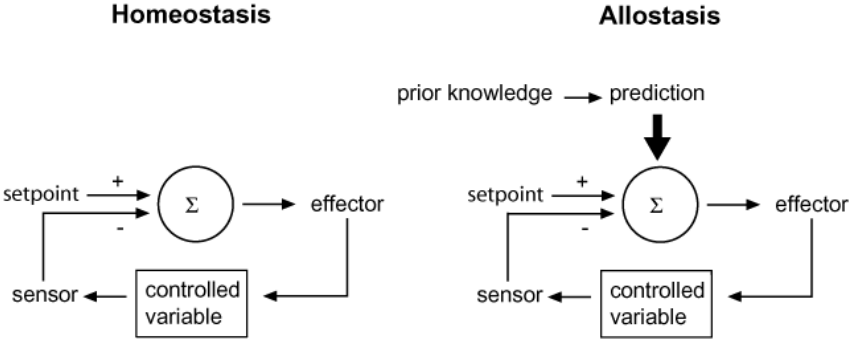

Figure 1. Alternative models of regulation. *Homeostasis* describes mechanisms that *hold constant* a controlled variable by sensing its deviation from a "setpoint" and feeding back to correct the error. *Allostasis* describes mechanisms that *change* the controlled variable by predicting what level will be needed and overriding local feedback to meet anticipated demand.

The second model, *allostasis* ("stability through change"), takes virtually the opposite view. It suggests that the goal of regulation is *not* constancy, but rather, fitness under natural selection. Fitness constrains regulation to be efficient, which implies preventing errors and minimizing costs. Both needs are best accomplished by using prior information to predict demand and then adjusting all parameters to meet it (Figure 1). Thus allostasis considers an

[1] This essay is dedicated to the memory of Howard A. Schneiderman, who recruited me to experimental biology and bailed me out of a Mississippi jail.
[2] Collected essays on this and related topics available at http://retina.anatomy.upenn.edu/allostasis/allostasis.html

unusual parameter value, not as a failure to defend a setpoint, but rather as a response to some prediction. And the model attributes diseases such as essential hypertension and type 2 diabetes to sustained neural signals that arise from unsatisfactory social interactions. Consequently the allostasis model would redirect therapy, away from manipulating low-level mechanisms, toward improving higher levels in order to restore predictive fluctuation – which under this model is the hallmark of health.

This essay comprises six main sections. The first provides a capsule history of the allostasis model, which by now extends back over 40 years. The second section offers a brief critique of the homeostasis model, focusing on blood pressure because of its broad medical significance. The third section presents key principles of allostasis. Here are introduced recent concepts of optimal matching and adaptive regulation which are then used to reconsider problems of human physiology, such as blood pressure. The fourth section describes how allostasis depends on higher neural mechanisms, and the fifth section suggests that these mechanisms interact with certain aspects of modern social organization to generate some of the major modern diseases. The last section treats the question of where to intervene.

Origins of allostasis.

For several decades, I combined research and teaching in neuroscience with social activism. In the mid-1960s, canvassing door-to-door in African-American ghettos such as Central and Hough in Cleveland, Ohio, I noticed that many people who answered my knock were partially paralyzed – faces sagging on one side, walking with a limp and a crutch. The cause was 'stroke', a rare affliction in my own community, and one that I never encountered later when canvassing in white, upper-class Brookline, Massachusetts. What caused so many strokes, I wondered, and how might they be connected to Cleveland's racial segregation? Arriving around 1970 at the University of Pennsylvania, I found that Joseph Eyer, another biologist/activist, had assembled clear epidemiological evidence that stroke and heart disease, and their precursor, hypertension, all accompany various forms of social disruption, including migration, industrialization, urbanization, segregation, unemployment, and divorce (Eyer and Sterling, 1977; Eyer 1975, 1977).

While publishing the epidemiological data, we began to investigate the possible biological mediators. The fury in Hough – which during the summer of 1966 exploded in riots and occupation by National Guard troops – would tend to activate Cannon's well-known, 'fight-or-flight' system (sympathetic nerves - adrenal medulla) and Selye's 'stress' system (hypothalamo-pituitary-adrenal cortex). But we were astonished by new evidence from fluorescence microscopy that all blood vessels are richly innervated by catecholamine nerve fibers, and new evidence from electron microscopy that most endocrine cells are also innervated. For example, sympathetic nerves contact the kidney cells that secrete renin, and parasympathetic nerves contact the pancreas cells that secrete insulin. Recent work has shown that nerves even contact cells that form bone and scavenger cells (macrophages) that serve inflammation and immune surveillance (Takeda et al., 2002; Bernik et al., 2002; Blalock, 2002; Flier, 2000; Tracy, 2002). This suggested that the brain has close access to essentially every somatic cell.

Furthermore, John Mason measured multiple hormones in awake, behaving monkeys – and found concerted shifts that made functional sense. A mild demand for focused attention raised hormones associated with catabolism and suppressed those associated with anabolism (Mason, 1968, 1971, 1972). Furthermore, prolonging these demands caused sustained elevations of blood pressure (Harris et al., 1973). Mason concluded that the broad metabolic patterns over

short and long time scales, and under mild as well as emergency conditions, are controlled by the brain. Subsequently, myriad studies of neuro-endocrine control have supported this conclusion (Schulkin, 1999).

Back then, standard medicine attributed essential hypertension and atherosclerosis to excessive consumption of salt and fat – as though what people choose to eat were unrelated to their internal physiological and mental states. So it was compelling to learn that the peripheral hormones that raise blood pressure, such as angiotensin, aldosterone, and cortisol, also modulate brain regions that stimulate hunger for sodium (reviewed by Schulkin, 1999; Fluharty, 2002). Similarly peripheral hormones that increase catabolism, such as cortisol, also modulate brain regions that stimulate hunger for energy-rich substrates – fat and carbohydrates (reviewed by Saper et al., 2002; Schulkin, 1999; Schwartz et al., 2000). Of course, such findings would not have surprised Pavlov, who had early demonstrated the brain's anticipatory control over many phases of digestion, nor Curt Richter, who had connected specific hungers to physiological regulation (Schulkin, 2003a, b).

But to a social activist this seemed immensely relevant: if the brain regulates both physiology and its supporting behavior, then treatments directed just at the peripheral physiology will tend to be countered by the behavior. So, rather than clamp blood pressure at some "normal" value by diuretics, vasodilators, and beta-adrenergic antagonists (the main antihypertensive drugs of the 1970s -'80s), wouldn't it be better to reduce social and psychological disruption? That is, address the higher-level signals that stimulate both the physiology and the behavior? We found a perfect example at the Philadelphia Child Guidance Clinic.

Diabetic children who experience chronic bouts of ketoacidosis had been widely treated with beta-adrenergic antagonists. This often proved ineffective, and it was hypothesized that the metabolic disturbance might be induced by parental conflict expressed through the child ("who is right, Daddy or Mommy?"). This was directly observed in "stress interviews". The parents' fatty acid levels would rise but soon return to baseline; whereas the child's would remain elevated for hours (Figure 2A). Clearly, potent psychological demands were driving multiple physiological mechanisms to override the beta-adrenergic mechanism. Salvador Minuchin, the clinic director, described this as a poignant demonstration that "behavioral events among family members can be measured in the bloodstream of other family members" (Minuchin, 1974).

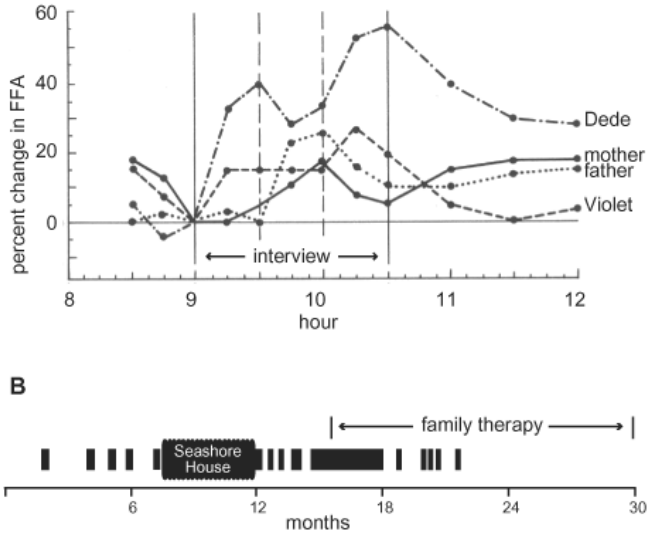

Figure 2. Parental conflict modulates a child's blood chemistry.
A. While parents expressed conflict during an interview, free fatty acid levels rose in all family members. Initially the children, both diabetic, watched through a one-way mirror. At 10 o'clock, they entered the room, whereupon both parents tried to enlist Dede on their side while Violet remained aloof. Violet's free fatty acid levels followed the parents', but Dede's were greatly elevated. Reprinted from Minuchin, 1974.
B. Child had been hospitalized (▌) for emergency treatment of ketoacidosis 23 times over two years, and beta-blocker treatment of her "superlabile" diabetes was unsuccessful. Family therapy that encouraged the parents to express their disagreements directly (rather than through the child) prevented further relapse. Modified from Baker et al., 1974.

Such children easily stabilized in the hospital but, upon reentering the family, soon relapsed. When the parents were helped to resolve their marital conflicts directly, the children stabilized at home without the beta-blocker (Figure 2B**;** Baker et al., 1974). This example of successful intervention *between people*, rather than between nerve and liver, seemed of broad socio-medical significance (Sterling and Eyer, 1981). But the idea upon which it rests, that the brain controls human physiology, still remains largely outside the realm of standard teaching in biology and medicine.

Later, while summarizing this material for another essay collection, it hit me that when you *name* an idea, it has a better chance. So, we coined a new word, "allostasis", to emphasize two key points about regulation: *parameters vary*, and *variation anticipates demand* (Sterling and Eyer, 1988). The idea did spread to some degree, largely through the prolific writings of experts on stress and neuroendocrine regulation, such as McEwen, Schulkin, Sapolsky, Koob, and their colleagues (Sapolsky, 1998; McEwen, 2002; Koob and Le Moal, 2001). But even these proponents of allostasis have been somewhat reluctant to abandon homeostasis as the core theory of regulation and have tended to view allostasis as a modulator of homeostatic mechanisms. Some have simply equated it with "stress" or "fight-flight" response and suggested that it is an anachronism. For example, "Allostasis has evolved as the response for running away from a predator, escaping acute danger, or fighting off a threat… However, a defense system that has its roots in an archaic fish can be absurd in a modern human." (Elbert and Rochstroh, 2003). If this were allostasis, it would be entirely justified to discount it as just a fancy word applied to an old idea (Dallman, 2003).

But the allostasis model has a more radical intent – to *replace* homeostasis as the core model of physiological regulation. There are solid scientific reasons: the allostasis model connects easily with modern concepts in sensory physiology, neural computation, and optimal design. Also, this model can begin to comprehend what homeostasis cannot: the main diseases of modern society, such as hypertension, obesity/diabetes, and drug addiction. There are also practical, socially relevant reasons: the allostasis model suggests a different goal for therapeutics and thus a different direction for medical education and treatment. Consequently, this essay starts by assuming that the original conjecture is proved – that physiology is indeed sensitive to social relations. The evidence for this is now vast and thoroughly summarized by McEwen (2003) and Sapolsky (1998). Thus I first describe some difficulties with the homeostasis model and then set out some core principles of the allostasis model that might justify the fancy name.

Problems with homeostasis as the primary model for regulation.

*Constancy is **not** a fundamental condition for life.*

It seems past time to acknowledge that when Bernard declared constancy to be the sole object of all vital mechanisms, he went too far. Most biologists now agree that the true object of all the vital mechanisms is not "constancy" but survival to reproduce. So what all the vital mechanisms actually serve is reproductive success under natural selection. Moreover, there is nothing magical about "constancy". We now know that the conditions of life span amazing extremes: thermophilic bacteria can thrive at 100° C., and the limit for their successful culturing extends to 113° C! (Hochachka and Somero, 2002). Cell temperatures in the desert can fluctuate by nearly 100° C, and even in complex metazoans the pH of blood and cytosol varies systematically with temperature (Hochachka and Somero, 2002).

Of course, some parameters are regulated quite closely. For example, the mammalian brain tolerates only small fluctuations in oxygen, glucose, temperature, and osmotic pressure. An acute insult that drives any one of these parameters beyond its design limit can trigger cascades of positive feedback that are quickly lethal. And such catastrophic departures from stability certainly require emergency treatment directed at low level processes (Buchman, 2002). But the purpose of such tight regulation may not be to defend "constancy" in the abstract. Rather, it may simply reflect specific design choices that optimize overall mammalian performance for successful competition.

In fact mammalian brain tissue, such as the intact retina or a slice of cerebral cortex, functions for hours in a simple medium at *room temperature*. A neuron's sensitivity is lower than for the optimal 37° C by two-fold for each ten degrees (Dhingra et al., 2003), but this follows the temperature sensitivity of most biochemical reactions. So the mammalian brain's normal operating temperature apparently reflects an early design decision: to move fast, we must think fast. This decision had myriad consequences; for example, to move fast, we must also *see* fast. This requires the photoreceptors to be small, which in turn sets the design of retinal circuits (Sterling, 2003). In short, close regulation of human cerebral temperature does not exemplify *the* condition for preserving *all* life – it is just one condition set by a particular design.

A mean value need not imply a setpoint but rather the most frequent demand.

It also seems past time to reevaluate the core hypothesis of the homeostasis model, that the average level of each parameter represents a "set point" which is "defended" against deviations (errors) by local feedback (Figure 1). This model captured much of the experimental truth in a simple "preparation" – such as an isolated organ or an animal whose brain has been silenced by anesthesia or decerebration – which were the primary experimental models for over 100 years. But regulation under natural conditions presents a response pattern that the homeostasis model cannot easily explain.

Consider the record of arterial blood pressure measured continuously over 24 hours in a normal adult (Figure 3). Far from holding steady, pressure fluctuates markedly around 110/70 mm Hg for two hours. Then in correlation with identified external stimuli and mental states, it varies more extremely. As the subject dozes in lecture, pressure falls to 80/50. When he is jabbed with a pin, pressure spikes to 150/70; then, having recognized the joke, he again relaxes, and the pressure sinks to 80/50. During sexual intercourse, pressure spikes to 170/90 and then falls profoundly during sleep to ~70/40 with one hour as low as 55/30. In the morning pressure steps up nearly to its level during sex and remains high for hours.

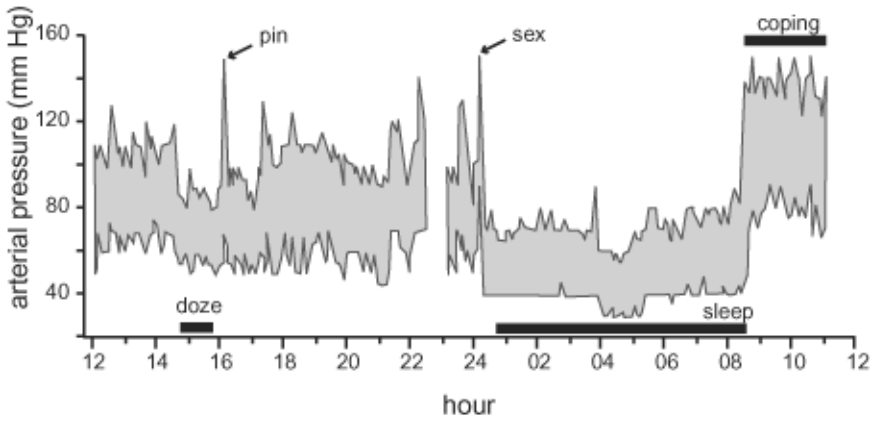

Figure 3. Arterial pressure fluctuates to meet predicted demand.
Pressure was plotted in a normal adult at 5 minute intervals over 24 hours. Note that pressure spends about equal time above and below the steady daytime level. This pattern suggests, not defense of a setpoint, but rather responsiveness to rising and falling demand. Upper trace, systolic; lower trace, diastolic. Redrawn from Bevan et al., 1969.

This record contains no hint that blood pressure is defended at particular setpoint. Quite the contrary, it fluctuates markedly and does so on multiple time scales – minutes, seconds, and hours. There are elevations, both brief and sustained, above the most frequent level. And there are also similar depressions below the most frequent level. If this level truly represented a "setpoint", we might expect it to fluctuate only mildly except when a particularly arousing signal would drive it higher (fight-or-flight). But the pressure spends about as much time *far below* the most frequent level as above it, and this is not predicted by a model of setpoint + arousal-evoked elevation. More parsimoniously, the record suggests that pressure is regulated to match anticipated demand, rising to certain signals and falling to others. This implies that the most frequent value, 110/70, occurs not because pressure is clamped there, but because that value satisfies the most frequent level of demand (see Figure 5).

Indeed, were pressure actually clamped at an average value, it would match some specific need only by sheer accident. This is true for all states and all parameters: average values are useless. The essential need is to occupy distinctly different states and to move flexibly between them. But how could this occur, given local negative feedback mechanisms that do tend to resist fluctuations?

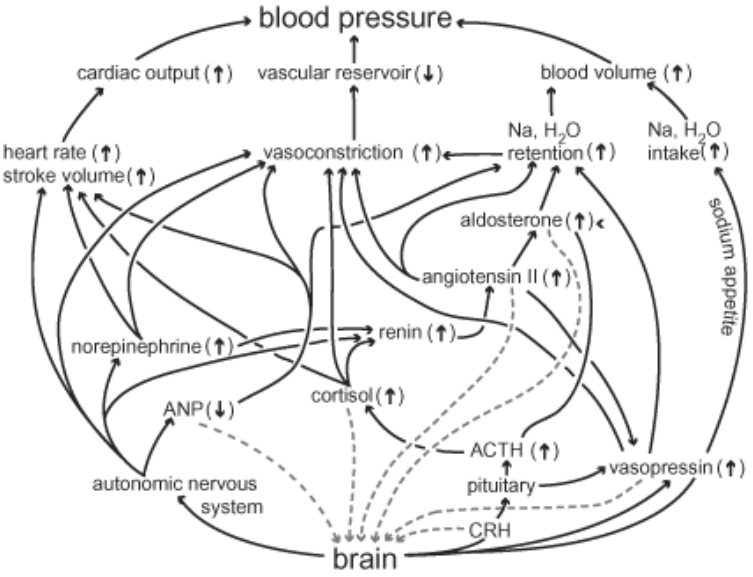

Figure 4. The brain sets blood pressure via multiple, mutually reinforcing mechanisms. Negative feedback mechanisms are acutely overridden. When demand persists, all mechanisms are reset to operate at the new level. Most hormones illustrated here are also sensed by brain (dashed arrows) in specific regions that control behaviors that support increased pressure. Thus, aldosterone and angiotensin II are sensed by brain regions that enhance salt appetite and drive salt-seeking behavior. Abbreviations: CRH, corticotrophic releasing hormone; ACTH, adrenal corticotrophic hormone; ANP, atrio-natiuretic peptide. Modified from Sterling and Eyer, 1988.

Once the brain predicts the most likely demand, it resets the blood pressure to match. To do so, the brain directly modulates all three primary effectors: nerves signal the heart to pump faster, blood vessels to constrict, and kidneys to retain salt and water. These direct neural messages are reinforced by additional signals acting in parallel (Figure 4). For example, the neural system that excites the primary effectors also releases multiple hormones that send them same message. Hormones signaling the opposite message are suppressed. This pattern: multiple, mutually reinforcing signals acting on multiple, mutually reinforcing effectors, overrides the various feedbacks that oppose change.[3] Recognizing such fluctuation, some authors have

[3] It is probably no accident that the error-correction model that Bernard adopted for physiology mimicked the simple device that inaugurated the 19th Century's industrial technology (the speed governor on Fulton's steam engine). But machines have evolved, and the 21st Century automobile now preempts driver errors. The myriad sensors in a BMW (~ 100) relay data to a central mechanism (computer chip) that calculates the power and braking that each wheel needs to optimize stability and skid-resistance. Data from other sensors are centrally integrated to control fuel,

proposed the idea of shifting setpoints termed, "rheostasis" (Mrosovsky, 1990). Shifting setpoints might seem to describe certain cases, e.g. sustained elevation of body temperature in fever, but even here temperature is responding to specific signals that fluctuate adaptatively.

The same is true for essentially *all* parameters: temperature, blood distribution, hormone levels, and so on. All fluctuate with different amplitudes and time constants, and these fluctuations all share a single goal. Yet the goal is not constancy, but coordinated variation to optimize performance at the least cost. This is the core idea of allostasis, whose essential principles are addressed next.

Principles of allostasis (predictive regulation).

This section discusses six interrelated principles that underlie allostasis: **(1)** organisms are designed for efficiency; **(2)** efficiency requires reciprocal trade-offs; **(3)** efficiency requires predicting what will be needed; **(4)** prediction requires each sensor to adapt its sensitivity to the expected range of input; **(5)** prediction requires each effector to adapt its output to the expected range of demand; **(6)** predictive regulation depends on behavior whose neural mechanisms also adapt.

Organisms are designed for efficiency.

Organisms must operate efficiently. Beyond escaping predators and resisting parasites, they must compete effectively with conspecifics. If you encounter a bear while hiking with a friend, you need not outrun the bear – just your friend. So natural selection sculpts every physiological system to meet the loads that it will most likely encounter in a particular niche plus a modest safety factor for the unusual load. No system can be "overdesigned" because robustness to very improbable loads will slow the organism and raise fuel costs. Nor can a system be "underdesigned" because, if it fails catastrophically to a commonly encountered load, well, that's it. In effect the organism resembles an elevator cable – which must be just sufficiently robust to prevent the cancellation of the manufacturer's insurance (Diamond, 1993).

It follows that all internal systems should mutually match their capacities. Thus our intestinal absorptive capacity supplies sufficient fuel for our most likely energy need – with modest excess to meet unusual demands (Hammond and Diamond, 1997). And our lung and circulatory capacities supply sufficient oxygen to burn the available fuel; and our muscles contain sufficient mitochondrial capacity to provide an adequate furnace (Weibel, 2000). Clearly it would be inefficient for an organ to provide more capacity than could be used downstream, or for an organ downstream to provide more capacity than can be supplied from upstream. This aspect of organismal design, where physiological capacities optimally match, is termed "symmorphosis" (Taylor and Weibel, 1981). It holds for digestive, respiratory, and muscular systems, and also for neural systems (Weibel, 2000; Diamond, 1993; Sterling, 2003).

Efficiency requires reciprocal trade-offs.

Efficiency requires that resources be shared. Otherwise, each organ could meet an unusual demand only by maintaining its own reserve capacity. To support this extra capacity would require more fuel and more blood – and thus more digestive capacity, a larger heart, and so on, thereby creating an expensive infrastructure to be used only rarely. Consequently, organs

---

oxygen, and spark timing for each cylinder to optimize fuel consumption at each power level. This resembles biology, where changing gait maximizes efficiency at different speeds (Alexander, 1996; Weibel, 2000). In short, for a car with a "brain", predictive regulation produces better stability and greater efficiency.

can trade resources – that is, make short-term loans. Regulation based on reciprocal sharing between organs is efficient, but for several reasons it requires a centralized mechanism: to continuously monitor all the organs; to compute and update the list of priorities; and to enforce the priorities by overriding all the local mechanisms (Figure 4).[4]

For example, skeletal muscle at rest uses about 1.2 liters of blood per minute (~ 20% of resting cardiac output), but during peak effort it uses about 22 l/min (~90% of peak cardiac output), an 18-fold increase. Much of the extra blood comes from increased cardiac output, but that is insufficient. And although tissues may store fuel, such as glycogen and fatty acids, they cannot store much oxygen. Nor would it be useful to maintain a large reservoir of de-oxygenated blood because peak demand completely occupies the pulmonary system's capacity to re-oxygenate. So a reservoir of de-oxygenated blood would require a reservoir of lung, heart, etc. In turn, these would require increased capacities for digestion, absorption, excretion, and cooling. Consequently for an unstorable resource, subject to variable demand, it is most efficient to share. So, at peak demand about 10 % of the total flow to muscle is *borrowed* (Weibel, 2000).

The loan cannot come from the brain, which requires a constant supply, that if interrupted for mere seconds causes loss of consciousness. So muscle borrows from the renal and splanchnic circulations, whose individual shares of cardiac output both drop from about 20% to 1%, and whose absolute supplies fall by 4- to 5-fold (Weibel, 2000, his Figure 8.6). The skin circulation also contributes. Kidney, gut, liver, and skin can generally afford to lend for the short-term – depending on circumstances. For example, skin can postpone re-oxygenation – but exercise in a warm environment may require flow to skin for cooling. Gut can also postpone re-oxygenation, but following a meal it requires blood to transport digests into the portal circulation and thus needs to reclaim a higher share.

Reciprocity requires central control.

The brain, though it represents 2% by weight in a 70 kg man, requires 20% of the resting blood flow. This proportion is so great that when a given brain region increases activity, the extra blood is requisitioned, not from somatic tissues, but from other brain regions (Lennie, 2003). Thus within the brain itself, resources are reciprocally shared.

Because the needs of muscle, gut, and skin can be irreconcilable, appropriate trade-offs between them (and all the organs) must be calculated. This requires a central mechanism, the brain, which must also enforce a specific hierarchy of priorities and shift them as needs change. When muscular effort is urgent, but you have just eaten and the environment is warm, the brain triggers a vomiting reflex; when cooling is more urgent than effort, the brain may reduce the priority for erect body and trigger the vaso-vagal reflex ("fainting"): the heart slows, vessels dilate, blood pressure falls, and muscle tone collapses. In short, the brain must decide the conditions for each loan and set the schedule for repayment. Furthermore, because such conflicts potentially threaten overall stability (survival), these solutions are accompanied by unpleasant sensations, such as nausea and dizziness, which the brain also provides. These sensations are vividly remembered in order to reduce the likelihood of repetition.

[4] Again, industrial analogies seem pertinent: consider the efficiencies achieved by sharing electricity in a power grid and by rapidly redistributing inventory in a factory system. This type of efficiency requires continual, rapid updating of information about current demand, plus prior knowledge of how demand will probably change with factors such as temperature, time of day, season, world market, and so on.

Efficiency requires predicting what will be needed.

We have already seen that blood pressure fluctuates according to match the ever-shifting prediction of what might be needed (Figure 3). This is true for essentially all physiological mechanisms. Consider an additional example, control of blood glucose by insulin.

This is usually presented as a core example of homeostasis: ingested glucose raises the blood level, stimulating pancreatic beta cells to release insulin, which stimulates muscle and fat cells to take up the glucose and restore blood levels to the standard ~ 90 mg/dl. And indeed a pancreas placed *in vitro* and exposed to glucose will release insulin. But when an intact person sits down to a meal, the sight, smell, and taste of food predict that blood glucose will soon rise, and this triggers insulin release via neural mechanisms well before freshly ingested glucose reaches the blood (Schwartz et al., 2000). This anticipatory pulse of insulin signals muscle and fat cells to take up glucose, and signals the liver to cease releasing it. Thus this prediction can prevent a large rise in blood glucose.

A different prediction can do the opposite, that is, can elevate blood glucose above the most frequent level. For example, Cannon reported that members of the Harvard rowing crew, anticipating a race, would elevate their blood glucose to levels that spilled into the urine (Cannon, 1929). In other words, predicting an intense need for metabolic energy can raise blood glucose to diabetic levels. Insulin and the myriad other hormones that regulate the fuel supply are modulated rigorously from the brain which bases its predictions on a continuous data stream regarding metabolic state that arrives via nerves from the liver and sensors in the cerebrovascular organs, such as the area postrema, and the hypothalamus (Friedman et al., 1998; Saper et al., 2002). The importance and challenge of predictive regulation is best appreciated by the type 1 diabetic who tries to prevent surges of blood glucose by injecting insulin *before* a meal, and who must inject insulin *before* exercise to allow glucose to enter his muscles.

Sensors must match the expected range of input.

Sensors are designed to transduce a range of inputs into a range of outputs (Figure 5A). Typically the input/output curve is sigmoid and set so that its midpoint corresponds to the statistically most probable input (Figure 5B). The curve's steep, linear region brackets a range of inputs that are somewhat likely, and its flatter regions correspond to inputs (very weak or very strong) that are relatively unlikely (Laughlin, 1981). This design has a clear advantage: the most likely events are treated with greatest sensitivity and precision (Laughlin, 1981; Koshland et al., 1982). When input events are small relative to noise, they may be amplified nonlinearly to remove the noise by thresholding (Field and Rieke, 2002), but most sensors amplify linearly as shown here (Rieke et al., 1999). Note that the design of each sensor embodies "prior knowledge", derived via natural selection, regarding the range of most likely inputs (Sterling, 2003).

This simple design is effective when the statistical distribution of inputs is steady. But environmental signals fluctuate enormously, for example, light intensity changes between day and night by ten billion-fold. The linear range of a visual sensor spans only ten-fold. So over the course of a day it would frequently confront a range of inputs far too large or too small for its response curve (Figure 5B). For much stronger inputs the detector would be too sensitive, and its output would saturate; for much weaker inputs, it would be too insensitive and would miss them.

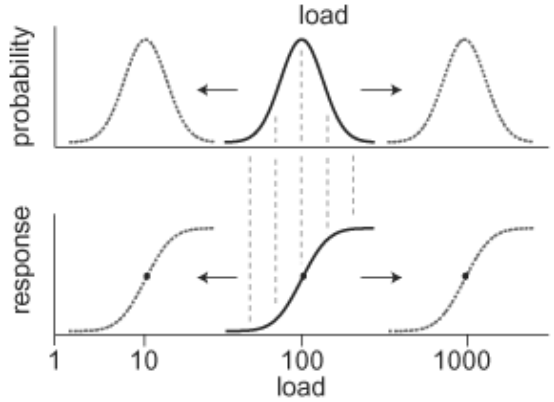

Figure 5. Regulatory mechanisms adapt to keep the input/output curves centered on the most probable loads.
Upper panel. Every system confronts some distribution of probable loads (bold curve). As conditions shift, so does the distribution (dashed). *Lower panel.* The input/output curve (bold) is typically sigmoid with its most sensitive region (steep part) matched to the most probable loads. When the distribution of probable loads shifts, the input/output curve shifts correspondingly (dashed). See Laughlin, 1981.

There is a remedy: sense the altered input statistics → calculate a new probability distribution → shift the response curve to rematch its steep region to the most likely loads (Figure 5B). This strategy for continually rematching outputs to expected inputs has been observed at all levels of biological organization from bacteria and somatic (nonneural) cells to neurons (Koshland, 1987; Sakmann and Creutzfeldt, 1969). At lower levels the process has been termed "adaptation", and recently "Bayesian" has been added to emphasize Bayes' insight that the best estimate of what is happening in the world combines data from our sensors with our prior knowledge about what is probably out there (Rieke et al., 1999). This principle operates at many levels. Thus, we rely on a single experience of the unpleasant sensations associated with regulatory conflict (dizziness, nausea) to permanently enlarge our store of "prior knowledge". And in the perceptual realm, we identify an ambiguous object by sight or touch by combining sensory inputs with our prior knowledge of what the context suggests is most likely (Geisler and Diehl, 2002, 2003). In case of conflict between vision and touch, we rely on prior knowledge of which sense is probably more accurate (Ernst and Banks, 2002).

Prediction requires each sensor to adapt its sensitivity to the expected range of inputs.

Prediction must be based on sensors that are both accurate and fast with respect to the processes that they help to regulate. How sensors maintain their accuracy and speed over a large dynamic range is now well understood for various neural systems, especially for vision. The basic principles seem likely to be generalizable to all sensors. Indeed the conclusions from analyzing vision in the fly and vertebrate retinas are similar to those reached by analyzing chemotaxis in bacteria (cf, Laughlin, 1994 and Rieke et al, 1999 vs. Koshland et al, 1982; Koshland, 1987). Therefore, this section summarizes the current understanding of how and why sensors adapt to their inputs.

Rate of adaptation matches the rate of changing input.

Input statistics can fluctuate very rapidly – for example, light intensity on the eye of a flying insect varies over milliseconds. For a visual neuron to match its responses to such shifts, it must adapt in milliseconds; otherwise it will always be optimized for a past condition and never for the input that it will most likely encounter next. Such a neuron measures the input very briefly – for just long enough to provide reliable statistics – and then shifts sensitivity accordingly. Because photons arrive stochastically, their intrinsic noise is set by the square root of the number counted; so a fly neuron that adapts over 10 milliseconds can sense a change of 100 photons captured over five milliseconds and then predict the most likely intensity of the next instant to within 10% – leaving five milliseconds to shift the response curve (Fairhall et al.,

2001). In short, natural selection ensures prediction down to the limit set by physical laws (Rieke et al., 1999; Laughlin, 1994; Sterling, 2003).

The time course of predictive adaptation differs for every system and depends partly on the length of time spent under a particular load. For example, after carrying groceries from your car to the kitchen, your sense of effort is reduced briefly – a coffee cup feels "lighter" than normal. But after wearing ski boots or a backpack for hours, the sense of weightlessness lasts longer. Over hours mechanoreceptors in muscle, tendon, and ligament have reduced their sensitivities to match the persistently increased load. But then over tens of minutes, sadly perhaps, we regain our usual sense of effort as these mechanisms re-adapt to predict the next round of most likely loads. Even astronauts, initially exhilarated by zero gravity, gradually cease to notice either their weightlessness or even their unusual orientation within the cabin (for us, "upside down") – because all their sensory mechanisms readjust to predict the most likely conditions.

There are two levels of prediction: (1) most likely state in the next moment – generally best captured by the current state and its rate of change; (2) most likely time course of the new state – generally best captured by length of time in the current state. This second factor, persistence, improves efficiency because each change requires a response, and each response has a cost. Many predictors reduce costs by anticipating regular shifts in demand. For example, circadian prediction proves so advantageous that every cell in the body uses it to regulate the expression of vast numbers of different genes according to predicted demand (Roenneberg and Merrow, 2003). On a longer time scale, seasonal variation in daylength predicts average environmental temperature and food availability, performing much more reliably than local temperature. Furthermore, for migratory species daylength predicts the most likely temperature thousands of miles away. Consequently, predictions based on daylength have been built into the brains of many species as "prior knowledge" that profoundly regulates their physiology (Mrosovsky, 1990).

Prediction requires each effector to adapt its output to the expected range of reward.

Effectors also shift their output curves to match a change in the expected range of demand (Figure 5). Of course, effectors change more slowly than sensors because their adaptations are more expensive.[5] The example most familiar, because we observe it directly, is skeletal muscle. One bout of intense effort, though fatiguing, little affects the response curve. But prolonged effort over days, weeks, months gradually evokes a panoply of gene modulations: increased synthesis of proteins for muscle, bone, and connective tissue, plus corresponding shifts of metabolic and respiratory enzymes. It would be a costly design that mobilized all these mechanisms just to deliver your groceries, or that completely *de*mobilized them after just a day's layoff from training. Even so, world-class athletes known for their superior fitness, such as Roger Clemens and Lance Armstrong, never reduce their physical demands more than momentarily, lest their effectors readapt even slightly to lower demand. And at zero gravity only a few weeks are sufficient to reduce a perfectly fit astronaut to jelly.

Internal effectors also adapt gradually. For example, although the brain's sensor of circadian time (suprachiasmatic nucleus) resets to a shift in daylength within one cycle, the liver, which synthesizes many gene products under circadian control, resets over six days (Roenneberg and Merrow, 2003). In fact all cells, via diverse molecular sensors on their surfaces(receptor

[5] If demand rises beyond current capacity for *long* periods, new power plants get built but, just as with the body, effector capacity lags)

proteins), regulate to meet predicted demand.[6] Furthermore, these receptors themselves regulate in number and sensitivity to match predicted demand over a range of time scales. Typically, prolonged exposure to high levels of its natural ligand (signaling molecule) reduces receptor number and sensitivity.

Note that downregulation of a receptor triggered by prolonged exposure to its ligand occurs by negative feedback. But this need not be caused by an "error"; rather the downregulation is simply a response to the anticipation of a higher level of the ligand. Thus, when blood glucose is persistently elevated and triggers persistent secretion of insulin, insulin receptors eventually anticipate high insulin and downregulate. The system learns that blood glucose is *supposed* to be high. Similarly, sustained demand for elevated blood pressure teaches all effectors to expect it, and gradually adapt: arterial smooth muscle cells hypertrophy; the carotid sinus wall thickens to reduce baroreceptor sensitivity; secretory cells whose products support the pressure rise hypertrophy (renin, norepinephrine, cortisol, etc.). In short, it seems inevitable that the sustained presence of high blood glucose would gradually reduce insulin sensitivity; i.e. cause "insulin resistance", and thus type 2 diabetes, and that sustained elevations of blood pressure would gradually cause essential hypertension. Such changes are the appropriate adaptations to predicted demand (Figure 5).

Predictive regulation relies on complex behavior whose neural mechanisms also adapt.

Few of the raw materials needed for regulation are stored in any quantity. Most of the body's sodium is in the blood and extracellular space, and sodium is lost daily, along with water, to tears, sweat, and urine.[7] Calcium is stored within various intracellular compartments, but there it is needed for signaling and must not be depleted. Only bone can loan calcium for the short term, but for obvious reasons, the loan must be repaid. Fuel is generally stored in modest quantity as glycogen and fat, whose rapidly mobilizable components within muscle cells are just sufficient to carry a trained runner to the end of a marathon (Weibel, 2000). Prolonged exertion, as in the *Tour de France*, soon depletes stored fat and is ultimately limited by the gut's maximum absorptive capacity – which can sustain energy consumption over basal levels by only about four-fold (Hammond and Diamond, 1997). In short, physiological regulation is inexorably tied to replenishing.

The most efficient way to update a grocery list is *immediately* – as an item is used.[8] There are two reasons. First, to be depleted is unpleasant – and can be quickly lethal. Second, supplies, such as salt, water, and fuel, are not always available. The brain's every command to consume a particular substance always accompanied by parallel commands to reduce its loss and to seek opportunities to replenish. This need to replenish generally involves a rich set of cognitive and emotional experiences and to make this concrete consider this example from an actual hike in the Arizona desert, a 7,000 foot descent into the Grand Canyon.

This hot, arid environment demands evaporative cooling through sweat which expends water and sodium. In anticipation the brain triggers release of antidiuretic hormone and

[6] This represents an important difference from machines. The BMW's central integrator can adapt, but its effector parts do not adjust; they only wear out.

[7] The moose provides an instructive exception. Its winter diet of tree browse contains hardly any sodium, but the moose can borrow sodium from its capacious rumen (stomach for fermentation) and then repay the debt, after the ice thaws, by feeding on water plants rich in sodium (Denton, 1993).

[8] Supermarkets have now adopted this practice – each item sold at the register by its bar code is automatically subtracted from the inventory and re-ordered.

aldosterone to stringently suppress salt and water loss through urine. Nevertheless, we soon felt thirsty and paused to replenish.

Drinking from the water bottle, which we had anticipated would be essential, our thirst was satisfied. But watching the level drop, as our companions also drank, we felt anxious that the supply might be insufficient and that others might drink our share. Responding to this concern, one group member suggested a rule to govern further consumption; another consulted the map for the next spring, and the third redistributed the weight among the packs, taking more for himself, to better match the strengths of the individual hikers. Upon reaching the spring, the anxiety dissipated, there was pleasure in the drinking, and rejoicing in the sense of solidarity that accompanied our successful cooperation. But before very long, gazing up the sheer walls to the Canyon's distant rim, we began to wonder and worry how to carry enough water to climb out.

Such an experience illustrates that human physiological regulation depends powerfully on a host of high-level neural mechanisms: retrieval of prior knowledge, multiple emotions, perception, planning, cooperation, and altruism. Such an experience can thrust us back, past Gortex and Polartec, to the root of human evolutionary success and cause us reflect on its neural basis. Somewhere in the brain all the critical factors must weighed, and a plan forcefully executed. The critical site turns out to be the prefrontal cortex.

How allostasis depends on higher levels.

Prefrontal cortex: where thought and feeling, past and present, meet.

Each sensory system projects to its own primary area in neocortex where elaborate computations begin to identify key environmental features and group them based on prior knowledge (Figure 6A; Geisler and Diehl, 2002; 2003). These areas relay to various higher order areas (30 or more areas for vision) that compute still higher order features. Eventually, data from the separate senses converge at particular cortical sites, such as the intraparietal lobule, to be compared and weighed against each other (Ernst and Banks, 2002). Ultimately the higher order and multimodal regions converge in cascades upon the prefrontal cortex (Figure 6A), and from this cascaded pattern emerges the best estimate, based on all the senses, and either conscious and unconscious, about the present.

Although it remains uncertain exactly where and how the brain stores specific memories, we do know that retrieval involves elaborate connections within "limbic" structures and that these also project in cascades to prefrontal cortex (Figure 6B). Thus, returning to our example, the desert hike, the prefrontal cortex connects the visually perceived level in the water bottle with the past experience of running out. This cortical region, which encompasses about a dozen specific areas, can even connect the present situation to a grandfather's tale about running out of water, or to the story of Hagar and Ishmael from Genesis.

But to simply compare past *vs.* present and calculate a plan is insufficient. To *execute* a plan, thoughts need to be driven – focused – by emotion. This is another task of the limbic system, to generate some combination of feelings, such as anxiety, rage, and love, that sustainedly connect our thoughts which otherwise tend to flicker about. We know that prefrontal cortex participates in the focusing of thought by emotion because for about 30 years (1940-1970) prefrontal lobotomy was widely performed as a treatment for mental illness. Following lobotomy a person can conceive of a plan but cannot hold it in focus for long enough to complete it, a deficit that has been succinctly termed "instability of intent" (Nauta, 1971; Sterling, 1978; Valenstein, 1973, 1986).

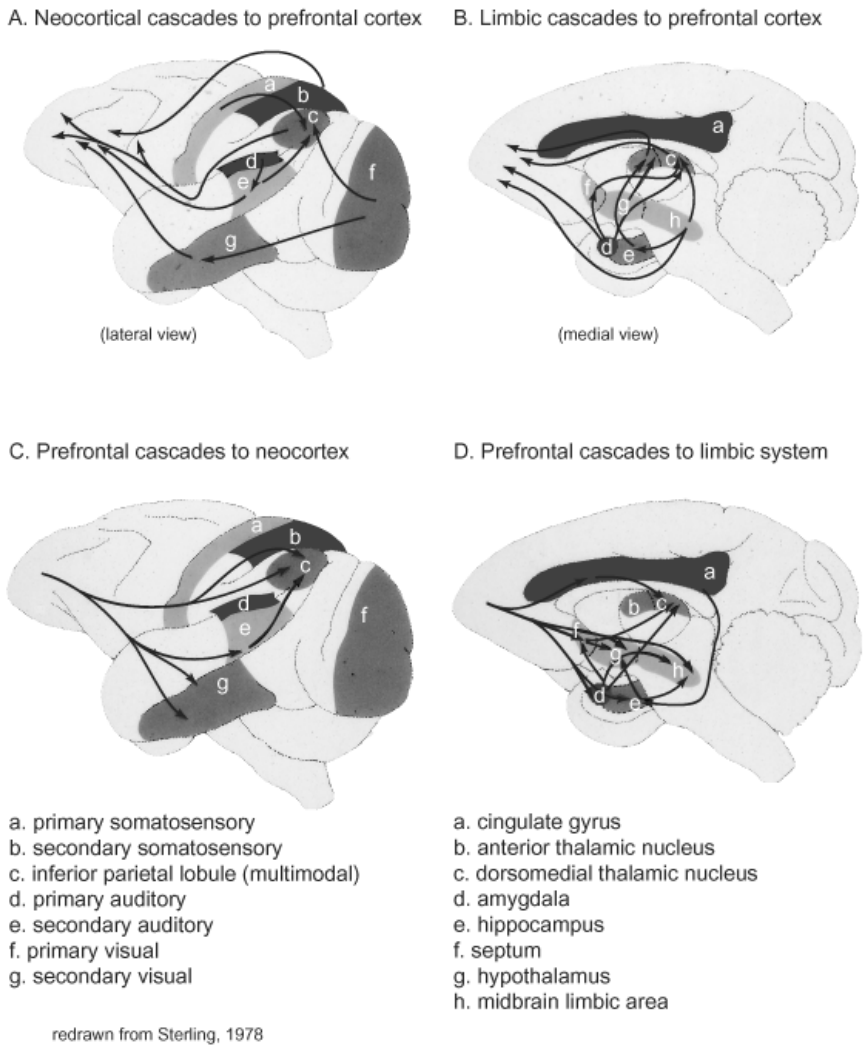

Figure 6. Prefrontal cortex integrates cascaded inputs from neocortical and limbic systems – and feeds back to both. This arrangement serves two functions: to imbue intellectual calculations with urgency and focus, and to modulate emotional expression by perceptual and cognitive context (Nauta, 1971). Diagram shows the brain of macaque monkey.

Just as thought must be focused by emotion to support physiological regulation, so must the experience and expression of emotion be governed by perception and thought. Thus, the perception of a diminishing water supply triggers anxious feelings, and our persistent expression of these feelings to our companions may curtail their rate of consumption. If that fails, anxiety may turn to fear – whose expression as anger will attract the attention of even the least sensitive companion. These mechanisms are served by cascaded outputs from prefrontal cortex back to the neocortical and limbic structures (Figure 6C, D). Following lobotomy, a person can still express emotion, but because it is no longer modulated by shifting perceptions, affect is "flattened", and when affect does shift, it is frequently inappropriate to the circumstance.

These deficits illuminate how profoundly human physiology depends on continual modulation of emotional expression. Inappropriate affect will fail to achieve the immediate goal – conservation and sharing of a limited resource. Affect, if too mild, will fail to persuade our companions; affect, if too strong, will anger them and threaten the group's stability. Flattened affect also degrades even the simplest exchange, whether of substance (water) or an idea (schedule for consumption), because every exchange requires an emotional acknowledgement. A plain, "thank you.", unaccompanied by a wink, a nod, a smile, or a touch, may be perceived as

insincere and thus fail to provide what is essential to every exchange between humans, some reciprocal emotional recognition. Such recognition both relieves anxiety and provides pleasure. This is so fundamental that an individual with flattened affect is rapidly identified as deviant and isolated from the social interactions essential for survival.

"Stick and carrot" for anticipatory regulation: neural mechanisms for anxiety and satisfaction.

The feelings of anxiety, fear, and anger depend on neural activity in the amygdala, a large complex of nuclei in the basal forebrain which connects reciprocally with prefrontal cortex (Figure 6B, D). The amygdala also serves our highly developed ability to predict these feelings in others based on their facial expression and body language. We know this partly because the amygdala was ablated during the 1960s-1970s as a therapy to attenuate anxiety and rage (cf (Mark and Ervin, 1970; Valenstein, 1973; Sterling, 1978), and partly because of recent studies by magnetic resonance imaging (Adolphs et al., 1994; LaBar et al., 1998). The amygdala collects and integrates myriad lower level signals concerned with physiological regulation: **(i)** steroid hormones and peptides that regulate blood pressure, mineral and energy balance; **(ii)** neural signals from the brainstem visceral areas, such as nucleus of the solitary tract and the hypothalamus; **(iii)** signals from brainstem raphe neurons that modulate levels of arousal and mood via the neural transmitter, serotonin (Schulkin et al., 1994; Schulkin, 1999).

Serotonin serves many different functions in the brain, because the raphe neurons project very widely, for example, down to sensory and motor columns of spinal cord and up to the cerebellum and neocortex. But serotonin released in the amygdala appears to suppress transmission of anxiety signals to prefrontal cortex. Consistent with this theory, drugs that increase release of serotonin or its persistence in the synaptic cleft, reduce anxiety, elevate mood, and enhance social pleasure[9] (Wise, 2003). Presumably, the powerful social signals transmitted via neocortical cascades to the prefrontal cortex (Figure 6A) also feed down to the amygdala and other limbic structures to regulate serotonin release in cortex (Figure 6A, D).

Many behaviors that serve physiological regulation are driven less by the promise of reducing anxiety than by the expectation of "reward" – some outcome that leads to a feeling of satisfaction. Satisfaction depends on activity of neurons in the midbrain's ventral tegmental area (VTA). This limbic region connects reciprocally with the basal forebrain (nucleus accumbens) and the prefrontal cortex (Figure 6B, D). The VTA, like the amygdala, integrates signals related to myriad appetites – for specific nutrients, water, fuels, sex, etc – and uses prior experience to establish specific predictions of how each behavior should be rewarded. A VTA neuron is quiet, until an outcome exceeds the expectation, but then it increases firing linearly with magnitude of the perceived reward (Schultz, 2002; Fiorillo et al., 2003). Each action potential releases a pulse of the neural transmitter, dopamine, whose binding by receptors in the nucleus accumbens and prefrontal cortex signals "satisfaction".

This mechanism provides a common pathway to sustain behaviors that serve *all* appetites (Montague and Berns, 2002). It works because what the accumbens/prefrontal cortex apparently "wants" is neither sodium, sugar, fat, nor sex, *per se*, but simply a pulse of dopamine. Because this mechanism is general and plastic, it can lock onto virtually any experience or behavior that might release dopamine: music, visual art, food, or social recognition. For one mechanism to

[9] MDMA (methylenedioxy-n-methylamphetamine, "Ecstasy") is used recreationally to enhance social interactions, especially in all night dancing, termed the "rave". Of course, similar ecstasy can be achieved during intense social interactions, as documented, for example, in the ethnographic film classic, "*Trance and Dance in Bali*" (Bateson and Mead, 1952).

serve so many core needs, the satisfaction that it brings must be brief; that is, the mechanism must adapt quickly. Thus the profound restlessness of the human spirit (Goethe, 1833) seems to arise mechanistically from the ceaseless call of the accumbens and prefrontal cortex for pulses of dopamine – to which they rapidly desensitize.

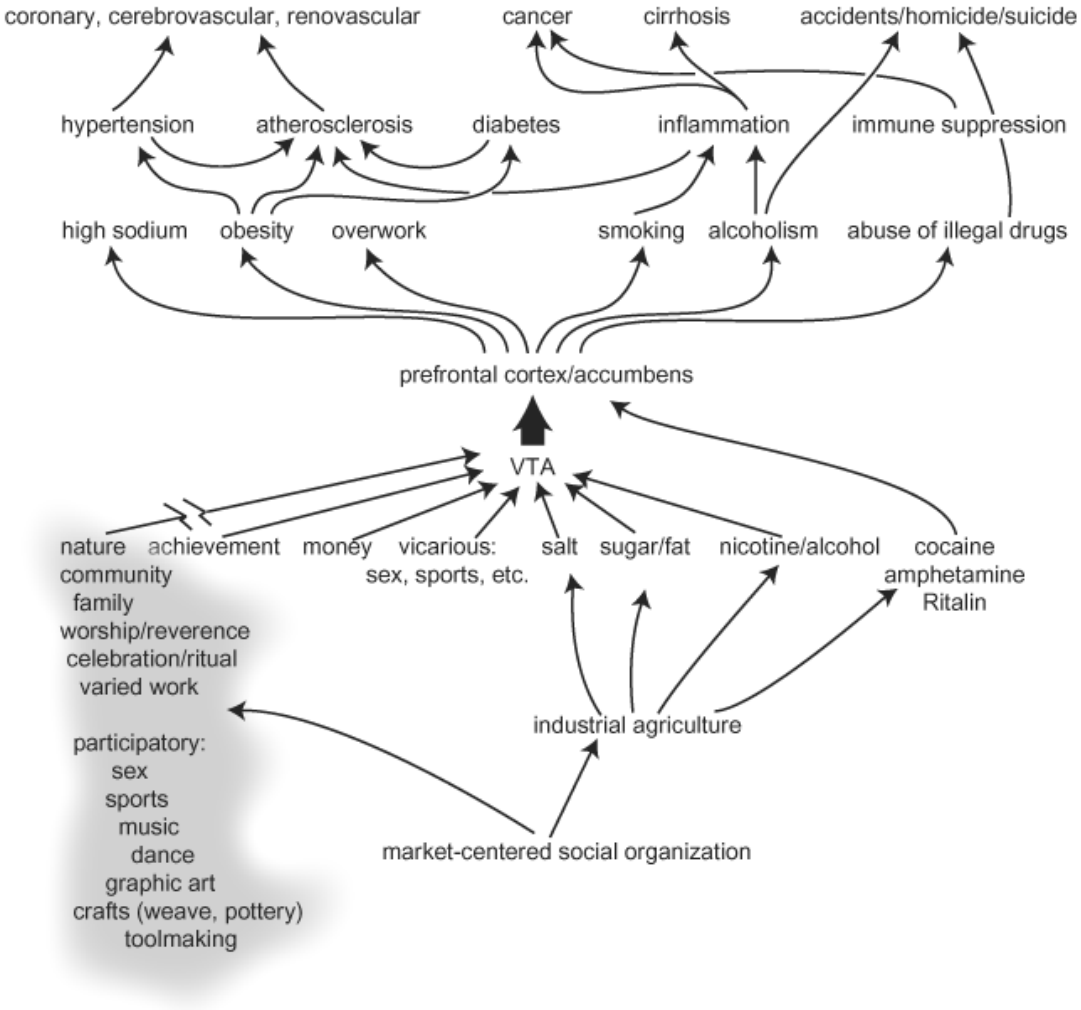

Figure 7. How market-centered social organization contributes to mortality from hypervigilance and hyposatisfaction.
Major causes of mortality stem from several co-occurring and mutually reinforcing pathogenetic processes. These arise partly from regulatory mechanisms (Figure 4) designed to meet the need for hypervigilance and also from behaviors that try to meet the need for daily satisfactions. Each of many potential sources of satisfaction (shaded on left) can cause neurons in the ventral tegmental area (VTA) to deliver a pulse of dopamine to the nucleus acumbens and prefrontal cortex, thereby briefly providing a sense of well-being (Montague and Bern, 2002; Schultz, 2002). But market-centered social organization narrows the sources of satisfaction; whereas satisfaction from a single source (work, food, nicotine) tends to adapt (Figure 5), requiring higher levels for the same relief. A chronically elevated appetite in the context of "industrial agriculture", which provides the key substances cheaply and markets them intensively (Nestle, 2002; Schlosser, 2002), lead to a panoply of pathogenetic mechanisms which have been grouped as "metabolic syndrome" (Zimmet et al., 2001; Moller, 2001). When the reward system remains "unsatisfied" by natural inputs, drugs are employed to "short circuit" the reward system by directly increasing cerebral dopamine (Wise, 2003; Schultz, 2002)

Consistent with this, many drugs that enhance the sense of well-being, including nicotine, ethanol, opioids, and cannabinoids, work by directly activating VTA neurons via specific molecular receptors to enhance dopamine release in the accumbens and prefrontal cortex. Other

drugs, such as amphetamine and cocaine, evoke these feelings by acting directly on the dopamine terminals in accumbens and cortex to enhance release or block reuptake, which allows dopamine to persist in the synaptic cleft (Wise, 2003; Schultz, 2002). When a signal from any one source is prolonged, this system desensitizes, like all other sensors and effectors, to keep its input/output curve centered on the most frequent demand (Figure 5). Thus, just as our enhanced sense of effortlessness fades quickly after setting down the groceries, so does our enhanced sense of well-being fade – and for the same computational reasons – all systems must adapt to a persistent signal.

This mechanism, as the "carrot" for anticipatory regulation and thus key to the behavioral regulation of physiology, harbors grave potential for pathology. The system is designed to serve myriad needs, each one contributing a small dollop of satisfaction. But satisfaction cannot be stored, so if the number of sources shrink, the task of driving the mechanism devolves to the few that remain. And the more frequently one source is called upon, because of desensitization, the less satisfaction it can deliver. Yet, the persistent demand of this one circuit calls for still stronger stimulation , insistently crying out, "feed me!". Under these circumstances the "reward circuit" can mediate addiction to essentially any source of satisfaction. But note that for this to occur, nothing inside the body need be "broken" or "dysregulated". The system can arrive at this state, locked on to a single source of satisfaction, simply because life circumstances have reduced all the other sources. Now we can summarize how the "stick" of hypervigilance and the "carrot" of satisfaction contribute to disorders of physiological regulation.

Pathophysiology from allostasis.

Hypertension: adaptation to sustained vigilance.

Roughly one-quarter of US adults are hypertensive (blood pressure >140/80 mm Hg on repeated measurement). A few cases arise from identifiably defective phenotypes (e.g., (Wilson et al., 2001), but 95% are classified as "essential" hypertension –cause unknown (Zhu et al., 2002). Prevalence is greater for African-Americans than for whites (32% vs. 23%; Carretero and Oparil, 2000a) . This difference of nearly 40% is commonly attributed to genetics, but this seems doubtful because the West African ancestors of US blacks were not hypertensive (Waldron, 1979). Furthermore, hypertension seems to be more strongly associated with various sources of social distress, rather than race *per se*. Thus its main sequelae – death from coronary heart disease, cerebrovascular disease, and atherosclerosis – are more prevalent for divorced *vs*. married men and for low *vs.* high employment grades, by factors as large or larger than for race (Figure 11). Thus, if in Cleveland I had canvassed *poor white* neighborhoods, I would also have seen men with limps and sagging faces.

The homeostasis model cannot explain essential hypertension because it attributes all pathology to a "defect" – to something "broken". But the allostasis model suggests that there is no defect. More parsimoniously, it proposes that hypertension emerges as the concerted response of multiple neural effectors to prediction of a need for vigilance (Figure 4). When this prediction is sustained, all the effectors, both somatic and neural, adapt progressively to life at high pressure. The adaptations all seem entirely explicable from our general knowledge of signaling and regulation (Figure 5). Although the endpoint may be tragic (Figures 7, 11), every step along the path seems perfectly "appropriate".

Vigilance starts when a child is delivered from its mother's protection to the care of strangers. Thirty years ago this occurred when US children first entered school at around age 6. Correspondingly, blood pressures were constant from the first year of life until age six (median

systolic level ~100 mm Hg). Then commenced a steady rise, so that by age 17 half of all boys showed systolic pressures above 130 mm Hg, and about 20 percent showed pressures above 140 mm Hg (hypertensive). The rise for girls was similar, though slightly milder (Figure 8).

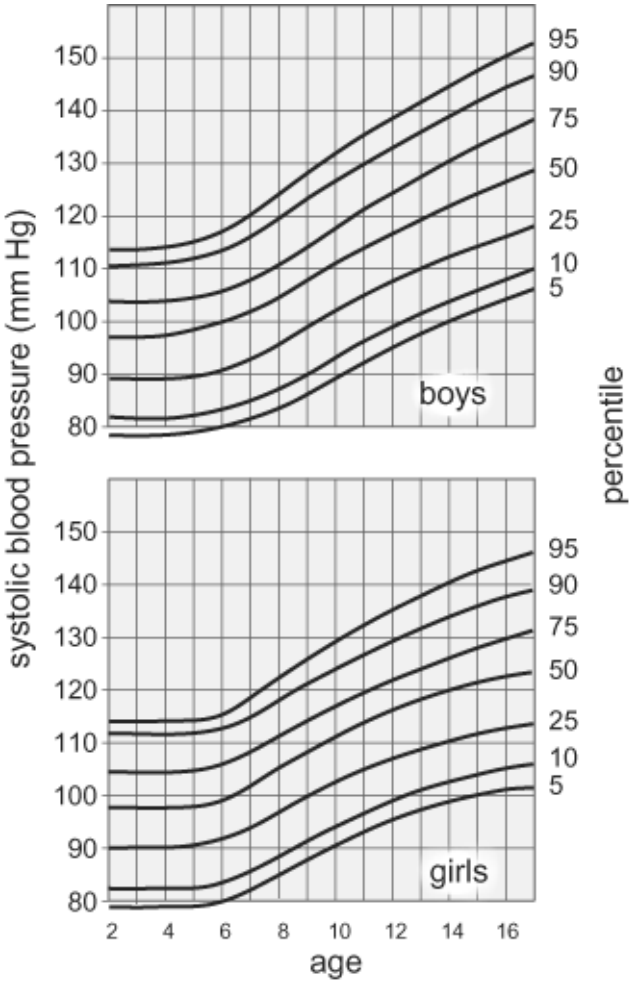

Figure 8. Systolic blood pressures were steady until school age and then rise continuously. Diastolic pressures also rise. Recent data show pressures rising in the first year, perhaps associated with the increase of "day care" (NIH report, 1997). Redrawn from Blumenthal et. al., 1977.

But now blood pressures begin to rise in the first year of life (NIH Report, 1997). This startling change might be associated with the rise of "day care" and the shift of mothers away from their infants and into the workforce. Consistent with this hypothesis, rat pups detached from their mothers show an eight-fold rise of corticosterone over 24 hours, and human toddlers detached from their parents show increased cortisol (Schulkin, 1999). As noted in Figure 4, the neural signals that call for increased blood pressure also call for salty foods – which the fast-food industry ("industrial agriculture") provides in prodigious quantity, both in the supermarket and as part of the Federally-funded school lunch program (Figure 7; (Nestle, 2002; Schlosser, 2002). Industrial agriculture does not *cause* hypertension by excessively salting prepared foods; it merely obliges the public's appetite for sodium, which is driven quite appropriately by intact regulatory systems. Indeed, if under present conditions of life, the food industry were to restrict sodium, we might see the development of public "salt licks" like those used to attract deer.

In a younger person if the predicted need for vigilance declines, effector adaptations can reverse promptly. But persistent demand leads to more profound and persistent effector adaptations. Over decades the constant call for vigilance adapts arterial muscle and carotid sinus to thicken and stiffen so that pressure rarely returns to normal levels. Probably there are also corresponding adaptations in the brain. We know now that adult synapses continuously adjust their molecular components and that "memories" are stored at all levels, even in the spinal cord (Lücher and Frerking, 2003; Ikeda et al., 2003). So the many hormones that feed back to the brain to sustain high pressure (Figure 4) probably entrain many levels to expect and support high pressure.

Thus, coordinated somatic and brain adaptations generate response patterns of "established" hypertension (Figure 9). The hypertensive pattern, like the normal pattern (Figure 3), does not seem to be "defended" at a particular level. Rather it is modulated up and down, apparently according to demand, with an overall range of 140 points. This pattern suggests

adaptation to chronic vigilance, and consistent with this the hypertensive pattern is absent in undisrupted preindustrial societies where children remain in contact with their parents and strangers are rare (Eyer and Sterling, 1977).

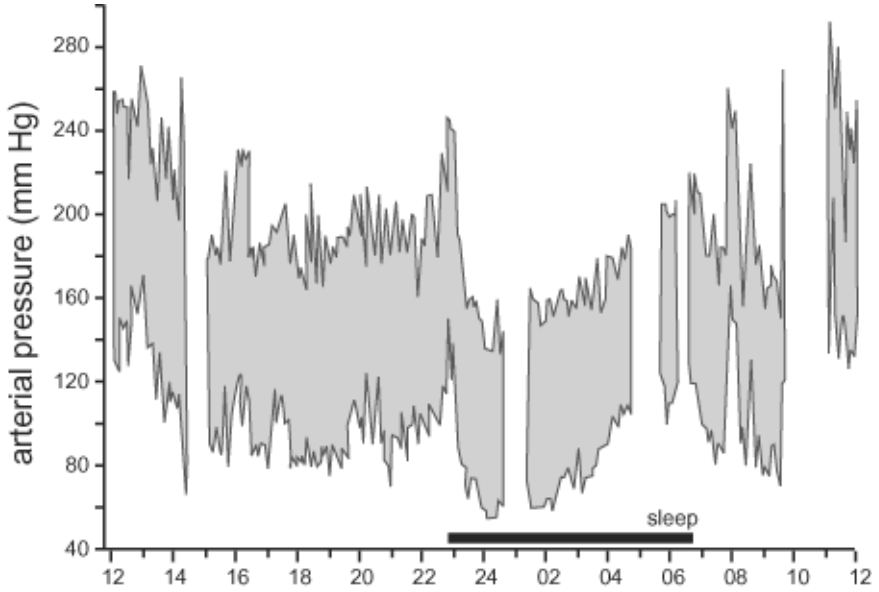

Figure 9. Arterial pressure fluctuates even in a subject with established hypertension. This pattern suggests, as for the normal subject in Figure 3, not defense of a setpoint, but rather responsiveness to fluctuating demand. Same conventions as Figure 3. Redrawn from Bevan et al., 1969.

Established hypertension is most common in segments of modern society where family structure is most disrupted, where children are least protected, and where they are marked from birth for suspicion and various forms of ill-treatment. One example stamped into memory is of the teenager I tutored in Hough who refused an opportunity to attend summer camp. As she explained, "I'd be in the sun a lot, and I'm already dark enough". Not to mention the universal experience of African-American men who are regularly stopped by the police for the offense known as "DWB" – driving while black.

To such "prior knowledge" – the well-founded expectation of suspicion or contempt – hypervigilance is a natural response. Although the most overtly repressive forms of racism, such as lynching and legally enforced segregation of public facilities, have declined since the 1960s, *de facto* segregation of neighborhoods and schools has actually expanded, as has the disparity in wealth between rich and poor (Massey and Denton, 1994; Orfield, 2001). These trends intensify the sense of wariness between inhabitants of the ghetto and the outside – which enhances everyone's vigilance. Finally, since "prior knowledge" embodied in communal attitudes serves regulation, the communal suspicion and fear engendered by experience can not only be stored for decades within individual brains but also transferred across generations (Figure 7). So we should expect that even when overt racial conflict declines, "prior knowledge" among African-Americans about what to expect from white people that was built into communal expectations over 400 years, will dissipate with a time constant of generations.

So to explain essential hypertension there is no need to postulate a "defect" in any particular regulatory pathway. Certainly we can create a hypertensive mouse by knocking out one gene or another (Wilson et al., 2001; Zhu et al., 2002). But we can also create hypertension and atherosclerosis in a whole colony of mice simply by introducing a stranger (Henry et al., 1967). Certainly we recognize that the variance of blood pressure within a community must be partially caused by genetic differences. But this cannot explain why blood pressures of essentially *all* our children rise with age. Nor why the rise is largest and most persistent in the

poorest and most socially disrupted communities. Nor why African-Americans are more hypertensive than genetically similar populations in West Africa. These observations certainly point to an environmental cause. Furthermore, hypertension represents only one of many similar threads from the quilt of predictive regulation.

Obesity and metabolic syndrome: adaptation to hypo-satisfaction.

Roughly half of US adults are obese, a condition that contributes to type 2 diabetes. Obesity and type 2 diabetes jointly contribute to a constellation of pathologies, recently termed "metabolic syndrome", which includes hypertension, glucose intolerance (diabetes), hyperinsulinemia, dyslipidemia, visceral obesity, atherosclerosis, and hypercoagulability (Zimmet et al., 2001). Together these factors create a profoundly lethal cascade (Figure 7), and all follow the familiar epidemiological pattern: elevated with divorce, low socio-economic status, and disrupted preindustrial communities (Figures 10, 11; Zimmet et al., 2001; Diamond, 2003). Like blood pressure, these conditions are rising in children, where the rate of obesity has reached 15% (Hill et al., 2003).

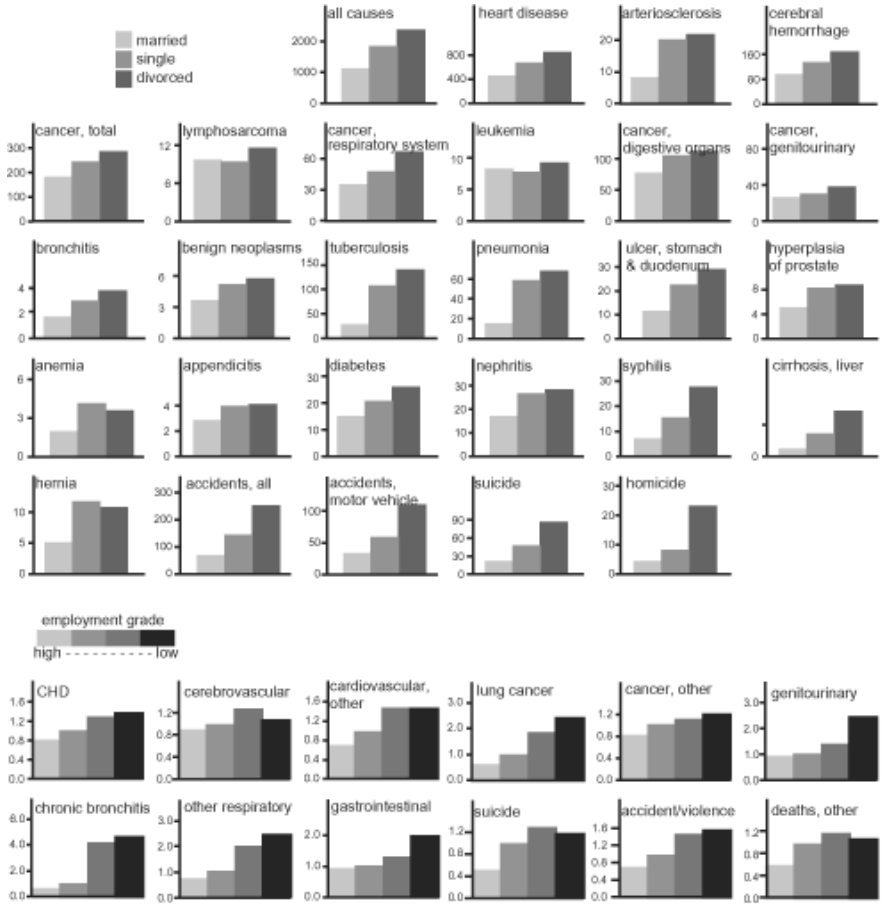

Figure 10. Death rates from diverse causes correlate with marital status and employment status. Such epidemiological patterns are consistent with the hypothesis (Figure 7) that hypervigilance and hyposatisfaction contribute broadly to pathogenesis. *Upper:* redrawn from Berkson, 1962. *Lower:* adapted from Marmot, 2000.

The homeostasis model cannot explain the prevalence of obesity. If metabolism were primarily controlled by negative feedback, then decreased energy expenditure would lead to decreased food intake. Yet presently in the US the opposite is so: the less we exercise, the more we eat. This conundrum could be caused by defects in the regulatory chain. For example, certain obese individuals are deficient in leptin, an important negative regulator of feeding, and when administered leptin their weight returns toward normal (Farooqi et al., 2002). But, just like

hypertension, specific defects in energy regulation are rare. They account for only a minor fraction of obesity and for none of its striking increase (Hill et al., 2003).

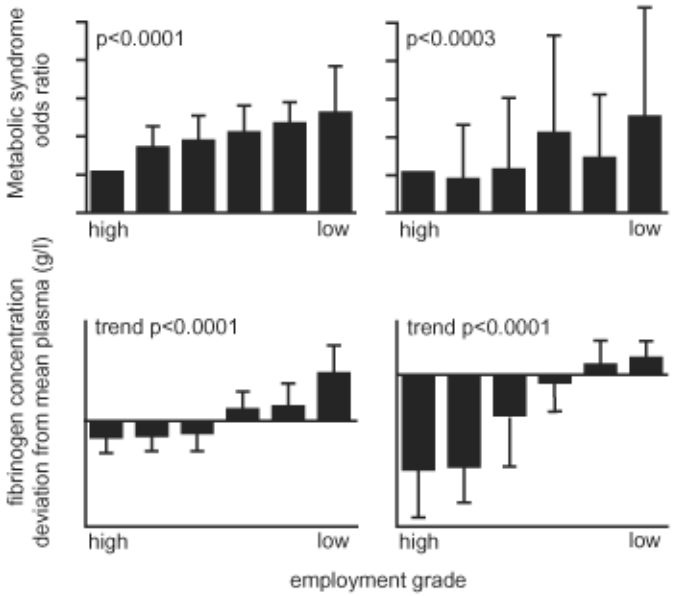

Figure 11. Metabolic syndrome and fibrinogen concentration are higher for lower employment grades.
Metabolic syndrome includes many factors that contribute to cardiovascular and cerebrovascular disease, and elevated fibrinogen directly increases probability of myocardial or cerebral infarction. Thus these patterns are also consistent with the hypothesis diagramed in Figure 7. Redrawn from Brunner, 2000.

Nor can homeostasis explain the growing prevalence of type 2 diabetes. Its core feature, insulin "resistance", involves changes at many levels, including decreased concentrations of insulin receptors, kinase activities, concentration and phosphorylation of IRS-1 and –2, PI(3)K activity, glucose transporter translocation, and the activities of intracellular enzymes (Saltiel and Kahn, 2001). Although these changes are loosely termed "defects" (Saltiel and Kahn, 2001), they do not arise from mutant alleles, so "defect" denotes, not their origins, but rather their unwanted effects.

The allostasis model can explain both obesity and insulin resistance without postulating any true defect. The standard signals for vigilance (such as cortisol), which raise the appetite for sodium, also raise the appetite for carbohydrate and fat (Schulkin et al., 1994). This makes adaptive sense – if we will soon need more salt, we will also soon need more fuel. Elevated cortisol also shifts the distribution of fat deposits toward the viscera – one feature of metabolic syndrome. And when chronically high levels of carbohydrate evoke chronically high levels of insulin, its receptors and their downstream mechanisms naturally reduce their sensitivities, just as every signaling system responds to prolonged, intense stimulation – the input/output curve inexorably shifts to the right (Figure 5). This does not explain mechanistically how myriad inter- and intracellular signals cause resistance, a problem now pursued by thousands of molecular and cell biologists. But the point to recognize is that bad things can happen as the natural outcome of predictive regulation.

Cortisol and related signals are elevated, not only during hypervigilance, but also during states of hyposatisfaction – when outcomes prove less than expectations. Because satisfaction cannot be stored, it must be continuously renewed. So if its potential sources become constricted, the brain must inevitably rely on those that remain: people needing a pulse of satisfaction will try to find it somehow (Figure 7). For those of higher socioeconomic status there are opportunities for satisfaction in work, achievement, and money. Mono-pursuit of such opportunities tends to spiral out of control ("workaholism", "type A" behavior, etc). This may occur especially when expectations inculcated by the family as "prior knowledge" are so high as to be intrinsically

unsatisfiable[10]. Another likely factor is that a stimulus which initially releases dopamine adapts, limiting the satisfaction obtainable from its repetition.

For people of lower socioeconomic status potential sources of satisfaction are less available, but food is abundant and cheap. So the allostasis model suggests that the brain overrides local negative feedback (metabolic satiety signals) – just as it overrides the negative feedback that would counter commands to raise blood pressure – and people eat. For the reasons just cited, satisfaction is fleeting – so people eat even more (Saper et al., 2002; Schultz, 2002).

Alcohol and drug addictions follow a similar pattern and apparently share many of the same mechanisms (Wise, 2003). For example, the neuropeptide NPY enhances feeding, and is also abundant in brain areas mediating these drug addictions. The acute effect of NPY resembles alcohol in reducing anxious behavior, and it is also associated with developing alcohol and cocaine dependence. Similarly leptin, identified primarily with feeding and energy balance, contributes to hypertension. Thus, there is considerable cross-talk between these systems along brain pathways that serve satisfaction (VTA- amygdala-accumbens-prefrontal cortex).

The rise in obesity and type 2 diabetes has recently been attributed to "thrifty genes". Noting that the most explosive increases are in populations that have suddenly changed from food scarcity to plenty, especially among Pacific Islanders, Diamond recalls Neel's hypothesis that certain human groups were selected to "eat up" in times of plenty to protect against times of famine (Diamond, 2003). This implies that body fat is not regulated to a setpoint, but varies according to some prediction – in this case, future hunger. This theory would be entirely consistent with the allostasis model, but there may be an additional explanation.

Consider that for these groups the sudden appearance of plentiful food is accompanied by the equally abrupt dissolution of the entire culture. Consequently, obesity is only one disorder of many that accompany disruption of a preindustrial society. Among Native Americans, Australian Aborigines, Inuit, and so on, the rise in obesity and type 2 diabetes invariably accompanies rises in essential hypertension, alcoholism, drug addiction, suicide, and murder (Eyer and Sterling, 1977). Furthermore, the same correlations are found in modern societies: the highest rates of all these afflictions appear in the most disrupted populations, those with the worst life experience, the lowest expectations, and the least hope. Thus over the period of rising racial segregation in urban neighborhoods and schools (Denton, 1993; Orfield, 2001), the prevalence of obesity in predominantly black elementary and middle schools has tripled (Gordon-Larsen et al., 1997).

In summary, the allostasis model attributes the pathogenesis of hypertension and metabolic syndrome to prolonged adaptation to hypervigilance and hyposatisfaction. The impact is strongest among populations with the best reasons for vigilance, the narrowest range of satisfactions, and expectations that are least often met (Figures 10, 11). Intensified genetic screening will undoubtedly identify defective or alternative alleles that render individuals more sensitive to one of these conditions than another. Such screens will help to explain ***who*** in a particular population gets ***what*** disorder. But since gene frequencies change over centuries, whereas the prevalence of these syndromes has risen over decades, their explosive increases cannot be attributed to genetic defects. So in considering where to intervene in order to prevent and treat this constellation of disorders, we should ask first, what changes in pre-industrial societies accompany the appearance of plentiful food?

[10] My mother in her eighties likes to tell my friends that she still feels guilty about smoking during her first pregnancy because, otherwise, "Peter might have been smarter".

Large-scale social organization: reasons for hypervigilance and loss of satisfactions.

Preindustrial communities, which upon disruption are so susceptible to modern disorders, have the following shared features. First, the communities are small, so most human encounters are between people who are completely familiar with each other and who are often closely related. Consequently each person can pretty well predict how every other person will behave, and all are constrained by clear rules to behave well. Second, every transaction between individuals is governed by the near certainty that there will be more transactions in the future. So, fairness and generosity tend to be rewarded, and cheating tends to be punished (Glimcher, 2002). Third, in a small community every significant transaction is widely known, discussed, and judged – providing another incentive to play fair. Finally, goods tend to be sparse and their ownership is well-known, so theft is pointless and consequently rare. All these circumstances reduce the need to sustain vigilance.

For industrial, market-dominated societies essentially the opposite is true. Communities are large and comprised mostly of strangers who come from different cultures with different rules and often with ample historical reasons for mistrust. Furthermore, many transactions occur between strangers who will never meet again, so trust is less rewarding, and the need to sustain hypervigilance is greater. Any urban skeptic, who might consider this a baseless speculation, should recall how remarkable it seems when on a trip to the country we find that rural people don't bother to lock the door or set a car alarm – and that enough trust persists that you can pump gasoline first and pay later.

What about satisfactions? People in preindustrial communities connect strongly with nature. Daily they experience dawn and dusk, the rising of the moon, the night sky, the murmur of a stream, the season. An adult's labor is highly varied: hunting, fishing, and gathering; clearing, sowing, weeding, and reaping; making tools and pots; spinning and weaving; building shelter. Each person develops many skills – all of the sort that in large, urban communities, we now seek as '*re*-creation'. Consider that each connection with nature, the practice of each skill, and the different rhythm of each activity might provide just the right signal to release the small pulse of dopamine that we experience as a satisfaction.

These small satisfactions are lost in market-dominated communities which tend to shrivel the richness of an individual's labor down to a few repetitive motions. Today a worker may be less likely to serve on an assembly line than to sit isolated in a cubicle and stare at a computer screen – but the variety and challenge of labor have not been recovered. This was thoroughly anticipated by early observers of the market-centered economy. Here is Adam Smith:

> "The man whose life is spent in performing a few simple operations has no occasion to exert his understanding, or to exercise his invention in finding out expedients for difficulties which never occur. He naturally loses, therefore, the habit of such exertion and generally becomes as stupid and ignorant as it is possible for a human creature to become".(Wealth of Nations, 1776).

And here is F.W. Taylor, the father of 'scientific management', describing with pride his contribution to this process of stupidification:

> Owing to the fact that workmen … have been taught … by observation of those immediately around them, there are … perhaps 40, 50, or hundred ways of doing each act in the trade. Now… there is always one best method … which is quicker and better than any of the rest. And this…can only be discovered through a scientific study and analysis of all the methods … together with minute, motion and time study. This involves the gradual substitution of science for rule of thumb throughout the mechanic arts." (Principles of Scientific Management, 1911).

Again, the urban skeptic might object that all these losses are simply the cost of today's many comforts, which are surely satisfying. But here is Herman Melville:

> We felt very nice and snug, the more so since it was so chilly out of doors; indeed out of bed-clothes too, seeing that there was no fire in the room. the more so, I say, because truly to enjoy bodily warmth, some small part of you must be cold, for there is no quality in the world that is not what it is merely by contrast.… If you flatter yourself that you are all over comfortable, and have been so a long time, then you cannot be said to be comfortable any more. But if, like Queequeg and me in the bed, the tip of your nose or the crown of your head be slightly chilled, why then, indeed … you feel most delightfully and unmistakable warm. For this reason, a sleeping apartment should never be furnished with a fire, which is one of the luxurious discomforts of the rich …. (Moby Dick, 1851).

These examples fit a hypothesis that our regulatory systems were selected to seek satisfactions that are small and brief. The best satisfactions occur when an experience exceeds prior expectations; for example, when it contrasts to a prior discomfort. And because of this when an experience is sustained, we soon adapt and it ceases to satisfy. Furthermore, because it is expectation (or hope) that counts, when we achieve a certain level of comfort it will cease to satisfy when we learn that others have more. Although these conclusions can all be learned from the classics of religion and literature, their neural basis is just now being confirmed by recordings from VTA neurons and measurements of dopamine release (e.g., Fiorello et al, 2003; Phillips et al, 2003). These studies may help understand why market-centered society loses the full range of satisfactions, and also why preindustrial societies abandon theirs so quickly when brief exposure to modern goods (shotgun, chainsaw, outboard motor) instantly raises expectations. This renders such groups profoundly vulnerable to hypo-satisfaction -- just as they are vulnerable to new germs (Diamond, 1999) -- and thus prey to the standard addictions of modern life: fast food, alcohol, and drugs.

Note that "adaptation" refers simply to the resetting of response sensitivity to a signal. Although it may turn out badly over time, the outcome is not caused by any low-level error or defect. Consequently, it should not be considered as "inappropriate" or as "dysregulation". Rather, disordered human relationships can drive perfectly normal adaptations of internal physiology into mass-scale pathogenesis – just like disorder within a family can drive a child into diabetic acidosis (Figure 2). The allostasis model clearly identifies a paradox: people are dying, but their internal regulatory mechanisms are intact. So where should we intervene?

Rational therapeutics: where to intervene?

Homeostasis treats low level targets.

Following the homeostasis model, physicians try to restore each parameter to what they consider an "appropriate" level. Therefore, hypertension is treated with drugs that target the three primary effectors of elevated pressure: **(i)** diuretics to reduce blood volume; **(ii)** vasoconstrictor antagonists to dilate the vascular tree; **(iii)** heart rate antagonists to reduce cardiac output. The pharmaceutical industry continues to target myriad molecules that regulate these three mechanisms, and fundamental research widely promises to identify new targets. Thus, WNK kinases and their associated signaling pathways "may offer new targets for the development of antihypertensive drugs" (Wilson et al., 2001); as might the β1subunit of the calcium-activated potassium channel (Brenner et al., 2000), and gene targets of the estrogen receptor β (Zhu et al., 2002).

The same is true for obesity. A recent review lists six neuromodulators that increase feeding and ten that decrease feeding, and then concludes, "a multi-drug regimen that targets multiple sites within the weight-regulatory system may be necessary to achieve and sustain

weight loss in many individuals" (Schwartz, et al., 2000) . Similarly, a study using RNAi in the nematode identifies 305 gene inactivations that reduce body fat and 112 gene inactivations that increase it – and concludes that many of these genes "are promising candidates for developing drugs to treat obesity and its associated diseases" (Ashrafi et al., 2001). The same strategy is proposed for type 2 diabetes and metabolic syndrome (Moller, 2001) and for drug addictions (Laakso et al., 2002).

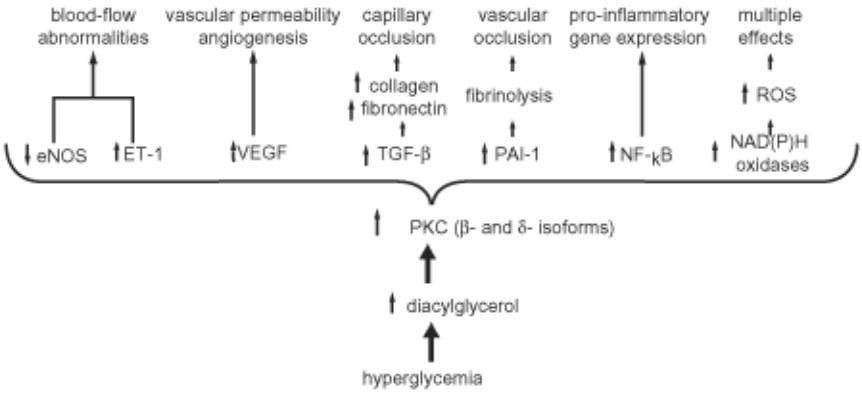

Figure 12. How one signaling molecule (diacylglycerol) can stimulate a cascade of pathogenesis.
Hyperglycemia, part of the "metabolic syndrome", affects numerous signaling molecules, each of which activates its own cascades. As exemplified here, diacylglycerol triggers protein kinase C, and thereby a host of signals, all of which contribute to vascular pathology. Abbreviations: eNos, endothelial nitric oxide synthase; ET-1, endothelin -1; VGF, vascular endothelial growth factor; PAI-1, plasminogen activator inhibitor; TGF-b, transforming growth factor-b. Adapted from Brownlee, 2001.

There are three problems with targeting low-level mechanisms. First, each signal evokes multiply cascaded effects, so even the most specific molecular antagonist will cause a cascade of effects. For example, in hypertension the angiotensin converting enzyme affects all of angiotensin's myriad downstream targets (arteriolar muscle, kidney, and multiple brain sites – see Figure 4), and so also does its widely prescribed inhibitor. Similarly, in type 2 diabetes one effect of hyperglycemia is to elevate the signaling molecule, diacylglycerol. This triggers a host of cascades, some of whose bad effects are shown in Figure 12. Although it might seem advantageous to antagonize an early step, such as the activation of protein kinase C, myriad other cascades with beneficial effects would also be affected (Figure 12). And it turns out that because of such cascading effects, low level inhibitors and antagonists tend to be strongly iatrogenic (Buchman, 2002; Sterling and Eyer, 1981).

Second, the variables targeted for treatment are being driven to their particular levels by concerted signals from the brain (Figure 4) in response to predicted needs (Figure 13A). Consequently, if one signal is suppressed by a drug, the brain compensates by driving all the others harder. Thus, when blood pressure is treated by a diuretic to reduce volume, there are compensatory increases in heart rate and vasoconstriction. These can be treated in turn by beta-adrenergic antagonists, calcium channel antagonists, etc.(Carretero and Oparil, 2000b; Sterling and Eyer, 1981). But adding more drugs to a complex system increases the frequency of iatrogenesis. This is why proposals to treat obesity by a multi-drug regimen at multiple brain sites or to screen 417 genes as drug targets for obesity seem implausible.

Third, there is a cost to performance in clamping a variable to some target level by blocking the effectors designed to modulate it. Clamping renders that variable insensitive to predicted need, which opposes the whole point of physiological regulation (Figure 13A). Thus

clamping blood pressure low with a beta-blocker commonly causes "exercise intolerance" – inability to increase cardiac output when it is needed (Figure 13C).

For all of these reasons, less than 25% of hypertensive patients in the US are controlled. The major problem is considered to be "the very high rate of discontinuance or change in medications: 50-70% … within the first six months... " (Carretero and Oparil, 2000b). These high discontinuance rates are considered to reflect, among other factors, "a combination of adverse drug effects, cost of drugs, and poor efficacy" (Carretero and Oparil, 2000b). Consequently despite their remarkable ingenuity, 30 years of low-level pharmacological treatments for hypertension have not worked. For the same reasons, it seems doubtful that low-level treatments for obesity and metabolic syndrome will be more successful – and already there have been serious adverse effects, for example from fenfluramine and amphetamines.

These problems also apply to pharmacotherapy for mental disorder. Certainly drugs are better than lobotomy: they can be titrated and are reversible over the short run. But when applied for long periods, the "antipsychotic" drugs, which primarily antagonize various dopamine receptors, cause motor disorders. These "tardive" (late appearing) dyskinesias eventually occur in most patients and persist after the drugs are withdrawn (Gelman, 1999). Beyond this devastating iatrogenic effect, drugs that work by antagonizing the major modulators of the nucleus accumbens, amygdala, and prefrontal cortex will, like beta-blockers for blood pressure, reduce responsiveness (Figure 13C). Such drugs would be predicted to cause in stability of intent and to flatten affect (Figure 6). In fact they do, and this is a major reason why patients often refuse to take them (Sterling, 1979; Gelman, 1999).

Allostasis emphasizes higher-level interventions.

The allostasis model defines health as *optimal predictive fluctuation*. A shift in the probability of demand should shift the response, and when the prediction reverses, so should the response (Figure 13A). A system becomes unhealthy when, during long periods of high demand, effectors adapt so strongly that they cease to follow promptly when the prediction reverses (Figure 13B). Drugs can force the response back to the original level, despite continued prediction of high demand, but this compresses responsiveness (Figure 13C). A more rational goal of intervention would be to shift the predicted distribution of demand back toward its original level. This would allow the effectors to naturally reestablish flexible variation around the predicted lower demand, thus preserving the range of responsiveness (Figure 13D).

Can this work for hypertension? Consider that the current authoritative recommendations for treatment are no longer drugs but: **(i)** weight loss; **(ii)** exercise; **(iii)** moderate alcohol consumption; **(iv)** diet reduced in sodium and fat and increased in calcium, potassium, and fiber; (**v)** cease smoking (Carretero and Oparil, 2000b). Weight loss is strongly correlated with reduced blood pressure and is considered to be the most effective of all nonpharmacological treatments. Moderate exercise, such as brisk walking or bicycling thee times per week, may lower systolic pressure by 4-8 mm Hg. The dietary recommendation is based on the "DASH" study, which found overall reductions in blood pressure of 11.4/5.5 mm Hg to a diet rich in fruits, vegetables, and low-fat dairy products, with further reductions of pressure to reduced sodium intake (Sacks et al., 2001). These reductions are said to be "comparable to or greater than those usually seen with monotherapy (i.e., 1 drug) for stage 1 hypertension" (Sacks et al., 2001).

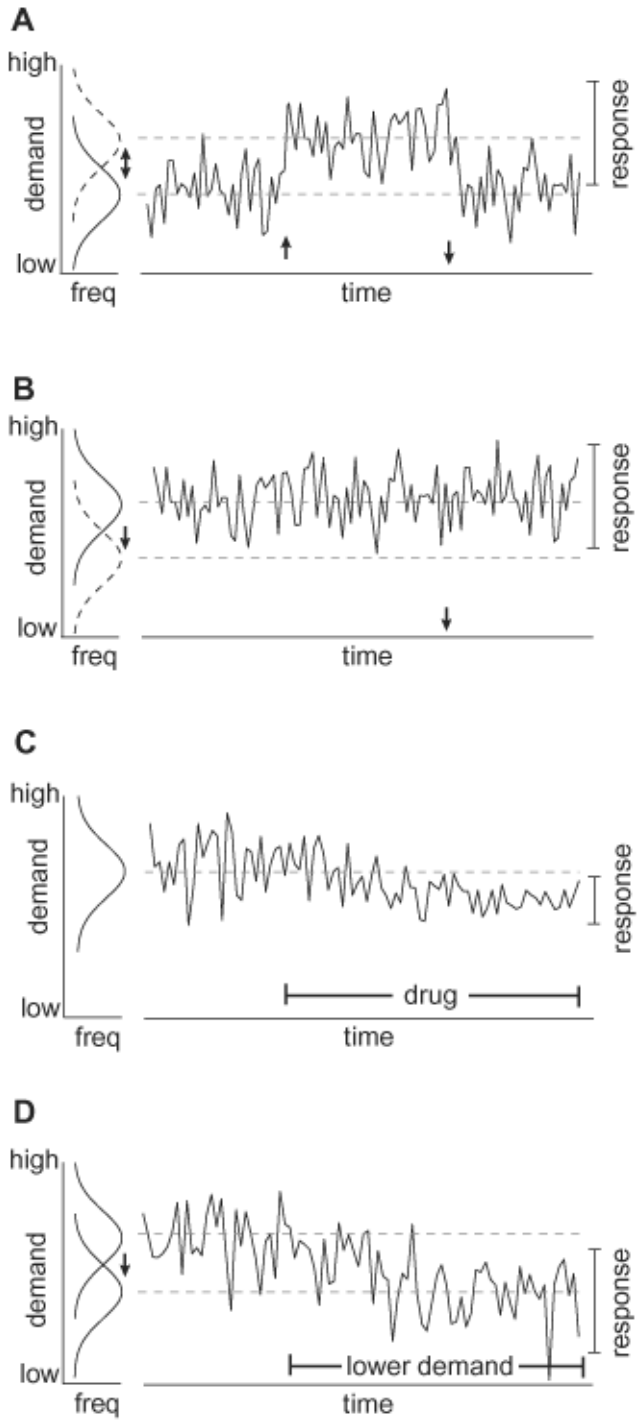

Figure 13. Where to intervene?
A. Healthy system. As demand distribution shifts upward briefly, the response distribution follows to maintain variation centered on most probable demand (see Figure 3). As demand distribution returns to its initial state, the response distribution follows.
B. Unhealthy system. When high demand predominates for long times, the system adapts to this expectation. When demand is reduced briefly, system does not return to the initial state.
C. Standard pharmacotherapy. While demand stays high, drugs that antagonize key effector mechanisms force the response distribution back toward its initial mean. But this reduces responsiveness and evokes iatrogenic effects. This should be expected because the organism must continue to meet elevated demand but with fewer or weaker effectors. This is a common complaint of patients on anti-hypertensive and psychotropic medications.
D. Rational therapy. When demand is reduced for long periods, the system re-adapts to the initial demand distribution. The mean response returns to its initial level while responsiveness is maintained.

These studies document that the distribution of responses can indeed return to lower levels (Figure 13D). But as the DASH study notes, long-term health benefits "will depend on the ability of people to make long-lasting dietary changes, including the consistent choice of lower-sodium foods" and "upon (their) increased availability" (Sacks et al., 2001). This requires, in effect, a sustained victory in the prefrontal cortex of abstract knowledge about what is "good for you" over all the unsatisfied appetites that cause the problem in the first place. Hold onto your McDonald's stock.

The most successful interventions do not deny the sense of need. Rather, they find ways to satisfy it by enlarging positive social interactions and revivifying the sense of connectedness. In the case of coronary heart disease, when patients combined diet and exercise in a group context with a charismatic leader, atherosclerotic plaques regressed over a year, as established by angiography (Ornish et al., 1990). Other outstanding examples are "therapeutic communities", such as the "Twelve-step" programs for treating addictions to alcohol and various illegal drugs.

Therapeutic communities formed the basis for the first mental asylums, such as the Quaker-organized York Retreat (England, 1796) and the first American asylums of the early 19th century, Pennsylvania Hospital and Worcester State Hospital in Massachusetts. These institutions offered "moral therapy", which included physical labor (farm work), well-lit rooms, good food, porter in moderation (alcohol), and lectures on varied topics such as astronomy and literature. Every feature of the program had but one goal, to enhance the patient's sense of well-being (Tuke, 1964; Grob, 1966; Bockoven, 1956). Extraordinarily detailed follow-up studies published in the late 19th Century showed these programs to be highly effective – as were subsequent programs with similar goals, for example, that of the Boston Veteran's Administration in treating posttraumatic stress disorders after WWII (Greenblatt et al., 1955).

This is *not* to argue against treating any mental disorder with a drug. Almost certainly, some disorders will be found to arise from specific molecular defects, just as specific mutations of ion channels, gap junctions, and signaling enzymes, etc. are being identified as causing various neurological disorders (Rosenberg et al., 1997). But just as those defects are fairly rare, and just as molecular defects account for a minor proportion of hypertension, there is likely to be a rather large residual group that will be considered "essential mental illness" – arising from the same core problems of social disruption/disconnection.

This seems particularly applicable to the large group of boys now diagnosed with "attention-deficit, hyperactivity disorder" (ADHD). The prevalence of this diagnosis among boys in the U.S. has reached ~10-30% and it varies inversely with socioeconomic status. The standard drug treatment is methylphenidate – "Ritalin" – an amphetamine analog, or dextroamphetamine. These drugs do help a rambunctious youngster to settle down in the classroom and concentrate for longer periods than he could normally manage. Should this surprise us?

These are the drugs that a street addict takes to obtain his small satisfactions – to quiet his restless prefrontal cortex (Figure 7). And these are the drugs that the long-distance trucker takes to concentrate on the road. So it seems entirely consistent that a boy dosed with amphetamine can concentrate on the assigned task. But over the long term these drugs will certainly cause brain adaptations whose specific consequences cannot be foreseen.

This example seems especially poignant because it arises from disrespecting our greatest evolutionary advantage: our intrinsic diversity of talent and temperament. A proto-scholar might sit still effortlessly in a classroom, whereas a proto-navigator or proto-comedian might not. The allostasis model would not administer the very drugs upon which (outside the classroom) we have declared "war". Rather it would investigate the possible causes of a youngster's restlessness and would intervene by finding activities – beyond sitting still with a book – that *would* absorb him.

Physiological mechanisms of high-level intervention.

Although healing has always relied on the power of positive human interaction (Sterling and Eyer, 1981), to proponents of the homeostasis model this has often seemed like so much mumbo-jumbo. But modern imaging begins to reveal some specific underlying mechanisms. For

example, administration of levo-dopa has been an important therapy for Parkinson's disease, apparently because it is converted to dopamine by midbrain neurons whose release of dopamine in the striatum relieves Parkinsonian symptoms. Yet it turns out that placebo can relieve these symptoms about as well as levo-dopa. Apparently the expectation of therapeutic benefit evokes release of endogenous dopamine, as measured by positron emission tomography (PET)(de la Fuente-Fernández et al., 2001; de la Fuente-Fernández and Stoessl, 2002).

Similarly, placebos can relieve pain, just like opioids. It was suggested that the placebo evokes release of endogenous opioids in the brain. This hypothesis is now supported by PET which shows placebo and opioid analgesia activating the same brain regions (Petrovic et al., 2002). Finally, despite the common wisdom that exercise improves mood and sharpens the mind, there was no clear physiological basis. Now it is found that exercise raises levels of brain-derived neurotrophic factor and other growth factors known to serve synaptic plasticity and learning (Cotman and Berchtold, 2002).

Some will object to the allostasis model by pointing out that death rates have plummeted from many causes, including cardiovascular disease, and that life expectancy has steadily risen. Certainly, diagnosis has benefited enormously from new imaging methods (CAT, MRI, angiography, ultrasound), and biological chemistry has benefited from monoclonal antibodies, PCR, fast assay by mass spectrometry, etc. And surgery has developed less invasive approaches and better management of acute physiology and shock (Buchman, 2002). So, if you are shot or stabbed, or injured in an automobile, your chances of survival are greatly improved, as they also are if you suffer myocardial infarction or stroke.

But by all accounts the financing and organizing of health care are a big problem. The roughly 15% of GNP spent for health goes largely for high-tech treatments, leaving a pittance for the low-tech, human contributions, like nursing and rehabilitation, that require time and emotional contact. Although high-tech medicine can be amazing, medical errors are frequent and lethal, causing annually an estimated 50,000 deaths. Furthermore, high-tech care is unevenly distributed along racial and socioeconomic lines; whereas mammoth pharmaceutical and insurance companies take huge bites for relatively small contributions. Finally, much ingenuity and expense is directed at treating casualties of excessive haste (accidents), aggression, and unsatisfied need.

The allostasis model hints that the biggest improvements in health might be achieved by enhancing public life. The guiding principle would be: do everything that promises to reduce the need for vigilance and to restore small satisfactions. Enhance contact with nature by building more parks and by providing communal opportunities to garden – i.e. not just to look at but to grow flowers and vegetables. Enhance opportunities to walk and cycle by restricting automobile traffic. Prevent this restriction from becoming an annoyance by improving public transportation. Encourage broader participation in sports especially among youth – by constructing public facilities for gymnastics, skating, skate-boarding, climbing, and swimming.

Improve work by acknowledging that no human can be satisfied by performing an unvarying task for eight hours a day, 40 hours per week, 50 weeks per year. Companies, and now even our National Institutes of Health, play a recording, "this phone call may be monitored for purposes of quality control". What this implies, of course, is that the task is so uninteresting that the operators need to be threatened with every call that their supervisor might be listening. For workers at the computer, every keystroke can be similarly monitored. Such humiliating and alienating procedures were introduced recently and could easily be eliminated. The astronomical disparities of income are also recent and could be narrowed while still preserving incentives for

the more energetic and clever. Such proposals are well within our capacity to organize and implement – for they would benefit the rich as well as the poor by reducing everyone's need for vigilance and by expanding everyone's range of small satisfactions.

Acknowledgements.

I thank Joseph Eyer for many wonderful years of collaboration, Charles Kahn for suggesting the Greek roots of allostasis, Jonathan Demb for help with Figures 1, 5 and 13, and Jay Schulkin for his encouragement and patience. I also thank Jay Schulkin, Paul Rozin, Mark Friedman, Ingrid Waldron, Sally Zigmond, Paul Glimcher, Gerd Blobel, Gino Segre, Simon Laughlin, Bettina Hoerlin, Nicole Neff, and Neil Krieger for stimulating discussions and for valuable comments on the manuscript. I am greatly indebted to Sharron Fina for preparing both the manuscript and the figures.

References:

Adolphs R; Tranel D; Damasio H; Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372:669-672.

Alexander RM (1996) Optima for Animals. Princeton NJ: Princeton University Press.

Ashrafi K; Chang FY; Watts JL; Graser AG; Kamath RS; Ahringer J; Ruvkun G (2001) Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. Nature 421:268-272.

Baker L; Minuchin S; Rosman B (1974) The use of beta-adrenergic blockade in the treatment of psychosomatic aspects of juvenile diabetes mellitus. In: Advances in Beta-Adrenergic Blocking Therapy (Snart JA, ed), pp 67-80. Princeton NJ: Princeton Excerpta Medica.

Berkson J (1962) Mortality and marital status. Reflections on the derivation of etiology from statistics. Am J Pub Hlth 52:1318.

Bernik TR; Friedman SG; Ochani M; DiRaimo R; Ulloa L; Yang H; Sudan S; Czura CJ; Ivanova SM; Tracy KJ (2002) Pharmacological stimulation of the cholinergic antiinflammatory pathway. J Exp Med 195:781-788.

Bevan AT; Honour AJ; Stott FH (1969) Direct arterial pressure recording in unrestricted man. Clin Sci 36:329.

Blalock JE (2002) Harnessing a neural-immune circuit to control inflammation and shock. J Exp Med 195:F25-F28.

Blumenthal S; Epps RP; Heavenrich R; Lauer RM; Lieberman E; Mirkin B; Mitchell SC; Boyar-Naito V; O'Hare D; McFate-Smith W; Tarazi RC; Upson D (1977) Report of the task force on blood pressure control in children. Pediatrics 59:797.

Bockoven JS (1956) Moral treatment in American psychiatry. J Nerv Ment Dis 124:167-321.

Brenner R; Peréz GJ; Bonev AD; Eckman DM; Kosek JC; Wiler SW; Patterson AJ; Nelson MT; Aldrich RW (2000) Vasoregulation by the beta1 subunit of the calcium-activated potassium channel. Nature 407:870-876.

Brownlee M (2001) Biochemistry and molecular cell biology of diabetic complications. Nature 414:813-820.

Brunner EJ (2000) Toward a new social biology. In: Social Epidemiology (Berkman L; Kawachi I, eds), pp 306-331. New York NY: Oxford University Press.

Buchman TG (2002) The community of the self. Nature 420:246-251.

Cannon WB (1929) Bodily Changes in Pain, Hunger, Fear and Rage: an Account of Recent Researches into the Function of Emotional Excitement. New York NY: Appleton.

Cannon WB (1935) Stresses and strains of homeostasis. American Journal of the Medical Sciences 189:1-10.

Carretero OA; Oparil S (2000) Essential Hypertension: part I; definition and etiology. Circulation 101:329.

Carretero OA; Oparil S (2000) Essential Hypertension part II: treatment. Circulation 101:446-453.

Cotman CW; Berchtold NC (2002) Exercise: a behavioral intervention ot enhance brain health and plasticity. Trends Neurosci 25:295-301.

Dallman MF (2003) Stress by any other name . . . ? Hormones and Behavior 43:18-20.

de la Fuente-Fernández R; Ruth TJ; Sossi V; Schulzer M; Calne DB; Stoessl AJ (2001) Expectation and dopamine release: mechanism of the placebo effect in Parkinson's disease. Science 293:1164-1166.

de la Fuente-Fernández R; Stoessl AJ (2002) The placebo effect in Parkinson's disease. Trends Neurosci 25:302-306.
Denton D (1993) Control mechanisms. In: The Logic of Life (Boyd CAR; Noble D, eds), pp 113-146. Oxford: Oxford University Press.
Dhingra NK; Kao Y-H; Sterling P; Smith RG (2003) Contrast threshold of a brisk-transient ganglion cell *in vitro*. J Neurophysiol 89:2360-2369.
Diamond J (1993) Evolutionary physiology. In: The Logic of Life (Boyd CAR; Noble D, eds), pp 89-111. New York: Oxford University Press.
Diamond J (1999) Guns, Germs, and Steel: the Fates of Human Societies. New York NY: W.W. Norton & Sons.
Diamond J (2003) The double puzzle of diabetes. Nature 423:599-602.
Elbert T; Rockstroh B (2003) Stress factors. Nature 421:477-478.
Ernst MO; Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.
Eyer J (1975) Hypertension as a disease of modern society. Int J Health Serv 5:539-558.
Eyer J (1977) Prosperity as a cause of death. Int J Health Serv 7:125-150.
Eyer J; Sterling P (1977) Stress-related mortality and social organization. Review of Radical Political Economics 9:1-44.
Fairhall AL; Lewen GD; Bialek W; de Ruyter van Steveninck R (2001) Efficiency and ambiguity in an adaptive neural code. Nature 412:787-792.
Farooqi IS; Matarese G; Lord GM; Keogh JM; Lawrence E; Agwu C; Sanna V; Jebb SA; Perna F; Fontana S; Lechler RI; DePaoli AM; O'Rahilly S (2002) Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. J Clin Invest 110:1093-1103.
Field GD; Rieke F (2002) Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. Neuron 34:773-785.
Fiorillo CD; Tobler PN; Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.
Flier JS (2002) Is brain smpathetic to bone? Nature 420:619-622.
Fluharty SJ (2002) Neuroendocrinology of body fluid homeostasis. In: Hormones, Brain and Behavior pp 525-569. USA: Elsevier Science.
Friedman MI; Ji H; Graczyk-Milbrandt G; Osbakken MD; Rawson NE (1998) Hepatic sensing in the control of food intake: unresolved issues. In: Liver and Nervous System (Häussinger D; Jungermann K, eds), pp 220-229. London: Kluwer Academic Press.
Geisler WS; Diehl RL (2002) Bayesian natural selection and the evolution of perceptual systems. Philos Trans R Soc Lond (Biol) 357:419-448.
Geisler WS; Diehl RL (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. Cognitive Science 118:1-24.
Gelman S (2003) Medicating Schizophrenia: a History. Piscataway NJ: Rutgers University Press.
Glimcher PW (2002) Decisions, decisions, decisions: choosing a biological science of choice. Neuron 36:323-332.
Goethe JW (1833) Faust.
Gordon-Larsen P; Zenel BS; Johnston FE (1997) Secular changes in stature, weight, fatness, overweight, and obesity in urban African American adolescents from the mid-1950's to the mid-1990's. Amer J Hum Biol 9:675-688.

Greenblatt M; York RH; Brown EL (1955) From Custodial to Therapeutic Patient Care in Mental Hospitals: Expllorations in Social Treatment. New York NY: Russell Sage Foundation.
Grob GN (1966) The State and the Mentally Ill. A History of Worcester State Hospital in Massachusetts, 1830-1920. Chapel Hill NC: University of North Carolina Press.
Hammond KA; Diamond J (1997) Maximal sustained energy budgets in humans and animals. Nature 386:457.
Harris AH; Gilliam WJ; Findley JD; Brady JV (1973) Instrumental conditioning of large magnitude. Daily, 12-hour blood pressure elevations in the baboon. Science 192.
Henry JP; Meehan JP; Stevens PM (1967) The use of psychosocial stimuli to induce prolonged systolic hypertension in mice. Psychosomatic Medicine 29:408.
Hill JO; Wyatt HR; Reed GW; Peters JC (2003) Obesity and the environment: where do we go from here? Science 299:853-855.
Hochachka PW; Somero GN (2002) Biochemical Adaptation: Mechanism and Process in Physiological Evolution. New York NY: Oxford University Press.
Ikeda H; Heinke B; Ruscheweyh R; Sandkühler J (2003) Synaptic plasticity in spinal lamina I projection neurons that mediate hyperalgesia. Science 299:1237-1240.
Johnson DDP; Stopka P; Knights S (2003) The puzzle of human cooperation. Nature 421:911-912.
Koob GF; Le Moal M (2001) Drug addiction, dysregulation of reward, and allostasis. Neuropsychopharmacol 24:172-189.
Koshland Jr. DE (1987) Switches, thresholds and ultrasensitivity. Trends Biochem Sci 12:225-229.
Koshland Jr. DE; Goldbeter A; Stock JB (1982) Amplification and adaptation in regulatory and sensory systems. Science 217:220-225.
Laakso A; Mohn AR; Gainetdinov RR; Caron MG (2002) Experimental genetic approaches to addiction. Neuron 36:213-228.
LaBar KS; Gatenby JC; Gore JC; LeDoux JE; Phelps EA (1998) Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron 20:937-945.
Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch C 36:910-912.
Laughlin SB (1994) Matching coding, circuits, cells, and molecules to signals: General principles of retinal design in the fly's eye. Prog Ret & Eye Res 13:165-196.
Lennie P (2003) The cost of cortical computation. Curr Biol 13:493-497.
Lücher C; Frerking M (2003) Restless AMPA receptors: implications for synaptic transmission and plasticity. Trends Neurosci 24:665-670.
Mark VH; Ervin FR (1970) Violence and the Brain. New York NY: Harper & Row.
Marmot M (2000) Multilevel approaches to understanding social determinants. In: Social Epidemiology (Berkman LF; Kawachi I, eds), pp 349-367. New York NY: Oxford University Press.
Mason JW (1968) Organization of endocrine mechanisms. Psychosomatic Medicine 30:565.
Mason JW (1971) A reevaluation of the concept of 'nonspecificity' in stress theory. J Psychiat Res 8:323.

Mason JW (1972) Organization of psychoendocrine mechanisms. In: Handbook of Psychophysiology (Greenfield NS; Sternbach RA, eds), New York NY: Holt, Rhinehart & Winston.
Massey DS; Denton NA (1994) American Apartheid. Cambridge MA: Harvard University Press.
McEwen B (2002) The End of Stress As We Know It. Joseph Henry Press/Dana Press.
Minuchin S (1974) Families and Family Therapy. Cambridge MA: Harvard University Press.
Moller DE (2001) New drug targets for type 2 diabetes and the matabolic syndrome. Nature 414:821-827.
Montague PR; Berns GS (2002) Neural economics and the biological substrates of valuation. Neuron 36:265-284.
Mrosovsky N (1990) Rheostasis The Physiology of Change. New York: Oxford University Press.
Nauta WJH (1971) The problem of the frontal lobe: a reinterpretation. J Psychiat Res 8:167-187.
Nestle M (2002) Food Politics. University of California Press.
NIH publication (1997) The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure.
Orfield G (2001) Schools More Separate: Consequences of a Decade of Resegregation. Harvard Civil Rights Project: Harvard University.
Ornish D; Brown SE; Scherwitz LW; Billings JH; Armstrong WT; Ports TA; McLanahan SM; Kirkeeide RL; Brand RJ; Gould KL (1990) Can lifestyle changes reverse coronary heart disease? Lancet 336:129-133.
Petrovic P; Kalso E; Petersson KM; Ingvar M (2002) Placebo and opioid analgesia - imaging a shared neuronal network. Science 295:1737-1740.
Phillips PEM; Stuber GD; Helen MLAV; Wightman RM; Carelli RM (2003) Subsecond dopamine release promotes cocaine seeking. Nature 422:614-618.
Rieke F; Warland D; de Ruyter van Steveninck R; Bialek W (1999) Spikes: Exploring the Neural Code. Cambridge MA: MIT Press.
Roenneberg T; Merrow M (2003) the network of time: understanding the molecular circadian system. Curr Biol 13:R198-R207.
Rosenberg RN; Prusiner SB; DiMauro S; Barchi RL (1997) Molecular and Genetic Basis of Neurological Disease. Boston MA: Butterworth-Heineman.
Sacks FM; Svetkey LP; Vollmer WM; Appel LJ; Bray GA; Harsha D; Obarzanek E; Conlin PR; Miller ERI; Simons-Morton DG; Karanja N; Lin P-H (2001) Effects of blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet. New Eng J Med 344:3-9.
Sakmann B; Creutzfeldt OD (1969) Scotopic and mesopic light adaptation in the cat's retina. Pflügers Arch 313:168-185.
Saltiel AR; Kahn CR (2001) Insulin signalling and the regulation of glucose and lipid metabolism. Nature 414:799-806.
Saper CB; Chou TC; Elmquist JK (2002) The need to feed: homeostatic and hedonic control of eating. Neuron 36:199-211.
Sapolsky RM (1998) Why Zebras Don't Get Ulcers. New York NY: W.H. Freeman & Co.
Schlosser E (2002) Fast Food Nation. New York NY: HarperCollins.
Schulkin J (1999) The Neuroendocrine Regulation of Behavior. New York NY: Cambridge University Press.

Schulkin J (2003a) Allostasis: a neural behavioral perspective. Hormones and Behavior 43:21-27.
Schulkin J (2003b) Rethinking Homeostasis: Allostatic Regulation in Physiology and Pathophysiology. Cambridge MA: MIT Press.
Schulkin J; McEwen BS; Gold PW (1994) Allostasis, amygdala, and anticipatory angst. Neurosci Biobehav Rev 18:385-396.
Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241-263.
Schwartz MW; Woods SC; Porte D; Seeley RJ; Baskin DG (2000) Central nervous system control of food intake. Nature 404:661-672.
Sterling P (1978) Ethics and effectiveness of psychosurgery. In: Controversy in Psychiatry (Bradie; Brodie, eds), pp 126-160. Philadelphia: W.B. Saunders Co.
Sterling P (1979) Psychiatry's drug addiction. The New Republic, December 8th.
Sterling P (2003) How retinal circuits optimize the transfer of visual information. In: The Visual Neurosciences (Chalupa LM; Werner JS, eds), pp 243-268. Cambridge MA: MIT Press.
Sterling P; Eyer J (1981) Biological basis of stress-related mortality. Soc Sci Med 15E:3-42.
Sterling P; Eyer J (1988) Allostasis: a new paradigm to explain arousal pathology. In: Handbook of Life Stress, Cogintion and Health (Fisher S; Reason J, eds), pp 629-649. New York, NY: J. Wiley & Sons.
Takeda S; Elefteriou F; Levasseur R; Liu X; Zhao L; Parker KL; Armstrong D; Ducy P; Karsenty G (2002) Leptin regulates bone formation via the sympathetic nervous system. Cell 111:305-317.
Taylor CR; Weibel ER (1981) Design of the mammalian respiratory system. I. Problem and strategy. Respir Physiol 44:1-10.
Tracy KJ (2002) The inflammatory reflex. Nature 420:853-859.
Tuke S (1964) Description of the Retreat, 1813. Reprinted with introduction by R. Hunter and I. Macalpine. London: Dawson of Pall Mall.
Valenstein ES (1973) Brain Control. New York NY: John Wiley & Sons.
Valenstein ES (1986) Great and Desperate Cures: the Rise and Decline of Psychosurgery and Other Radical Treatments for Mental Illness. Basic Books
Waldron I (1979) A quantitative analysis of cross-cultural variation in blood pressure and serum cholesterol. Psychosomatic Medicine 41:582.
Weibel ER (2000) Symmorphosis. Cambridge MA: Harvard University Press.
Wilson FH; Disse-Nicodème S; Choate KA; Ishikawa K; Nelson-Williams C; Desitter I; Gunel M; Milford DV; Lipkin GW; Achard J-M; Feely MP; Dussol B; Berland Y; Unwin RJ; Mayan H; Simon DB; Farfel Z; Jeunemaitre X; Lifton RP (2001) Human hypertension caused by mutations in WNK kinases. Science 293:1107-1112.
Wise RA (2003) Brain reward circuitry: insight from unsensed incentives. Neuron 36:229-240.
Zhu Y; Bian Z; Lu P; Karas RH; Bao L; Cox D; Hodgin J; Shaul PW; Thorén P; Smithies O; Gustafsson J; Mendelsohn ME (2002) Abnormal vacular function and hypertension in mice deficient in estrogen receptor beta. Science 295:505-508.
Zimmet P; Alberti KGMM; Shaw J (2001) Global and societal implications of the diabetes epidemic. Nature 414:782-787.