

Wanting, Liking, and Learning: Neuroscience and Paternalism

Colin F. Camerer[†]

A large number of young children die every year from ingesting poisons by accident. Revealed-preference theory, the foundation of microeconomics, has only two concepts to explain these accidents: beliefs and preferences (that is, utilities revealed by choices). In the belief-preference language, all we can say is that a child who died of accidental poisoning either believed the poison was safe, and made a mistake, or that the child preferred death to life, and committed suicide. While economic theory may be comfortable calling these accidents only mistakes or expressions of preference, I am not. My goal in this Essay is to acquaint the legal audience with ideas emerging in neuroscience that could potentially be a richer language for talking about cases like accidental child poisoning and, more broadly, about welfare and paternalism in some limited cases. Further development of the framework could lead to a broader view with wider applicability.

The idea is that three separable neural systems are relevant for choice and welfare: a hedonic “liking” system (welfare), a “wanting” system that guides choice, and a learning system that, ideally, links information stored in the other systems so that people choose to learn what they truly like. Other multiple-process neuroeconomic accounts use different components but have similar implications.¹

[†] Rea A. and Lela G. Axline Professor of Business Economics, California Institute of Technology. Thanks to conference participants, to Meghana Bhatt, Ming Hsu, and Ian Krajbich for research assistance, and to Alice Lin for a presentation on the Berridge work.

¹ See Drew Fudenberg and David K. Levine, *A Dual Self Model of Impulse Control* (Aug 17, 2005), online at http://post.economics.harvard.edu/faculty/fudenberg/papers/dual_self.pdf (visited Jan 17, 2006) (arguing that a simple “dual-self” model explains empirical irregularities in self-control problems and can give a value for commitment in decision problems); Isabelle Brocas and Juan D. Carillo, *The Brain as a Hierarchical Organization* (Aug 2005), online at <http://www-rcf.usc.edu/~juandc/PDFpapers/wp-brain.pdf> (visited Jan 17, 2006) (constructing a model of consumption and time allocation that is solved using tools from mechanism design and economics of information); B. Douglas Bernheim and Antonio Rangel, *Addiction and Cue-Triggered Decision Processes*, 94 *Am Econ Rev* 1558, 1561 (2004) (proposing a model of addiction that is premised on a pathological divergence between choice and preference consisting of mistakes, environmental cues, and attempts by users to minimize their mistakes); George Loewenstein and Ted O’Donoghue, *Animal Spirits: Affective and Deliberative Processes in Economic Behavior* (July 27, 2004), online at <http://gsbwww.uchicago.edu/research/workshops/behavioral/lowenstein.pdf> (visited Jan 17, 2006) (discussing a model of human behavior based on the interaction of a deliberative system and an affective system); George Loewenstein, *Out of Control: Visceral Influences on Behavior*, 65 *Org*

The Essay is organized in several parts. Part I sets the historical stage and then describes a canonical experiment (with mice) on wanting and liking. Part II describes anecdotal evidence of wanting-liking gaps from psychological disorders and choices over time (addiction and credit card use). Paternalism enters in Part III, not with much vigor, but as a potentially efficient response to gaps in wanting and liking. Part IV suggests some directions for positive analysis. Part V concludes.

I. REVEALED PREFERENCES AND WANTING-LIKING

A. History of Thought: The Revealed-Preference Approach Is a Convention

In the early 1900s, Vilfredo Pareto won a debate among economists about the epistemological basis of utility,² a victory later cemented in the 1930s by the ordinalist revolution.

Pareto and his contemporaries, such as Edgeworth and Jevons, all believed in the “concrete deductive method” of John Stuart Mill, which starts with simple empirical regularities or “laws” and makes deductions from those regularities.³ Furthermore, all of these scholars believed that understanding the hedonic basis of utility was necessary, in a form that is now called psychology (but was not called psychology then). As Jevons wrote: “But it is surely obvious that economics does rest upon laws of human enjoyment; and that, if those laws are developed by no other science, they must be developed by economists. . . . [T]he theory of economics must begin with a correct theory of consumption.”⁴

Pareto’s view was the opposite: he thought rooting economic preferences in psychological detail was an unnecessary burden. In an 1897 letter Pareto wrote:

Behav & Hum Dec Proc 272 (1996) (attributing actions against self-interest despite full knowledge to visceral factors defined by a direct hedonic impact and an effect on the relative desirability of different goods and actions).

² Consider Luigino Bruni and Robert Sugden, *The Road Not Taken: Two Debates about the Role of Psychology in Economics* 1, 2 (unpublished manuscript 2002) (on file with author) (explaining that the behavioral economics movement is trying to reverse the approach initiated by Pareto and others to remove psychology from economics).

³ Later, Milton Friedman’s essay, *The Methodology of Positive Economics*, in Milton Friedman, *Essays in Positive Economics* 3 (Chicago 1953), liberated economists from even having to start with *empirical* regularities. He suggested that simple assumptions with no empirical backing could also serve as a starting point, because they might lead to accurate predictions despite being “wrong.” His “F-twist” is best seen as encouraging a reduced-form style of modeling in which incorrect foundational assumptions might lead to good predictions because they approximate predictions from more accurate foundational assumptions that are left unspecified.

⁴ W. Stanley Jevons, *The Theory of Political Economy* 102–03 (Penguin 1970) (arguing that knowledge of the conditions of human utility are necessary to understand economics).

It is an empirical fact that the natural sciences have progressed only when they have taken secondary principles as their point of departure, instead of trying to discover the essence of things. . . . Pure political economy has therefore a great interest in relying as little as possible on the domain of psychology.⁵

Pareto advocated divorcing economics from psychology by simply assuming that unobserved utility is necessarily revealed by choice (though stay tuned): “[W]e are concerned only with certain relations between objective facts and subjective facts, principally the tastes of men. *Moreover, we will simplify the problem still more by assuming that the subjective fact conforms perfectly to the objective fact.*”⁶

While Pareto was comfortable severing economics from psychology, some of his contemporaries were not. Writing about him in 1926, shortly after Pareto died, Edgeworth said:

[Pareto’s] The *Manuale* is distinguished by the original idea of treating the laws of demand and supply, or rather the “curves of indifference” from which those may be deduced, as objective, capable of being ascertained by external observation without the psychological knowledge obtained through sympathy. In short, the economist may be a *solipsist*. The conception has been criticised . . . as a needless abandonment of one large source of information.⁷

Remember that Pareto did not win this argument about how to proceed by mathematical proof or data. The equation of utility with choice was not a scientific discovery on par with a powerful theorem or solid empirical regularity. Pareto simply asserted that, as a matter of convenience, it was okay to give up on understanding the “essence of things.” Pareto thought economics could start with “secondary principles” like utility-maximization (which, ironically, are now taught to graduate students as “*first* principles”). Pareto’s turn—the definition of utility as a quantity revealed by expressed preference equation—

⁵ Bruni and Sugden, *The Road Not Taken* at 20 (cited in note 2).

⁶ Vilfredo Pareto, *Manual of Political Economy* 103 (Augustus M. Kelley 1971) (emphasis added) (arguing in favor of simplifying problems to the objective facts by considering only repeated actions where the actor has corrected previous mistakes and has determined his or her exact preferences so that subjective preferences are identical to objective facts).

⁷ Francis Y. Edgeworth, *Pareto, Vilfredo*, in Henry Higgs, ed, 3 *Palgrave’s Dictionary of Political Economy* 711 (MacMillan 1926) (surveying Pareto’s works and criticism of those works) (internal citation omitted). Solipsism is “a theory holding that the self can know nothing but its own modifications and states . . . [and] that the self is the only existent thing.” *Webster’s Third New International Dictionary* 2170 (Merriam-Webster 1993). That is, the world is a figment of one’s imagination. Solipsism is not taken seriously as a philosophical stance; it is just a challenge for philosophers to explain precisely why it is wrong.

was an agreement on a convention for how to do economics, like the rules of tennis, or assuming away friction in physics.

Of course, the equation of utility and choice was also historically justifiable in the early 1900s because measuring utility directly, at that time, was as far-fetched as, say, sending people to the moon or cloning animals. In Pareto's day there was no ability to place electrodes into a monkey or human brain and record firing rates of individual neurons, to alter genes in mice and see what those mice can't do, to record online brain activity with 3-mm spatial resolution using PET or fMRI, to identify areas of brain damage (lesions) in humans and see what skills those humans lack, to genotype people and correlate genes with behavior and skill, to give humans drugs and see how their behavior changes, or to create "temporary lesions" using transcranial magnetic stimulation (TMS).

The huge advances in all these tools force the reader to draw one of two conclusions: (1) Pareto was right in his own time—he and his peers shouldn't have tried to measure utility directly, it was hopeless—and his intellectual heirs are still right today, because their tools can't possibly measure utility, or (2) Pareto was right *then* because the tools didn't exist, but the neo-Paretians are wrong *now* because we have tools to measure utility that didn't exist in Pareto's time (and the tools will only get better). Conclusion (1) might be right, but betting that something is impossible in science is dangerous and neglects the option value of exploring wild ideas. Conclusion (2) is the progressive conclusion with upside potential.

The ideas in this Essay are not an insistence that revealed-preference theory has not been useful. It has, and will continue to be. But the usefulness of one theory tells us nothing about whether an alternative approach might prove equally useful.

B. Beyond Revealed Preference: Types of Utility

The distinction between utility and choice has been revived in two ways.

First, Kahneman, Wakker, and Sarin noted that one can conceptually distinguish four types of utility: experienced utility (the hedonic sensation at the time of consumption that Jeremy Bentham had in mind); remembered utility; forecasted utility (a forecast of experienced utility); and finally, the familiar notion of decision utility (numbers an observer could use to rank an agent's revealed preferences).⁸

⁸ See Daniel Kahneman, Peter P. Wakker, and Rakesh Sarin, *Back to Bentham? Explorations of Experienced Utility*, 112 Q J Econ 375, 376–77 (1997); Daniel Kahneman, *New Challenges*

Thinking like a psychologist, it is easy to see why these types of utility might differ. Remembered and experienced utility can differ because of nostalgic biases in memory (for example, the inability to recreate the pain of childbirth in imagination). Forecasted utility could be subject to “projection biases” relative to experienced utility—people buy more groceries when they’re hungry,⁹ and overestimate how badly they will feel if their favored candidate loses an election.¹⁰ Most fundamental for my purposes, decision utility could deviate from experienced utility if people have not yet learned what they like (for example, children), or if there is some dissociation between wanting and liking.

The revealed-preferences view basically equates welfare (experienced utility) with choice (decision utility), as a matter of the *definition* of welfare. Is this the best we can do? If my neighbor thumps his head repeatedly with a ball-peen hammer, do I have no alternative but to infer that hammering his head with a ball-peen hammer is the most fun he can have? A more sensible direction is to study the relation between experienced and decision utility empirically.

The second revival of the debate about utility is work by the psychologist Kent Berridge and colleagues on “wanting,” “learning,” and “liking” that addresses the measurement challenge, suggesting some ways in which experienced utility (liking) and decision utility (wanting) might differ, depending on learning that potentially links the two.¹¹

My suggestion in this short Essay is that the wanting-learning-liking distinction provides a potential basis, grounded in neuroscience, for asserting that choices are not always utility-maximizing. In fact, the learning part of this framework was clearly endorsed by Pareto as a justification for assuming that the “subjective fact” of utility and the “objective fact” of revealed preference are the same. A longer quotation containing the short excerpt quoted above, reads:

We will study the many logical, *repeated* actions which men perform to procure the things which satisfy their tastes. . . . [W]e are concerned only with certain relations between objective facts and subjective facts, principally the tastes of men. Moreover, we will simplify the problem still more by assuming that the subjective

to the *Rationality Assumption*, 150 *J Institutional & Theoretical Econ* 18, 21 (1994) (discussing the distinction between the two primary notions of utility: experienced and decision).

⁹ See George Loewenstein, Ted O’Donoghue, and Matthew Rabin, *Projection Bias in Predicting Future Utility*, 118 *Q J Econ* 1209, 1215 (2003).

¹⁰ See Daniel T. Gilbert, et al, *Immune Neglect: A Source of Durability Bias in Affective Forecasting*, 75 *J Personality & Soc Psych* 617, 625 (1998).

¹¹ See Kent C. Berridge and Terry E. Robinson, *Parsing Reward*, 26 *TRENDS in Neurosciences* 507, 510–12 (2003) (discussing the interactions of wanting, learning, and liking in conditioned stimuli as motivational magnets, cue-triggered “wanting,” response reinforcement, and cognitive incentives).

fact conforms perfectly to the objective fact. This can be done because *we will consider only repeated actions to be a basis for claiming that there is a logical connection uniting such actions.*¹²

Pareto clearly implicates learning as a mechanism that brings wanting and liking together:

A man who buys a certain food for the first time may buy more of it than is necessary to satisfy his tastes, price taken into account. But in a second purchase he will correct his error, in part at least, and thus, little by little, will end up by procuring exactly what he needs. *We will examine this action at the time when he has reached this state.* Similarly, if at first he makes a mistake in his reasoning about what he desires, he will rectify it in repeating the reasoning and will end up by making it completely logical.¹³

Thus, the standard assumption, that choices reveal “true” utilities, is clearly endorsed by Pareto only for choices that are repeated often enough to ensure that learning has occurred. The wanting-learning-liking framework can therefore be seen as a proper generalization of revealed-preference theory and, in fact, it is perhaps the kind of general theory Pareto had in mind in the first place. Bruni and Sugden note that the “discovered preference hypothesis” is a revival of Pareto’s idea.¹⁴ They assert that rational choice theory applies *after* preferences have been “discovered” by trial-and-error learning and reflection. But this leaves a large and important hole: what happens *before* learning?

Besides generalizing revealed-preference theory, the wanting-learning-liking framework provides a potential way to talk scientifically about how restricting or imposing choices (paternalistically) could improve welfare. The core idea is simple: if there are separate systems for recording liking, expressing wanting, and for learning to want what the brain likes, then paternalism could be justified if the wanting system produces choices that are not later liked, *and* if a paternalistic correction produces choices that are unwanted by an agent but will be liked by her, *or* that are wanted but not liked, *and* if the correction does not cause other harms (or much harm to rational agents).

¹² Pareto, *Manual of Political Economy* at 103 (cited in note 6) (emphasis added).

¹³ *Id.* (emphasis added).

¹⁴ See Bruni and Sugden, *The Road Not Taken* at 7 (cited in note 2) (labeling the claim that rational-choice theory can predict behavior when certain criteria are satisfied as the “discovered preference hypothesis”). See also Charles R. Plott, *Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis*, in Kenneth J. Arrow, et al, eds, *The Rational Foundations of Economic Behavior* 225, 226 (St. Martin’s 1996) (suggesting a theory that rational choices evolve through three stages reflecting experience and practice).

By the way, note that I am truly not an eager paternalist. I could care less about paternalism, per se. Paternalism is only interesting as a challenge for behavioral economics, both normatively and positively (to make sense of paternalism that is currently practiced, and changes in paternalism across time and geography).

At this point, the neuroscientific basis for any practical paternalism is flimsy. But progress in neuroscience is very rapid. And of course, there is a genuine danger of a slippery slope in moving from paternalism that is designed to be unobjectionable, because it helps some and doesn't harm others, to paternalism that is scary, intrusive and wrong-headed. The political economy of capture by special interest groups also should make us wary, in practice, that opening the door to even the mildest paternalism could get out of hand. But it can't hurt to instigate a purely academic debate about how limited paternalism might someday be scientifically justified, subject to a reasonable consensus of scientists, lawyers, and citizens.

C. Wanting-Learning-Liking Illustrated

I will motivate my discussion by starting with details of a simple experiment with mice by Peciña and colleagues.¹⁵ Their experiment uses mice with a dopamine transporter (DAT) “knockdown” due to genetic engineering. As I understand it (which is not well), the knockdown mice produce only about 10 percent as much dopamine transporter (which “takes up” or recycles dopamine) as the control mice, but as a result have 70 percent more extracellular dopamine sloshing around between neurons.

When freely allowed access to food and drink, the DAT-knockdown mice consumed twice as much as the control mice. This difference, by itself, does not tell you whether they liked the consumption or just wanted it more.

The mice were also put through a standard “runway” task, in which they are kept hungry for eight hours, then trained to move down a short runway from a start box to a goal box containing food. During a learning period, the DAT-knockdown mice ran about twice as quickly down the runway as the control mice and were much less likely to reverse course or pause and sniff. Because the knockdown and control mice are about equally fast in a pretraining phase, and

¹⁵ See Susana Peciña, et al, *Hyperdopaminergic Mutant Mice Have Higher “Wanting” But Not “Liking” for Sweet Rewards*, 23 *J Neuroscience* 9395 (2003) (finding that chronically elevated extracellular dopamine in mutant mice facilitates “wanting” but does not increase “liking”).

equally fast after a lot of training, this difference during learning suggests that extra dopamine facilitates faster learning.¹⁶

These experiments show that the DAT-knockdown mice eat more, and learn there's food waiting in the goal box faster than the control mice do, but do they actually *like* food and liquid more? Here we have no choice but to skate out onto thin empirical ice: how do you measure liking?

If you are open-minded about the possible difference between choice, which is observable, and liking (true utility), which is not as readily observable, then you should accept that any measure of true utility won't be as sharp and agreeable as observing choice is. If you aren't willing to accept this sad fact, stop reading now.

Peciña and colleagues measure liking and disliking using physical movements of the mice—tongue protrusions and paw licking to represent liking, and gaping (mouth-opening, like a yawn), screwing up the face, and arm flailing to represent disliking. Your first reaction might be that these are not serious scientific measures. But they are taken very seriously by people who study animals, because they are the best available measures of liking¹⁷ (putting observed choice aside, of course; if you use choice to infer liking we're back to square one with Pareto). These expressions and movements are also very similar across mice, primates, and human infants; in fact, they are so similar that researchers use an equation ($d = .26(w)^{.32}$) to predict the duration (d) of a facial expression (in milliseconds) in terms of a species' body weight (w) (in kilograms).¹⁸

In any case, these measures of liking and disliking show that the DAT-knockdown mice do *not* like the sucrose solution more than the control mice. In fact, 1.0 M sucrose, the largest concentration, is liked significantly less by the knockdown mice.¹⁹

¹⁶ See Samuel M. McClure, Nathaniel D. Daw, and P. Read Montague, *A Computational Substrate for Incentive Salience*, 26 *TRENDS in Neurosciences* 423, 424 (2000).

¹⁷ Methods like this are used by pediatricians to judge how much pain children are in. An "oucher" scale is used by pediatricians to map facial expressions onto a numerical scale. See, for example, *How to Use the Oucher*, online at <http://www.oucher.org/differences.html> (visited Jan 17, 2006). See also, for example, Rita Rubin, *Kids' Distress Is No Small Thing*, *USA Today* D1–2 (May 9, 2005). This can be thought of as a utility function that maps facial expressions into numerical "true disliking." If the goal is to prescribe an optimal dose of painkiller, you can't just let a child "choose" a dose. Therefore, inferring children's disutility—pain—from choice is a bad idea. Inferring their pain from facial expressions is a good idea.

¹⁸ See Berridge and Robinson, 26 *TRENDS in Neurosciences* at 509 (cited in note 11).

¹⁹ A concern in the study is that the eating task used regular food (Purina Rat Chow) and water; the learning task used Froot Loops; and the liking measures used sucrose solution. I presume Peciña and colleagues are comfortable with the working hypothesis that knockdown mice do not have different tastes for any of these substances than control mice do. See also Kent C. Berridge and Terry E. Robinson, *What Is the Role of Dopamine in Reward: Hedonic Impact*,

This is just one study that shows a potential dissociation between wanting and learning, which the knockdown mice reveal more of, and liking, which they reveal less of. They eat and drink more, but don't seem to like it.

The same kinds of wanting-liking dissociations have been shown in experiments with humans.²⁰ Dopamine-receptor antagonists appear to suppress cigarette-smoker wanting without changing (self-reported) pleasure from cigarettes.²¹ Environmental cues create learned association with pleasurable drug use; when these cues are presented they can create rapid craving, just like the conditioned stimulus produced salivation in Pavlov's dogs. These cues are clearly affecting wanting, though they may not affect eventual liking.

An important property of the wanting and liking systems is that they may react implicitly, without consciousness. For example, drug addicts will work (lever-pressing) for doses of stimulants or morphine that are so low that the addicts report no subjective effects and have no measurable autonomic responses. That is, the addicts detect enough of the drug to keep wanting the low doses, but don't think they are getting any dose at all.²²

Berridge and Robinson suggest general regions of the mouse brain that are responsible for wanting, liking, and learning.²³ Nobody has constructed such a neurogeographical map of utility for the human brain, and it will be a long time before we have a good one. (Furthermore, utility for food, shopping, drugs, sex, and golf may all work differently, so we'll need a book of brain maps.) But the point is that such a map has already been proposed for mice, based on a relatively recent and modest amount of research. So it is conceivable that at least a rough

Reward Learning, or Incentive Salience?, 28 *Brain Rsrch Revs* 309, 350 (1998) (finding that dopamine-depleted rats still "like" rewards and simply fail to "want" rewards).

²⁰ See Berridge and Robinson, 26 *TRENDS in Neurosciences* at 508 (cited in note 11) (discussing how drug addicts will work for low doses of stimulants or morphine despite no subjective effect). See also Carl L. Hart, et al, *Methamphetamine Self-administration by Humans*, 157 *Psychopharmacology* 75, 80 (2001) ("[S]everal investigators have reported data demonstrating that drug-related 'positive' subjective effects do not entirely account for drug-taking behavior and some have suggested that other factors should also be considered.").

²¹ See Lisa H. Brauer, et al, *Haloperidol Reduces Smoking of Both Nicotine-containing and Denicotinized Cigarettes*, 159 *Psychopharmacology* 31, 34-35 (2001) (discussing experimental results where haloperidol, a dopamine-receptor antagonist, was administered and significantly reduced smoking behavior but did not affect how much the subjects liked smoking).

²² See Berridge and Robinson, 26 *TRENDS in Neurosciences* at 508 (cited in note 11).

²³ See id at 510 (containing a schematic explanation of regions of the brain that are linked to wanting, liking, and learning). Their analogous terms are motivation (wanting) and emotion or affect (liking). Id at 508.

understanding of wanting, learning, and liking could emerge for humans, probably at first for a specialized domain like food.²⁴

Having separate systems for wanting, learning, and liking makes sense from an evolutionary point of view. Take food as a simple case. First note that unless tastes for specific foods are truly genetically hardwired, some kind of learning system is necessary to connect the sensory properties of food—how it smells, feels, and looks—to how much an animal likes specific foods. This basic learning is what infants do for months—putting everything in sight in their mouths to see how things taste. An infant who tastes dirt doesn't anticipate *liking* dirt—she just exhibits *wanting*, which is really just wanton exploration (to create inputs for the learning system). Then she gradually learns the difference between brown, crumbly stuff that tastes bad—dirt—and brown, crumbly stuff that tastes good—gooey brownies.

The fact that the learning system exists at all is *prima facie* evidence that wanting and liking are not automatically linked at birth; learning is there for a reason, and the reason is to educate the wanting system about what is liked. But why would wanting and liking occur in separate brain regions? Berridge and Robinson propose that liking for food, sex, and warmth rely, at least to some extent, on quite different sensory systems for processing smell, heat, and taste. A separate wanting system is needed to combine these signals into a “common currency,”²⁵ in order to make tradeoffs between goods that are liked in (sensorily) different ways.

Wanting and liking may also deviate, usefully, in the special case of pregnancy. When my wife was pregnant, she had a fierce craving for lamb—one of her favorite foods—then almost vomited after taking one bite. Her wanting system had tuned to her own tastes but the circuitry protecting the delicate fetus registered rapid dislike.

Learning is the mechanism that trains wanting about what is actually liked by trial-and-error. However, in the modern economy, many of the goods people purchase are much more complex to process and rep-

²⁴ A similar body of research in humans focuses attention on “actor-critic” models of temporal difference (TD) learning. In TD learning, an organism learns a value function according to an adaptive updating equation driven by the prediction error (or temporal difference). John O'Doherty and colleagues have located areas of striatum (in the brain's temporal lobe) that correspond to “critic” temporal differences, and other areas that correspond to an “actor's” expected reward or forecast. John O'Doherty, et al, *Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning*, 304 *Science* 452 (2004). In terms of this Essay, the actor is wanting, and the critic system is learning, which adjusts wanting so it learns the true value (liking).

²⁵ See Peter Shizgal, *On the Neural Computation of Utility: Implications from Studies of Brain Stimulation Reward*, in Daniel Kahneman, Ed Diener, and Norbert Schwarz, eds, *Well-Being: The Foundations of Hedonic Psychology* 500, 509–10 (Russell Sage 1999) (discussing the necessity and existence of a currency function to break down different stimuli into decision utility (wanting)).

resent than are simple foods. The learning necessary to link wanting and liking of such goods through trial-and-error reinforcement is either impossible (planning the perfect once-in-a-lifetime wedding) or very slow.

Other mechanisms may substitute for direct learning about complex choices (personal advice, imitation, advertising). But it is not likely that the human brain has developed a special patch for unique important decisions that links imitation directly to actual liking. At best, humans would have a capacity to learn at a young age whose advice to take (i.e., they develop a tightly coupled wanting-liking loop for good advice; and advice that is taken then triggers wanting for a good, which is a sensible forecast of later liking because that's what "good advice" means). Such a system might work well, but probably not perfectly, and might be vulnerable to exploitation by profit-maximizing firms (for example, celebrity endorsements might hijack such a system).

II. HUMAN MISFIRES IN WANTING AND LIKING

A. Disorders

The most compelling examples of wanting-liking gaps are pathological disorders that people clearly would like to change if they could.

Obsessive-compulsive disorder (OCD) is a psychiatric disorder characterized by obsessive thoughts and compulsive actions, such as cleaning, checking, counting, or hoarding.²⁶ "Cleaners" report an intense feeling that their hands are not clean, and must be washed over and over to reduce that feeling. "Checking" is similar. My father used to compulsively check the four oven stove dials in our house before leaving the house—he would point to each and say "off, off, off, off," partly coping with embarrassment by making fun of his mild checking compulsion. Sometimes he would stop the car, halfway out the driveway, and get out to double-check that the front door was locked.

In the revealed-preferences approach, we would infer that compulsive cleaners have a high utility for clean hands. Some sensible predictions may come from this inference—for example, cleaners spend a larger income share on cleaning than noncleaners do.

But the revealed-preferences view only gets you so far. If you give compulsive cleaners a serotonin reuptake inhibitor (SSRI) like fluoxetine (Prozac) or paroxetine (Paxil), a majority of cleaners find

²⁶ Note that the DAT knockdown mutant mice exhibit "sequential super-stereotypy" in grooming that is remarkably similar to OCD cleaning in humans. See Kent C. Berridge, et al, *Sequential Super-stereotypy of an Instinctive Fixed Action Pattern in Hyper-dopaminergic Mutant Mice: A Model of Obsessive Compulsive Disorder and Tourette's*, 3(4) BMC Biology (2005), online at <http://www.biomedcentral.com/content/pdf/1741-7007-3-4.pdf> (visited Jan 17, 2006).

that their symptoms are diminished. Of course, we can easily stretch revealed-preference language to allow “state-dependence” of cleaning-utility, which is altered by these drugs. But patients actively seek out these drugs to change their preferences. To talk about the demand for a cure, we need to posit a “metapreference” in which cleaners prefer to be symptom-free than to be compulsive. Their metapreference is expressed in demand for Prozac or Paxil.

A big advantage of revealed-preference theory is simplicity. But adding concepts of state-dependence and metapreference quickly makes the theory complicated, not simple. And even if enriched revealed-preference theory can accommodate these phenomena, the details of how preferences are linked to genes and childhood behavior, how “states” like SSRI drugs work, and whether agents understand the source of their behavior and the state-dependence, which all matter for good prediction, require a lot of empirical input from outside economics per se.

The language of wanting and liking could be even simpler. In wanting-liking terms, an OCD patient wants to clean but does not like it. Most say that they do not like cleaning per se—it can be very time-consuming, doesn’t bring pleasure, and direct measurement (facial expressions, self-report, imaging of striatum activity) would probably show that liking is low. But such a patient is motivated, compulsively, to clean (she “wants” to). SSRI drugs reduce wanting. Patients take medication to override their wanting system.

The OCD example is a relatively easy one. Of course, the wanting-liking framework is useful only if there are more common behaviors, with substantial economic consequences, that can be seen as reflecting wanting-liking gaps. Most of the obvious examples have to do with time preference, which I’ll discuss in the next Part.

Another interesting example is the behavior of “shopaholics.” Compulsive shopping is a disorder with formal diagnostic criteria analogous to those for compulsive drug use or gambling.²⁷ I will use it as a casual example just because it is a name for a possible disconnect between the transitory pleasure from buying goods and the later pleasure from consumption. A different class of behavior involves “virtuous intentions” like buying exercise equipment or a health club plan.²⁸

²⁷ See Lorrin M. Koran, et al, *Citalopram Treatment of Compulsive Shopping: An Open-Label Study*, 63 *J Clin Psychiatry* 704, 704 (2002) (showing that treatment with the antidepressant citalopram (Celexa) reduces shopping compulsion).

²⁸ Stefano Della Vigna and Ulrike Malmendier, *Paying Not to Go to the Gym* 22 (unpublished manuscript 2005), online at <http://emlab.berkeley.edu/users/sdellavi/wp/gymemp05-04-20.pdf> (visited Jan 17, 2006) (explaining the irrational behavior of gym users’ contract choices based on their overestimation of future efficiency or self-control).

An economist's definition of shopaholism is systematically buying goods that are not later consumed.²⁹ Such a disorder could result from a wanting system buying goods that the liking system doesn't appreciate (like closets full of shoes that are never worn). Of course, in the language of belief and preference, one could argue that shopaholics are mistaken in their beliefs about future consumption (or their budget of time to use goods). Then the challenge is explaining why they don't learn from a chronic pattern of purchase and disuse.

An extreme form of shopaholism is an OCD disorder called "collecting" or "hoarding" in which people collect items to an extreme. A milder form is probably present in all of us: most people have at least one class of goods or services they enjoy shopping for, but they don't have enough time or inclination to actually consume. Mine is books—I love getting books, but never get around to reading them all. My wife is obsessed with buying marinades. We have around twenty, which are invariably opened once and then sit patiently for years, three-quarters full.

B. Wanting and Integration of Momentary Liking over Time

Food is a useful kind of good for exploring the neural basis of wanting and liking, because mechanisms are common across species and consumption is immediate. But for most goods and services of economic interest to humans—education, savings, durables, credit card use—costs and benefits are spread out over time. It may be very uncomfortable to learn to windsurf or to struggle through college (the liking system is not happy during the learning), but rewards come in the form of later liking (windsurfing) or future liking of goods that can be bought with increased wages from a college education. The wanting system has to integrate these likes and dislikes over time to determine overall liking.

One method for integrating liking is exponential discounting of future rewards, which corresponds to dynamic consistency in choice and has much normative appeal.³⁰ A more general approach is models of present-biased preferences (for example, quasihyperbolic discounting),³¹ which splice a preference for immediacy and conventional dis-

²⁹ The closest clinicians come to an economic definition is the "[f]requent buying of more than can be afforded, more than is needed, or for longer periods of time than intended." Kim Bullock and Lorrin Koran, *Psychopharmacology of Compulsive Buying*, 39 *Drugs of Today* 695, 696 (2003).

³⁰ For short-lived episodes where discounting is likely to be weak (for example, evaluating an operation, a movie or a weekend getaway), Kahneman, Wakker, and Sarin propose and axiomatize a rule that takes the integral of momentary liking over time. 112 *Q J Econ* at 390 (cited in note 8).

³¹ See Ted O'Donoghue and Matthew Rabin, *Doing It Now or Later*, 89 *Am Econ Rev* 103, 106 (1999) (discussing preferences as time-inconsistent or present-biased, meaning that a per-

counting of future rewards.³² These preferences can be characterized as reflecting behavior of a wanting system that weights current liking too heavily, compared to exponential discounting. Let's consider two examples: addictive drugs and credit card spending.

1. Addictive drugs.

The hallmarks of addictive drugs are tolerance (larger doses are needed to produce equal hedonic sensations) and withdrawal (unpleasant physical sensations during abstinence). A natural and fruitful way to model these effects is through state-dependence of current utility on past consumption;³³ or a richer view in which states are numerical summaries of use histories that go up or down when addicts use or abstain.³⁴ In any case, a central issue is whether potential addicts rationally anticipate how current consumption creates tolerance and possible future withdrawal. A key prediction of the rational addiction model is that current consumption responds to expected future price changes, as it appears to do.³⁵ Aggregated data are too coarse, however, to tell whether some potential addicts refuse to start, or whether active users quit, in the face of future price increases.

In the wanting-liking view, addiction means that wanting can be created by a desire to avoid the pain of withdrawal, or is triggered by craving ("self-medication"), but the drug administration is not actually

son's relative preference for well-being at an earlier date versus a later date will change depending on when he or she is asked); David Laibson, *Golden Eggs and Hyperbolic Discounting*, 112 Q J Econ 443, 445–46 (1997) (discussing how decisionmakers foresee their own dynamically inconsistent discount rates over short horizons (high) and long horizons and commit themselves to nonliquid assets to constrain their future selves).

³² See Samuel M. McClure, et al, *Separate Neural Systems Value Immediate and Delayed Monetary Rewards*, 306 Science 503, 506 (2004) (describing how separate functions exist in the brain associated with immediate preferences (possibly reflecting evolutionary adaptations) and abstract future preferences and how these separate functions may explain the idiosyncrasies of human preferences).

³³ See generally Gary S. Becker and Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J Polit Econ 675 (1988) (describing a theory of rational addiction that is partially based on what state of addiction an addict is in).

³⁴ See Bernheim and Rangel, 94 Am Econ Rev at 1565 (cited in note 1) (modeling addictive states by incrementing or decrementing the state of addiction based on decisions to use or not to use—a higher state corresponds to greater addiction, for example, someone in state 8 is more addicted than someone in state 5).

³⁵ See Gary S. Becker, Michael Grossman, and Kevin M. Murphy, *An Empirical Analysis of Cigarette Addiction*, 84 Am Econ Rev 396, 397 (1994) (discussing how a one-period price increase of 10 percent reduces consumption in the previous period by 0.6 percent). See also Jonathan Gruber and Botond Köszegi, *Is Addiction "Rational"?* *Theory and Evidence*, 116 Q J Econ 1261, 1273 (2001) (showing that consumers increase consumption by hoarding cigarettes in anticipation of future tax increases); Angela A. Hung, *A Theoretical and Empirical Study of Addiction* 86–87, unpublished PhD dissertation, California Institute of Technology (2001) (finding that cigarette consumers are forward-looking in respect to prices).

pleasurable. (At best, it cures the withdrawal symptoms; but other cures, like rehabilitation or methadone, which do not feed the addiction as self-medicating doses do, may be available.) As Bernheim and Rangel suggest, it is sensible to entertain the idea that when craving, addicts' wanting systems are not fully accounting for the future dislike (or internality) created by current use. This phenomenon is not easily explained by hyperbolic discounting, because exogenous cues—like seeing drug use in a film—can trigger craving.³⁶

2. Credit card spending.

For the sake of argument, suppose that at least some consumers charge too much on their credit cards, in the sense of revealing a wanting-liking gap (at least ex post). This may be due to poor understanding of how interest rates work (for example, how rapidly interest compounds) or insufficient appreciation for the status quo bias of sticking with one card after an introductory “teaser rate” period lapses.³⁷

Another possibility, in wanting-liking terms, is that the brain likes walking out of a store (or clicking an internet radio button, or calling an 800-number and imagining goods being shipped), but the wanting system does not anticipate the “pain of paying” when the credit card statement arrives.³⁸ That is, current wanting does not integrate future dislike.

III. PATERNALISM

I adopt the textbook definition of paternalism as “the interference of a state or an individual with another person, against his will, and justified by a claim that the person interfered with will be better off or protected from harm.”³⁹

In my terms, paternalism could be justified, in terms of a person's own welfare, if the wanting system does not produce what the liking

³⁶ See David Laibson, *A Cue-Theory of Consumption*, 116 Q J Econ 81, 82–83 (2001) (explaining a cues model that assumes that the presence of cues (sensory inputs) linked to past consumption affects habit formation).

³⁷ See Haiyan Shui and Lawrence M. Ausubel, *Time Inconsistency in the Credit Card Market* 1, 27–28 (14th Annual Utah Winter Finance Conference, May 2004), online at http://papers.ssrn.com/sol3/papers?.cfm?abstract_id=586622 (visited Jan 17, 2006) (explaining that a hyperbolic model is necessary to explain consumer behavior in the credit card market).

³⁸ See Drazen Prelec and George Loewenstein, *The Red and the Black: Mental Accounting of Savings and Debt*, 17 Marketing Sci 4, 23 (1998) (describing mental methods including prospective accounting and decoupling that allow consumers to enjoy acquisitions at the moment of purchase with a credit card without considering the unpleasantness of paying their credit card bill when it comes due).

³⁹ Gerald Dworkin, *Paternalism*, Edward N. Zalta, ed, *The Stanford Encyclopedia of Philosophy* (2006), online at <http://plato.stanford.edu/entries/paternalism> (visited Jan 17, 2006).

system likes and if the intervention creates more liking than the person would achieve on his own or with market-supplied help.⁴⁰

Ideally, such paternalism would be “libertarian” in the sense that it would only intervene in choice when *some* kind of regulation is necessary (for example, setting a default),⁴¹ and would be “asymmetric” in the sense that little harm would be done to people whose wanting and liking systems are well-synched.⁴²

Paternalism could take two mild forms, licensing and dramatizing. “Licensing” is checking whether wanting and liking are sufficiently synchronized to permit a person to make choices on his own (presumably in a particular domain). “Dramatizing” is using mechanisms to remind the wanting system, at the time of choice, about liking that may be overlooked or missynchronized with wanting.

A. Licensing

Driving tests provide a model for how licensing might occur, as a way to certify sufficient education of the wanting system. Because adolescents develop the hand-eye coordination and patience to drive safely at different ages, and learning to drive takes time and supervision, the current policy makes sense: prohibit driving by all children up to a certain age;⁴³ issue learner’s permits that allow driving only when a supervising adult is present (so the learners can learn); then force would-be drivers to pass written and physical driving tests, preferably with periodic retesting and other constraints (such as license suspension after accidents).

One can imagine a “credit-card permit” test that is conceptually similar to a driving test. Treat a credit card as a piece of equipment that

⁴⁰ A shadow is always cast on this type of analysis by the nagging question, “How can we be so sure that *they* won’t like it?” There are two central problems: a slippery slope, and capture. The slippery slope argument is that interfering with OCD patients might make it legally or politically easier for more interference in similar behaviors that are less scientifically grounded as disorders. The capture concern is that paternalistic policies will be guided or shaped (in their detail) by special interests who either profit from interference (for example, prohibiting price advertising for eyeglasses) or who want to “save others” (for example, religious fanatics). A full analysis of any policy should, of course, anticipate both slippery slope escalation and capture. It is too much to ask at this early stage, however, when details of even a single policy are speculative, to anticipate these problems and respond to them, at least in this short Essay.

⁴¹ See generally Richard H. Thaler and Cass R. Sunstein, *Libertarian Paternalism*, 93 *Am Econ Rev* 175 (2003); Cass R. Sunstein and Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 *U Chi L Rev* 1159 (2003).

⁴² See generally Colin Camerer, et al, *Regulation for Conservatives: Behavioral Economics and the Case for “Asymmetric Paternalism,”* 151 *U Pa L Rev* 1211 (2003).

⁴³ Some states have a “farm license” that allows very young children to drive in certain situations. For example, in North Dakota, fourteen-year-olds can drive farm vehicles within 150 miles of their farm. *License Classifications* (North Dakota Dept of Transportation 2005), online at <http://www.state.nd.us/dot/dlclass.html> (visited Jan 17, 2006).

a user should know how to operate. Require potential borrowers to take a quiz with some basic questions, like how rapidly interest compounds and how long it takes to pay off a \$1,000 balance if minimum payments are made and interest accumulates. Perhaps have them speak with a counselor or with experienced adults, some of whom used cards wisely and others who declared bankruptcy and regret it. The test could also use the opportunity to educate potential borrowers about basic borrowing rates and phenomena (much as driving tests remind people of traffic laws and some facts). The paternalism here is asymmetric because if borrowers know these facts, the test can be designed to be easy to pass.

B. Dramatizing

Another mild paternalistic intervention could be used in cases where one suspects the wanting system is not accounting for future liking or disliking correctly, by bringing future liking and disliking vividly into the present. This is often done in “scared straight” films and talk-show interventions that show juvenile delinquents what prison is like.⁴⁴ Alcoholics Anonymous (AA) (one of the more successful treatment programs) uses vivid testimonials by other AA members that remind sober alcoholics of all the crazy and destructive things that alcoholics do when they are drinking. Teenagers in high school programs can be made to carry around a baby who periodically simulates an infant’s typical needs (crying, changing, nursing), to dramatize the constraint that results from early parenthood.

A combination of nutritional modeling and computer morphing of a body image could be used to show a person, for example, what they would look like in one year if they continue their steady diet of fast food or, oppositely, if they stuck with their personal trainer three times a week for a year.

C. Calibration

A strong paternalistic intervention is to try to calibrate liking directly, and then require choosers to consider their liking ratings when making choices (weak form) or to force them to choose the goods they liked most (strong form). Of course, the success of this technique depends on how well liking is measured.

⁴⁴ Unfortunately, the available evidence suggests these programs don’t work very well. See Surgeon General, *Youth Violence: A Report of the Surgeon General* 95 (2001), online at <http://media.shs.net/ken/pdf/surgeon/SG.pdf> (visited Jan 17, 2006) (discussing results that show shock programs such as Scared Straight have equal or higher rearrest rates than exposure to no program at all).

The kind of thing I have in mind is a blind taste test—tasters rate objects, and then automatically choose the one they rated most highly. This could improve choice if there is a component of wanting, like a brand name, that is not a genuine component of liking. But of course, if taste is not the only component of liking, then true liking is mis-measured by this kind of calibration.

D. Delegating Choice

When there are wanting-liking gaps, it is possible that somebody else knows more about what you would really like than you do. Having the local wine connoisseur at dinner choose the wine for the table is a common and good practice. From our point of view, delegating choice is wise if the expert knows about typical patterns of mistakes—commonly wanted choices that are typically disliked later (“don’t fill up on bread; save room for dessert”). Allowing experts some say in a person’s choice is a form of paternalism; in the weak form, the chooser is forced to hear an expert opinion (for example, a marriage counselor’s report); in the strong form, the expert can overrule the choice or impose a choice.⁴⁵

Arranged marriages are an interesting example of various degrees of this sort of delegation. Ignoring externalities like family pride and business connections cemented by marriage, the argument for arranging a marriage is that members of a prospective couple know less about what kind of people they will truly like in the long run than a matchmaker (often a parent) does. The couple’s judgments might naturally overweigh observable short-run attributes (for example, attractiveness and sex appeal) and underweigh less observable long-run attributes (for example, fidelity or good parenting), while wiser and older matchmakers weight these attributes more accurately.

Indeed, a small part of this concept is already built into law: in all but three American states, minimum marriage ages are lower with parental consent than without consent. This gap in consent ages can be seen as a policy that permits marriage at age fifteen (in Utah, for example), but with a parental veto that can delay marriage until age eighteen.⁴⁶

E. Promoting Learning

In the face of status quo bias, a common type of wanting-liking mistake might be underexperimenting to find out what you really like most. If this is so, a paternalistic policy that subsidizes or forces experi-

⁴⁵ One argument about the influence of defaults is that they implicitly convey information about the best choices.

⁴⁶ Utah Code Ann § 30-1-9 (2005).

mentation could be justified, especially for momentous decisions with irreversible consequences—like choosing a profession, or a spouse.

Another example is hospital rotations, in which interns are exposed to many different types of medicine (presumably to help them learn what specialty appeals to them most).

The Amish tradition of “rumspringa” is a model for forced experimentation (although it serves a different purpose than I have in mind). At the age of sixteen, Amish youths are released from the strictures of the church and allowed to “run around” (which is the meaning of the Pennsylvania Dutch term). After some months or years, they can choose to be baptized and enter the church as an adult, but must then accept restrictions on behavior and dress; if they don’t, typically the disappointed Amish family cuts off all contact. (Around 90 percent return to the church.) The philosophy behind rumspringa is that paternalism is acceptable until sixteen, but after that point, a young adult should be entitled to make an informed decision about lifelong membership, and can only do so by exploring the “Devil’s Playground” (Amish slang for the decadent outside world—such as my neighborhood in Los Angeles, where the reality TV show “Amish in the City” was filmed).

A “minirumspringa” could include mandatory field trips to foreign countries for high school students, so they can see another part of the world, exposure to arts and sports they are unfamiliar with, and “international day” food smorgasbords. This kind of experimentation could be bundled into educational experiences in public schools or universities.

F. Can Markets Equate Wanting and Liking?

Of course, paternalism may be unnecessary if private market transactions promote learning, or bring wanting and liking together. So it is important to ask whether markets can equate wanting-liking gaps. The answer is likely to depend on details of the phenomenon and on some aspects of industry structure and technology.

A crucial feature is self-awareness: Do people *know* they want what they don’t like (or vice versa)? Put differently, do they have a metapreference for restraining wanting (do they want not to want)? In cases like addiction, individuals seem to vary on this dimension. So there is a hope for at least some people to contract privately for help (for example, voluntary drug rehab, or gastric bypass “stomach stapling” surgery for obesity).⁴⁷

⁴⁷ Stomach stapling is an interesting case. It is an extreme surgery in which the stomach is “stapled” down to the size of a shot glass, and part of the duodenum is removed to create a di-

In general, when consumers have limited rationality or willpower, there will be a tug-of-war in which firms compete to profit from “exploiting” them—for example, showing tempting hamburger commercials—and to profit from protecting them. What happens in equilibrium will take some careful thought and a lot of information about the nature of the wanting-liking gap, consumer self-awareness, and other details.

For example, firms could profitably exploit shopaholics by making it very easy to feel the pleasure of buying (for example, QVC home shopping network—“operators standing by!”), or by creating stringent return policies. Alternatively, firms could help shopaholics with other institutional fixes: rental markets for goods that are often little-used, like gym equipment; and low-cost trading institutions for getting rid of unused goods, like flea markets and eBay (and eBay consignment firms that sell on eBay for you).⁴⁸ The latter institutions enable economic bulimia—shopaholics can enjoy the pleasure of buying goods, while disgorging the goods later if they are unused.

IV. TOWARD A POSITIVE WANTING-LIKING THEORY OF PATERNALISM

The arguments above are a very rough sketch—science fiction, at this point—of how neuroscientific understanding of wanting and liking could eventually inform a reasoned approach to selective paternalism. But this framework might also have some value as a positive theory to explain paternalism that is widely accepted, and also explain differences in paternalism across history and cultures.

Below are some examples of where paternalistic restrictions are already in place. I know too little about any of these domains to do the proper scholarship, but the ideal positive analysis would work like so: ask whether regulations are justified in wanting-liking terms, and look

gestive shortcut. Postsurgery patients report having to eat small amounts regularly. A large majority lose a large amount of weight and keep it off. Some are such compulsive eaters that they eat too much and bust the staples. Interestingly, the large majority of patients are female. It is also notable that these \$25,000 surgeries are very profitable. So hospitals are eager to do the surgery, which shows how supply-side considerations are important in determining wanting-liking gaps in equilibrium. See Lisa Girion, *Hospitals Hunger for Losses*, *LA Times* A1 (June 7, 2005) (discussing the popularity, profitability, and potential risks of stomach stapling surgery). See also Atul Gawande, *Complications: A Surgeon's Notes on an Imperfect Science* 178–79 (Metropolitan 2002) (noting that gastric-bypass surgery is effective for more than 80 percent of patients).

⁴⁸ In the revealed-preference approach, generous return policies presumably indicate that returns are rare, or that returned goods are random forecasting mistakes (which is consistent with a view of rational choice in which people can't always guess perfectly what they'll like).

for historical or cross-sectional variations in wanting-liking gaps as justification for different regulations.

A. Minors

People under the age of eighteen are restricted from a great many activities. In the United States, different states have various ages at which minors can legally drive cars, buy cigarettes, drink alcohol, consent to sex, quit school, and get married. I'm not an expert on the history and basis for these restrictions, but presumably they are all at least loosely grounded in some presumption that wanting and liking systems in minors are not fully connected by learning (and in concerns for externalities, such as those caused by reckless teenage drivers).

As the brain develops, it is very unlikely that all minors develop wanting-liking synchrony at the same rate. Of course, there is obviously an advantage to having an easily enforced step function rule, like prohibiting alcohol use by people under twenty-one, even if it is understood that some teenagers are able to "handle it" (and some adults over twenty-one are not). Nonetheless, a litmus test for wanting-liking synchrony could conceivably be used as a substitute for uniform rules based on age.

B. Sexual Age of Consent

Sex is a challenge for paternalism. Because there is little externality (other than potential pregnancy, family honor, and so forth), consent age restrictions must have to do with either limiting competition or pure paternalism. Legal consent ages are remarkably similar worldwide, averaging sixteen years old.⁴⁹ In the United States, consent ages vary, and the minimum ages across the world range from twelve (Mexico, the Philippines, Panama, and Zimbabwe) to twenty (Tunisia).⁵⁰

C. Mental Incompetence

Tests of mental competence might be understood in this way. A typical criterion in psychiatric evaluation of involuntary commitment is whether a patient is a "danger to self and others." A danger to self can easily be construed as the capacity to choose (want) what is harmful (disliked).

⁴⁹ *Worldwide Ages of Consent* (2005), online at <http://www.avert.org/aofconsent.htm> (visited Jan 17, 2006).

⁵⁰ Few parts of the world have different ages for boys and girls (despite striking evidence of differences in teenage development, or perhaps because of that evidence). Bolivia, Hong Kong, and Iran have lower ages for girls, and Botswana has a lower age for boys. Id.

An interesting example, although not intended as paternalism *per se*, comes from a most unlikely source—Judge Richard Posner’s blog. In thinking about mandatory retirement, Posner wrote,

I wish to make a suggestion that would achieve the principal benefits of mandatory retirement without the principal costs. It is simply this: beginning at age 70, require every life-tenured professor and every life-tenured judge to take a test of mental acuity every five years. (I use these simply as examples of “light” jobs from which the occupant is unlikely to be forced to retire by the demands that the job places on him.) The test results would be available to the members of the professor’s department or the judge’s court but to no others. The results would not be a basis for a determination of incapacity; they would not even be admissible in a competence hearing. The expectation rather is that a poor test result would persuade the individual, perhaps by persuading his colleagues who would in turn persuade him, or persuade members of his family to persuade him, to retire voluntarily.⁵¹

As he notes, Posner is not eager to use such tests paternalistically to restrict a person’s employment opportunities, but one could imagine doing so in extreme cases or in conjunction with other measures of workplace competence.

D. Cooling Off

“Cooling off” laws allow consumers to breach contracts for purchases within some period (typically short, such as three days).⁵² Such laws are easily understood as an expression of the belief that consumers in a “hot” emotional state may want something, but later realize they don’t need it, can’t afford it, or won’t like it. (Such laws might also help in resolving household disputes, but the law typically makes no restrictions on married couples.)

E. Sin Taxes

“Sin taxes” are designed to internalize externalities, and possibly “internalities” from consumption of sinful goods like alcohol and cigarettes. Obviously, an ideal policy should be grounded in a specific

⁵¹ Richard Posner, *Refusing to Retire: What Can Be Done When People Overstay Their Welcome*, The Becker-Posner Blog (June 5, 2005), online at http://www.becker-posner-blog.com/archives/2005/06/refusing_to_ret.html (visited Jan 17, 2006).

⁵² See, for example, Camerer, et al, 151 U Pa L Rev at 1238–47 (cited in note 42) (discussing “cooling off” in regards to consumer protection, family law, settlement agreements, and other potential applications).

model of how much taxes will reduce consumption, and how large externalities and internalities are. Gruber and Köszegi's⁵³ approach is a good modern example that computes an optimal sin tax from a particular model. It is easy to quibble with the details of their work or the (large) sin tax they deduce. The point, however, is that sin taxes do exist. In the wanting-liking framework, they are a blunt instrument that substitutes a wanting-system forecast of future-disliking with a current budget constraint that presumably enters into wanting.

CONCLUSION

The idea in this highly speculative Essay is simple: there is evidence from mice with gene “knockdowns” that wanting (choice) and liking (hedonic reaction) are dissociated. There are many types of evidence with humans too—like addicts who report a compulsion to use drugs, but say they get no pleasure (admittedly, all this evidence relies on direct measurement of liking).

This type of evidence invites a wanting-learning-liking interpretation of choice and utility. Wanting guides choice, and liking is the same as the hedonic (Benthamite) concept of utility. Learning is a process by which the wanting system comes to know what the liking system likes.

A historical discursion on the development of revealed-preference theory⁵⁴ is a reminder that equating unobserved utility with observed choice—assuming wanting forecasts liking correctly—was a conventional simplification, not the result of any proof or data. At the same time, Pareto, who advocated such a simplification, readily admitted that assuming that choices match the “subjective fact” of utility is justified by restricting attention to repeated actions, where learning has had a chance to work. (His view is echoed by recent advocates of a “discovered preference hypothesis.”) By emphasizing repetition and learning as the conditions under which wanting and liking coincide, Pareto is actually a surprising progenitor for the ideas in this Essay.

Distinguishing wanting, liking, and learning does two things: It generalizes revealed-preference theory, because conventional revealed-preference is the special case where learning has taught wanting what is liked. And it provides a language for talking about paternalism. Government paternalism is justified if there is a verifiable gap between wanting and liking, markets do not close the gap, and no other harms are created.

Here are a few very rough ways paternalism might work to equate wanting and liking:

⁵³ See Gruber and Köszegi, 116 Q J Econ at 1285–93 (cited in note 35).

⁵⁴ See Bruni and Sugden, *The Road Not Taken* 1 (cited in note 2).

- Licensing—checking that a person has informed beliefs that influence expected liking (for example, a “credit card driver’s test”);
- Dramatizing—making future components of liking vivid;
- Calibrating—seeing if a person’s wanting and liking are matched, and if not, forcing choice based on measured liking;
- Delegating choice—allowing an expert who can forecast liking to make or constrain a choice (for example, matchmaking); and
- Promoting learning—preventing nonwanting of goods that people would actually like by forcing experimentation (à la Amish rumspringa).

Finally, paternalistic policies are already in place regulating the minimum age of marriage and sexual consent, judgment of mental competence, cooling off provisions in sales, and prohibition or sin taxation of drugs and alcohol. The wanting-liking-learning framework could be used positively to understand how these policies came about, and why they vary historically and across jurisdictions.