Best of LessWrong: April 2012

1. Attention control is critical for changing/increasing/altering motivation
2. Be Happier
3. How can we get more and better LW contrarians?
4. How accurate is the quantum physics sequence?
5. Crowdsourcing the availability heuristic
6. SotW: Be Specific
7. LessWrong downtime 2012-03-26, and site speed
8. Cryonics without freezers: resurrection possibilities in a Big World
9. Hofstadter's Superrationality
10. Why I Moved from AI to Neuroscience, or: Uploading Worms
11. Let's create a market for cryonics
12. Peter Singer and Tyler Cowen transcript
13. The Quick Bayes Table
14. A question about Eliezer

Best of LessWrong: April 2012

1. Attention control is critical for changing/increasing/altering motivation
2. Be Happier
3. How can we get more and better LW contrarians?
4. How accurate is the quantum physics sequence?
5. Crowdsourcing the availability heuristic
6. SotW: Be Specific
7. LessWrong downtime 2012-03-26, and site speed
8. Cryonics without freezers: resurrection possibilities in a Big World
9. Hofstadter's Superrationality
10. Why I Moved from AI to Neuroscience, or: Uploading Worms
11. Let's create a market for cryonics
12. Peter Singer and Tyler Cowen transcript
13. The Quick Bayes Table
14. A question about Eliezer

Attention control is critical for changing/increasing/altering motivation

I've just been reading Luke's "Crash Course in the Neuroscience of Human Motivation." It is a useful text, although there are a few technical errors and a few bits of outdated information (see [1], updated information about one particular quibble in [2] and [3]).

There is one significant missing piece, however, which is of critical importance for our subject matter here on LW: the effect of attention on plasticity, including the plasticity of motivation. Since I don't see any other texts addressing it directly (certainly not from a neuroscientific perspective), let's cover the main idea here.

Summary for impatient readers: focus of attention physically determines which synapses in your brain get stronger, and which areas of your cortex physically grow in size. The implications of this provide direct guidance for alteration of behaviors and motivational patterns. This *is* used for this purpose extensively: for instance, many benefits of the Cognitive-Behavioral Therapy approach rely on this mechanism.

I – Attention and plasticity

To illustrate this properly, we need to define two terms. I'm guessing these are very familiar to most readers here, but let's cover them briefly just in case.

First thing to keep in mind is *the plasticity of cortical maps.* In essence, particular functional areas of our brain can expand or shrink based on how often (and how intensely) they are used. A small amount of this growth is physical, as new axons grow, expanding the white matter; most of it happens by repurposing any less-used circuitry in the vicinity of the active area. For example, our sense of sight is processed by our visual cortex, which turns signals from our eyes into lines, shapes, colors and movement. In blind people, however, this part of the brain becomes invaded by other senses, and begins to process sensations like touch and hearing, such that they become significantly more sensitive than in sighted people. Similarly, in deaf people, auditory cortex (part of the brain that processes sounds) becomes adapted to process visual information and gather language clues by sight.

Second concept we'll need is *somatosensory cortex* (SSC for short). This is an area of the (vertebrate) brain where most of the incoming touch and positional (proprioceptive) sensations from the body converge. There is a map-like quality to this part of our brain, as every body part links to a particular bit of the SSC surface (which can be illustrated with silly-looking things, such as the sensory homunculus). More touch-sensitive areas of the body have larger corresponding areas within the SSC.

With these two in mind, let's consider one actual experiment [4]. Scientists measured and mapped the area of an owl monkey's SSC which became activated when one of his fingertips was touched. The monkey was then trained to hold that finger on a tactile stimulator – a moving wheel that stimulates touch receptors. The monkey had to pay attention to the stimulus, and was rewarded for letting go upon detecting

certain changes in spinning frequency. After a few weeks of training, the area was measured again.

As you probably expected, the area had grown larger. The touch-processing neurons grew out, co-opting surrounding circuitry in order to achieve better and faster processing of the stimulus that produced the reward. Which is, so far, just another way of showing plasticity of cortical maps.

But then, there is something else. The SSC area expanded only when the monkey had to *pay attention* to the sensation of touch in order to receive the reward. If a monkey was trained to keep a hand on the wheel that moved just the same, but he did not have to pay attention to it… the cortical map remained the same size. This finding has since been replicated in humans, many times (for instance [5, 6]).

Take a moment to consider what this means.

A man is sitting in his living room, in front of a chessboard. Classical music plays in the background. The man is focused, thinking about the next move, about his chess strategy, and about the future possibilities of the game. His neural networks are optimizing, making him a better chess player.

A man is sitting in his living room, in front of a chessboard. Classical music plays in the background. The man is focused, thinking about the music he hears, listening to the chords and anticipating the sounds still to come. His neural networks are optimizing, making him better at understanding music and hearing subtleties within a melody.

A man is sitting in his living room, in front of a chessboard. Classical music plays in the background. The man is focused, gritting his teeth as another flash of pain comes from his bad back. His neural networks are optimizing, making the pain more intense, easier to feel, harder to ignore.

II – Practical implications: making and breaking habits, efficacy of CBT

Habitual learned behaviors are often illustrated with the example of driving. When we are learning to drive, we have to pay attention to everything: when to push the pedals, when to signal, where to hold our hands… A few years later, these behaviors become so automatic, we hardly pay attention at all. Indeed, most of us can drive for hours while carrying on conversations or listening to audiobooks. We are completely unaware, as our own body keeps pushing pedals, signaling turns, and changing gears.

We can therefore say that *driving behaviors*, through practice and attention, eventually become automatic – which is, most of the time, a good thing. But so do many other things, including some destructive ones we might want to get rid of. Let’s take a simple one: nail biting. You are reading, or watching a movie, or thinking, or driving… when you suddenly notice some minor pain, and realize that you have chewed your nail into a ragged stump. Ouch!

You catch yourself biting, you stop. Five minutes later, you catch yourself biting again. You stop again. Repeat *ad infinitum*, or *ad nauseam*, whichever comes first.

Cognitive-Behavioral Therapy has a highly successful approach for breaking habits, which requires only a very subtle alteration to this process. You notice that you are biting your nails. You immediately focus your attention on what you are doing, and

you stop doing it. No rage, no blaming yourself, no negative emotions. You just stop, and you focus all the attention you can on the act of stopping. You move your arm down, focusing your attention on the act of movement, on the feeling of your arm going down, away from your mouth. That's it. You can go back to whatever you were doing.

Five minutes later, you notice yourself biting your nails again. You calmly repeat the procedure again.

By doing this, you are training yourself to perform a new behavior – the "stop and put the hand down" behavior – which is itself triggered by the nail-biting behavior. As you go along, you will get better and better at noticing that you have started to bite your nails. You will also get better and better at stopping and putting your hand down. After a while, this will become semi-automatic; you'll notice that your hand went to your mouth, a nail touched your tooth, and the hand went back down before you could do anything. Don't stop training: focus your attention on the "stop and drop" part of the action.

After a while, the nail-biting simply goes away. Of course, the more complex and more ingrained a habit is, the more effort and time will be needed to break it. But for most people, even strong habits can be relatively quickly weakened, or redirected into less destructive behaviors.

It's probably obvious that habits can be *created* in this way as well. We don't become better at things we do – we become better at things we *pay attention to while we're doing them.* If you want to make exercise a habit, your efforts will be much more effective if you focus your attention on your exercise technique, rather than repeatedly thinking how painful and tiring the whole process is.

There is also a direct implication for training in any complex skill. Start with the well-known learning curve effect: we gain a lot of skill relatively quickly, and then improvements slow down incrementally as we approach our maximum potential skill level. It is relatively easy to go from a poor to a mediocre tennis player; it is much, much harder to go from mediocre to good, and even harder to go from good to excellent.

Complex skills have many different aspects, which we usually attempt to train simultaneously. We can become very good at some, while staying poor at others. The optimal approach would be to focus most of our attention on those aspects where our abilities are weakest, since smaller investments of time and effort will lead to larger improvements in skill.

To keep with the tennis metaphor, one could become very good at controlling the ball direction and spin, while still having a poor awareness of the opponent's position. Simply playing more will improve both aspects further, but our hypothetical player should optimally try to focus her attention on opponent awareness [7].

Finally, there is another implication which I'll leave as an exercise for the readers. Mindfulness meditation, which essentially boils down to training control of attention, has been shown to exert a positive effect on many, many different things (lowering depression, anxiety and stress, as well as improving productivity [8, 9, 10]). In the light of the previous text, one obvious reason why better control over attention can produce all these beneficial effects should immediately come to mind.

References

[1] I have several quibbles, but let's stick to one (to prevent this note from becoming longer than the above text). Luke presents a view of dopamine reward system which is stuck in the early 2000's – ages ago by the pace of neuroscientific research. Dopamine actually has a very, very complex effect on motivation, and is able to strengthen or weaken single synaptic connections based on timing of the signal relative to the signals from the sensory systems. Endocannabinoid neurotransmission (i.e. signaling through chemicals that stimulate the same receptors that are affected by active ingredients in marijuana) is being shown as more and more important in this system as well, and the relative timing of the two signals appears critical.

The complexity of the effects increases by several orders of magnitude when networks are concerned. Consider this: a planning-related network in the prefrontal cortex can influence the motivation-generating networks in the striatum. A stimulus from the outside is perceived by the sensory networks and transmitted to the dopamine system, to the prefrontal cortex, and to the striatum. The same dopamine signal can, depending on exact timing of action potential bursts, strengthen synapses in the striatum, while weakening synapses in the prefrontal cortex. The result? The link between the stimulus and the actual motivation can increase or decrease, depending on exact connectivity between networks, on the relative sensitivity and on the exact topology of the meta-network in question.

See the following two references for a broad overview of the subject area.

[2] Calabresi P, Picconi B, Tozzi A, Di Filippo M. "Dopamine-mediated regulation of corticostriatal synaptic plasticity" Trends Neurosci. 2007 30(5):211-9.

[3] Wickens JR. "Synaptic plasticity in the basal ganglia" Behav Brain Res. 2009 199(1):119-28.

[4] Recanzone GH, Merzenich MM, Jenkins WM, Grajski KA, Dinse HR. "Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency-discrimination task" *J Neurophysiol.* 1992 67(5), 1031-56.

[5] Heron J, Roach NW, Whitaker D, Hanson JV. "Attention regulates the plasticity of multisensory timing" *Eur J Neurosci.* 2010 31(10), 1755-62.

[6] Stefan K, Wycislo M, Classen J. "Modulation of associative human motor cortical plasticity by attention" *J Neurophysiol.* 2004 92(1), 66-72.

[7] I'm not finding good papers directed exactly on this point, so I'll just throw this out as a personal opinion (although I'll say it appears well supported by indirect research). We all like to appear competent and skillful, especially in those areas where we have invested a lot of time and effort. This can lead to a bias where we focus on using those aspects of complex skills we are best at, and training those aspects most intensely. In other words, a tendency appears to exist to do exactly the opposite of what we should be doing. (If anyone has encountered a name for this bias, or has references to suggest, I would be very grateful to hear from you.)

[8] Brown KW, Ryan RM. "The benefits of being present: mindfulness and its role in psychological well-being" J Pers Soc Psychol. 2003 84(4):822-48.

[9] Davidson RJ, Kabat-Zinn J, Schumacher J, Rosenkranz M, Muller D, Santorelli SF, Urbanowski F, Harrington A, Bonus K, Sheridan JF. "Alterations in brain and immune function produced by mindfulness meditation" Psychosom Med. 2003 65(4):564-70.

[10] Shao RP, Skarlicki DP. "The role of mindfulness in predicting individual performance" Canadian J of Behavioral Sci 2009 41(4): 195–201.

Be Happier

This started as an assignment to find out about the science of ‘buying happiness’ (using money to become happier) — hence the emphasis on money-and-happiness. I learned a great deal more than how to buy happiness, however, and the project became somewhat more generalized. It is not meant to be comprehensive, but perhaps it makes for a useful supplement to Luke’s How to be Happy. This post consists mostly of quoted material.

In A Nutshell

Money and Happiness

- Spend on others, especially people you are close to. Positive feedback loop: Prosocial spending makes you happier, and happiness makes you more likely to spend prosocially.
- Don’t be stingy. It's bad for your health.
- Don’t think too much about money. It will impair your savoring ability. It's also bad for your family life.
- Be time-aware, but don’t think of time in terms of money.
- Being richer will not necessarily make you happier.
- Do not live in wealthy enclaves.
- Avoid conspicuous consumption.

Work Satisfaction

- Coping with Stress: React pragmatically rather than emotionally.
- Go for ‘approach’ goals instead of ‘avoid’ goals.
- Autonomy: Make a point of prefering autonomous goals rather than heteronomous goals (goals imposed by others).
- Autonomy: Make sure you have spare discretionary time — even at financial cost.
- Be passionate, but don’t obsess.
- Do work that you enjoy doing. Flow.
- Set goals that are reasonably challenging and reasonably achievable.

Materialism and Purchasing

- Prefer experiential purchases; avoid materialistic goals.
- Keep your goals intrinsic.
- Don’t do ‘comparison shopping.’ And don’t place much stock in the happiness potential of any one positive change.
- Follow the herd. “The best way to predict how much we will enjoy an experience is to see how much someone else enjoyed it.”

Interpersonal

- Socialize with close others.
- Associate with happy people.
- Give the people around you opportunities to be generous. Ask them for favors.
- Be actively kind (and occasionaly reminisce about your recent acts of kindness).

Stretching Happiness (fighting hedonic adaptation)

- Go for smaller, more frequent successes rather than larger ones.
- Go for variety and surprise. Don't keep doing the same thing.
- Savor anticipation. Delay consumption. Actively anticipate good experiences.
- Divide positive experiences into smaller pleasures, if possible.
- Corollary: Conclude negative experiences as soon as possible.
- Make a point of avoiding experiences that make you feel bad.

Appreciation

- Be grateful and count your blessings (literally). Recycle happiness by reminiscing about good experiences.
- Think of counterfactuals. ("If I didn't have this positive thing, what do I lose?")
- Breathe deeply. Expand your time — by slowing down.
- Stay in the present.

Optimal Happification

- Actively *want* to be happier. Motivation and investment matter.
- Learn about the science of happiness. Internalize the lessons in this article and in here.

Some Key Terms

- Subjective Well Being (SWB) aka happiness.
- Hedonic Adaptation — the phenomenon of (rapidly) diminishing positive or negative affect from any one experience or thing.
- Hedonic treadmill — the phenomenon of neverending aspirations for materialistic acquisitions that results from hedonic adaptation.

Money and Happiness

Spend on others, especially people you are close to.

> Past research in our lab has repeatedly shown that people are happier when they use financial resources to benefit others rather than themselves [Aknin, Dunn, Sandstrom & Norton, submitted, 1,14].
>
> ... findings suggest that to reap the greatest emotional reward from spending on someone else, one should direct their purchases to close others

> These findings should not be taken to suggest that people should avoid spending on weak social ties. Indeed, treating an acquaintance from yoga to a coffee after class might help to build a new strong tie. Thus, spending money on a weak social tie might help facilitate the development of new strong ties in the longer term.
>
> … research on reciprocal altruism and the evolution of cooperation demonstrates that people ultimately benefit from behaving generously and cooperatively toward individuals with whom they are likely to interact in the future
>
> The current results […] shed novel insight into translating spending choices into happiness: the next time you find a few spare dollars in your pocket, you will be happiest if you treat your best friend.

(Aknin, Sandstrom, Dunn, & Norton, 2011b)

> This research also supports the broaden-and-build theory of positive emotions by demonstrating that higher levels of happiness may expand an individual's mindset to include thoughts of others.

(Ahuvia, 2002)

> … prosocial spending may be particularly promising route to prosocial behavior because it has been shown to increase happiness immediately after spending (Dunn et al. 2008, 2010) and later upon reflection, as demonstrated here.
>
> ... how money is spent may matter more than how much money is spent. That is, participants who recalled spending on others felt happier than those who spent money on themselves, and the benefits of prosocial spending were the same regardless of whether they spent $100 or just $20. Recent work suggests that prosocial behavior leads to emotional gains by providing opportunities for positive social contact (Aknin et al. 2011b); therefore, prosocial spending should promote happiness if the spending opportunity fosters positive relations with others, which may be largely independent of the specific amount of money spent. (Aknin, Dunn, & Norton, 2011a)
>
> ... participants assigned to spend a small windfall on someone else by purchasing a gift or making a donation to charity (prosocial spending) were significantly happier at the end of the day than participants assigned to spend the same size windfall by paying for a bill, expense, or gift for themselves (personal spending)

(Aknin, Dunn, & Norton, 2011a)

Positive feedback loop:

> Taken together, our results show that
> (a) recalling a past prosocial spending experience leads to higher levels of happiness,
> (b) higher levels of happiness increase the likelihood of engaging in prosocial spending, and
> (c) recalling a past experience of prosocial spending increases the likelihood of spending a new windfall on others to the extent that happiness levels are elevated in the interim. This suggests that spending money on others may be self-reinforcing as long as this prosocial experience provides happiness.

(Aknin, Dunn, & Norton, 2011a)

Being generous will make you happier.

> experiments within two very different countries (Canada and Uganda) […] show that spending money on others has a consistent, causal impact on happiness.
>
> In contrast to traditional economic thought—which places self-interest as the guiding principle of human motivation—our findings suggest that the reward experienced from helping others may be deeply ingrained in human nature, emerging in diverse cultural and economic contexts.
>
> [In both] Canada and Uganda — [which] differ dramatically in national-level income and donation frequency, we find that individuals report significantly greater well-being after reflecting on a time when they spent money on others rather than themselves. This effect emerged consistently across these two cultures, even though the specific prosocial spending experiences participants described differed considerably. Thus, although prosocial spending differs in both frequency (Study 1) and form (Study 2) in poor versus wealthy countries, its emotional consequences are remarkably consistent.

(Aknin et al., 2010)

> we found that spending more of one's income on others predicted greater happiness both cross-sectionally (in a nationally representative survey study) and longitudinally (in a field study of windfall spending). Finally, participants who were randomly assigned to spend money on others experienced greater happiness than those assigned to spend money on themselves.

(Dunn, Aknin, & Norton, 2008)

> …prosocial spending is consistently associated with greater happiness.
> The robustness of this mechanism is supported by our finding that people seem to experience emotional benefits from sharing their financial resources with others not only in countries where such resources are plentiful, but also in impoverished countries where scarcity might seem to limit the possibilities to reap the gains from giving to others.

(Aknin et al., 2010)

Don't be Stingy.

Aside from the positive effect of generosity on your own happiness, stinginess makes you less healthy; it is easier to be happy when you are healthy.

> The present research suggests that stingy economic behavior can produce a feeling of shame, which in turn drives secretion of the stress hormone cortisol.
>
> … the present research provides support for Social Self-Preservation Theory, which posits that acute threats to the 'social self' induce shame and lead to increased cortisol, as part of a coordinated response to social threats (Dickerson, Gruenewald, & Kemeny, 2004a).
>
> Our findings provide initial, suggestive evidence that shame and cortisol represent plausible emotional and biological pathways that might link everyday decisions

> about whether to help others with downstream consequences for one's own health.
>
> ... stingy economic behavior predicts cortisol secretion only to the extent that stinginess provokes shame.

(Dunn et al., 2010)

Caveat: hedonic adaptation moderates the deleterious effect of bad health on well-being, but not entirely — and negative experiences are more powerful than positive experiences:

> "To sum up almost two decades of research, bad is stronger than good."

(Chancellor & Lyubomirsky, 2011)

(More on negative experiences farther down in 'STRETCHING HAPPINESS'.)

Think about time, but don't think of time in terms of money ("An hour of my time is worth…").

> ... thinking about time in terms of money can influence how people experience pleasurable events by instigating greater impatience during unpaid time.
> In three separate experiments we have demonstrated that bringing individuals' effective hourly wage to their attention impairs the ability to derive happiness from pleasurable experiences.
> One possible explanation is that impatience discourages savoring. Savoring is a form of emotional regulation which augments the happiness individuals derive from experiences (e.g. Bryant et al., 2005; Quoidbach, 2009; Tugade & Fredrickson, 2007).
> ... recent ethnographic research […] found that people who are paid by the hour narrowly evaluate their time use in terms of its economic returns. As a consequence, they tend to discount the worth of activities with non-economic benefits (Evans et al., 2004).
> ... the present findings suggest that thinking about time in terms of money is poised to affect our ability to smell the proverbial roses.

(DeVoe & House, 2012)

> ... there is a bi-directional relationship between the scarcity of time and its value: not only does having little time make it feel more valuable, but when time is more valuable, it is perceived as more scarce (DeVoe & Pfeffer, 2010).

(Aaker et al., 2010)

In an word priming experiment done in a cafe:

> Pair-wise comparisons showed that individuals primed with time spent more of their time at the café socializing than those primed with money. Further, individuals primed with time spent less of their time working than those primed with money
>
> Participants primed with money worked more than those in the control condition and participants primed with time worked less than those in the control condition

> … participants primed with time were happier than those primed with money. […] Participants primed with time were also happier than those in the control condition [and] the happiness levels of those primed with money and those in the control condition did not differ significantly,
>
> These results suggest that increasing the relative salience of time (vs. money) can increase happiness by leading people to behave in more connecting ways [and] can nudge someone to spend that extra hour at home rather than at the office, there finding greater happiness.
>
> Focusing on money motivates one to work more, which is useful to know when struggling to put in that extra hour of work to meet a looming deadline. However, passing the hours working (although productive) does not translate into greater happiness. Spending time with loved ones does, and a shift in attention toward time proves an effective means to motivate this social connection.
>
> … the relevant question may be not how much money people have, but rather how much attention people put on money.
>
> Despite the belief that money is the resource most central to Americans' pursuit of happiness, increased happiness requires a shift in attention toward time.

(Mogilner, 2010)

Being richer will not necessarily make you happier.

> The belief that high income is associated with good mood is widespread but mostly illusory. People with above-average income are relatively satisfied with their lives but are barely happier than others in moment-to-moment experience, tend to be more tense, and do not spend more time in particularly enjoyable activities. Moreover, the effect of income on life satisfaction seems to be transient. We argue that people exaggerate the contribution of income to happiness because they focus, in part, on conventional achievements when evaluating their life or the lives of others.
> The latter finding might help explain why income is more highly correlated with general life satisfaction than with experienced happiness, as tension and stress may accompany goal attainment, which in turn contributes to judgments of life satisfaction more than it does to experienced happiness.
> Despite the weak relation between income and global life satisfaction or experienced happiness, many people are highly motivated to increase their income. In some cases, this focusing illusion may lead to a misallocation of time, from accepting lengthy commutes (which are among the worst moments of the day) to sacrificing time spent socializing (which are among the best moments of the day) (28, 29). An emphasis on the role of attention helps to explain both why many people seek high income—because their predictions exaggerate the increase in happiness due to the focusing illusion—and why the long- term effect of income gains become relatively small, because attention eventually shifts to less novel aspects of daily life.

(Kahneman, 2006)

> It is found that higher income aspirations reduce people's satisfaction with life. In Switzerland and the New German Laender, the negative effect of an increase in the aspiration level on well-being is of a similar absolute magnitude as the

positive effect on well-being of an equal increase in income. This suggests that subjective well-being depends largely on the gap between income aspirations and actual income and not on the income level as such. the higher the ratio between aspired income and actual income, the less satisfied people are with their life, ceteris paribus. This supports the notion of a relative utility concept.

(Stutzer & Frey, 2010)

Emotional well-being also rises with log income, but there is no further progress beyond an annual income of ~$75,000. Low income exacerbates the emotional pain associated with such misfortunes as divorce, ill health, and being alone.

[The data suggest that] above a certain level of stable income, individuals' emotional well-being is constrained by other factors in their temperament and life circumstances.

(Kahneman & Deaton, 2010)

Pitfall of being wealthy: your ability to savor positive emotions and experiences will be impaired. Don't make money your priority.

The present study provides the first evidence that money impairs people's ability to savor everyday positive emotions and experiences.

In a sample of working adults, wealthier individuals reported lower savoring ability.

... the negative impact of wealth on savoring undermined the positive effects of money on happiness.

... moving beyond self-report, participants exposed to a reminder of wealth spent less time savoring a piece of chocolate and exhibited reduced enjoyment of it. The present research supplies evidence for the previously untested notion that having access to the best things in life may actually undercut the ability to reap enjoyment from life's small pleasures.

In other words, one need not actually visit the pyramids of Egypt or spend a week in the legendary spas of Banff—simply knowing that these peak experiences are readily available may increase the tendency to take the small pleasures of daily life for granted.

... having access to the best things in life may actually undermine the ability to reap enjoyment from life's small pleasures.

... our research demonstrates that a simple reminder of wealth produces the same deleterious effects as actual wealth, suggesting that perceived access to pleasurable experiences may be sufficient to impair everyday savoring. (Quoidbach et al., 2010)

This perspective is consistent with the intriguing theoretical notion that hedonic adaptation may occur not only in response to past experiences, but also in response to anticipated future experiences (Frederick & Loewenstein, 1999).

Our studies provide a novel contribution by demonstrating that the emotional benefits that money gives with one hand (i.e., access to pleasurable experiences),

> it takes away with the other by undercutting the ability to relish the small delights of daily living.
>
> … experimentally exposing participants to a reminder of wealth produced the same deleterious effect on savoring as did actual individual differences in wealth.

(Quoidbach et al., 2010)

> Across nations, placing a higher importance on money is associated with lower well-being (Kirkcaldy, Furnham, & Martin, 1998).

(Diener & Seligman, 2004)

Financial aspirations are bad for family life (and the quality of interpersonal relationships is a strong predictor of happiness).

> The negative consequences [of financial aspirations] were particularly severe for the domain of family life; the stronger the goal for financial success, the lower the satisfaction with family life, regardless of household income.

(Nickerson et al., 2003)

Don’t live ‘high’.

> Not only materialism, but wealth itself has been found in a few studies to produce negative effects. Hagerty (2000) found that when personal income was statistically controlled, individuals living in higher-income areas in the United States were lower in happiness than people living in lower-income areas. (Diener & Seligman, 2004)
>
> This suggests that wealthy individuals are fortunate if they live in middle-class areas rather than in wealthy enclaves.
>
> The negative effects of wealthy communities might partly be explained by their higher materialism (Stutzer, in press).

(Diener & Seligman, 2004)

(Perhaps the more important point here is that you must surround yourself with low-materialism people, which means surrounding yourself with happy people, since materialism correlates negatively with happiness. A caveat to the advice of living in middle class areas if wealthy: the presence of a wealthy neighbor can make people more materialistic; it makes them aspire for more. A wealthy person can make his less wealthy neighbors less happy. See below.)

Avoid conspicuous consumption.

> The ‘relative income hypothesis’ was formulated and econometrically tested by James Duesenberry (1949), who posited an asymmetric structure of externalities. People look upwards when making comparisons. Aspirations thus tend to be above the level already reached. Wealthier people impose a negative external effect on poorer people, but not vice versa. Fred Hirsch (1976), in his book Social Limits to Growth, emphasised the role of relative social status by calling attention to ‘positional goods’ which, by definition, cannot be augmented, because they rely

> solely on not being available to others. This theme was taken up by Robert Frank (1985, 1999), who argued that the production of positional goods in the form of luxuries, such as exceedingly expensive watches or yachts, is a waste of productive resources, as overall happiness is thereby decreased rather than increased.

(Frey & Stutzer, 2002)

(This relates to the recommendation to associate with happy people — farther down.)

More Recommendations

WORK SATISFACTION

Coping with Stress: React pragmatically rather than emotionally.

> Coping can be divided into two broad engagements – either to trigger the individual to approach the problem or to regulate the emotional reactions arising from the challenge at hand (Andersson & Willebrand 2003). The literature typically differentiates two broad strategies of coping (for a review, see Lazarus & Folkman 1984). First, problem-based coping refers to a cognitively-based response behaviour that includes efforts to alleviate stressful circumstances. This coping strategy includes defining the problem, generating alternative solutions, determining the costs and benefits of such solutions, and actions taken to solve the problem. Second, emotion-based coping involves behavioural responses to regulate the affective consequences of stressful events, which may include avoidance, minimisation and distancing oneself from the problem (Lazarus & Folkman 1984).
>
> It seems that problem-based coping strategies are more instrumental than emotion-based ones for attaining successful entrepreneurial outcomes. This implies that entrepreneurs who are more inclined toward emotion-based coping could be trained to employ more problem-based coping, since coping can be learned just like any other competence.

(Drnovšek et al., 2010)

Leaders and Entrepreneurs: Don’t take on too many business partners. (See also AUTONOMY below)

> … entrepreneurs who had lower perceived role centrality and were part of a larger founding team were more inclined to use emotion-based coping than those who started their venture in smaller teams. We believe these insights can help in training entrepreneurs in the development of effective coping strategies. Individuals with perceived high centrality of their entrepreneurial role are more likely to effectively engage in coping to optimise their venture”s performance and their own psychological well being.

(Drnovšek et al., 2010)

Prefer the ‘approach’ path instead of the ‘avoid’ path.

It is good for your well-being to work towards *achieving* something, rather than *preventing* something from happening.

> [One] concern is whether one’s goal activities are characterized by approach or avoidance motivational systems. Elliot & Sheldon (1997), for example, classified goals as approach or avoidance and then examined the effects of goal progress over a short-term period. Pursuit of avoidance goals was associated with both poorer goal progress and with lower well-being. Elliot et al (1997) similarly showed that people whose personal goals contained a higher proportion of avoidance had lower SWB [Subjective Well Being]. They also demonstrated the association between neuroticism and avoidance goals, but showed that the impact of avoidance regulation was evident even when controlling for neuroticism. Carver & Scheier (1999) also presented research linking approach goals (positively) and avoidance goals (negatively) to well-being outcomes.

(Ryan & Deci, 2001)

AUTONOMY: Make a point of prefering autonomous goals rather than heteronomous goals (goals imposed/expected by others).

> Another actively researched issue concerns how autonomous one is in pursuing goals. SDT in particular has taken a strong stand on this by proposing that only self-endorsed goals will enhance well-being, so pursuit of heteronomous goals, even when done efficaciously, will not. The relative autonomy of personal goals has, accordingly, been shown repeatedly to be predictive of well-being outcomes controlling for goal efficacy at both between-person and within-person levels of analysis (Ryan & Deci 2000). Interestingly this pattern of findings has been supported in cross-cultural research, suggesting that the relative autonomy of one’s pursuits matters whether one is collectivistic or individualistic, male or female (e.g. V Chirkov & RM Ryan 2001; Hayamizu 1997, Vallerand 1997).
>
> Sheldon & Elliot (1999) developed a self-concordance model of how autonomy relates to well-being. Self-concordant goals are those that fulfill basic needs and are aligned with one’s true self. These goals are well-internalized and therefore autonomous, and they emanate from intrinsic or identified motivations. Goals that are not self-concordant encompass external or introjected motivation, and are either unrelated or indirectly related to need fulfillment. Sheldon & Elliot found that, although goal attainment in itself was associated with greater well-being, this effect was significantly weaker when the attained goals were not self-concordant. People who attained more self-concordant goals had more need-satisfying experiences, and this greater need satisfaction was predictive of greater SWB. Similarly, Sheldon & Kasser (1998) studied progress toward goals in a longitudinal design, finding that goal progress was associated with enhanced SWB and lower symptoms of depression. However, the impact of goal progress was again moderated by goal concordance. Goals that were poorly integrated to the self, whose focus was not related to basic psychological needs, conveyed less SWB benefits, even when achieved.

(Ryan & Deci, 2001)

> … freely chosen activities increase happiness, while obligatory activities lower it (Csikszentmihalyi & Hunter, 2003).

(Aaker et al., 2010)

> ... we find additional evidence that entrepreneurs also derive utility from things other than financial success. In particular, the achievement of independence and creativity is highly correlated with start-up satisfaction.
>
> ... our results indicate that forcing people into situations when they cannot choose among alternatives is likely to result in significant utility losses, independent of other factors.

(Block & Koellinger, 2009)

AUTONOMY: Make sure you have spare discretionary time — even at financial cost.

> having spare time and perceiving control over how to spend that time (i.e. discretionary time) has been shown to have a strong and consistent effect on life satisfaction and happiness, even controlling for the actual amount of free time one has (Eriksson, Rice, & Goodin, 2007; Goodin, Rice, Parpo, & Eriksson, 2008).
>
> Therefore, increase your discretionary time, even if it requires monetary resources. And if you can't afford to, focus on the present moment, breathe more slowly, and spend the little time that you have in meaningful ways.

(Aaker et al., 2010)

Be passionate, but don’t obsess. “Passion Does Make a Difference to People’s Well-Being” (Philippe, Vallerand, & Lavigne, 2009)

Key terms: hedonic well-being; eudaimonic well-being

> Recent research has begun to distinguish two aspects of subjective well-being. Emotional [hedonic] well-being refers to the emotional quality of an individual’s everyday experience—the frequency and intensity of ex- periences of joy, stress, sadness, anger, and affection that make one’s life pleasant or unpleasant. Life evaluation [eudaimonic well-being] refers to the thoughts that people have about their life when they think about it.

(Kahneman & Deaton, 2010)

> The results of two studies provided support for the idea that being harmoniously passionate for an activity contributes significantly to both hedonic and eudaimonic well-being, while being obsessively passionate or not being passionate for any activity does not contribute to well-being at all.
>
> Indeed, merely engaging in a given activity without passion (i.e. being non-passionate) led to the lowest scores on both hedonic and eudaimonic well-being in Study 1 and to the highest decreases in vitality in Study 2 (although no significant differences were found between non-passionate and obsessively passionate people in these studies).
>
> harmoniously passionate people scored significantly higher than obsessively passionate and non-passionate people on hedonic and eudaimonic well-being (Study 1).

only harmoniously passionate people showed a significant increase in vitality over a 1-year period, while obsessively passionate participants showed a slight decrease and non- passionate participants an even larger decrease (Study 2).

only harmonious passion positively predicts well-being over time, while obsessive passion is either negatively associated or unrelated to it (Rousseau & Vallerand, 2003, 2008; Vallerand et al., 2008, Study 2; Vallerand et al., 2007, Studies 1 and 2).

it would appear that an obsessively passionate or non-passionate engagement does not contribute to well-being, and may even have a cost, as shown by the decreases in vitality found in Study 2 for obsessively passionate and non-passionate people.

(Philippe et al., 2009)

Do work that you enjoy doing. Flow.

the accomplishment of goals and the ability to be lost in a task (Csikszentmihalyi and Csikszentmihalyi 1988) seem to be correlated with happiness.

(Nicolao, Irwin, & Goodman, 2009)

Sheldon and Lyubomirsky (2007) have posited that there is much room for improvement in one's happiness. They suggest that while the largest part of our level of happiness is preset by our genetic endowment (around 50%), some 40 per cent is still modifiable (the last 10% is due to uncontrollable circumstances) and the best way to do this is through what they call "intentional activity engagement".

They recommend engaging in interesting, fun activities that fit one's personality and dispositions, that can vary in content, and that are not merely engaged in as a routine but when people feel like doing it. We agree with such a recommendation, especially as Sheldon and Lyubomirsky's definition of intentional activity is rather close to that of harmonious passion.

(Philippe et al., 2009)

In Kasser's view, the secret to SWB is meeting one's intrinsic needs, which means pursuing intrinsic goals out of an intrinsic motivation. In this way, it is similar to Csikszentmihalyi's (1999) view that happiness stems from "flow" experiences, which are also intrinsically motivated. I contend that the shift toward individualistic cultures that accompanies economic development helps people create life-styles that are consis- tent with their preferences and aptitudes (Veenhoven, 1999), and in so doing pursue their intrinsic needs.

(Ahuvia, 2002)

Set goals that are reasonably challenging and reasonably achievable.

One issue concerns the level of challenge posed by one's goals. When life goals are nonoptimally challenging—either too easy or too difficult—positive affect [emotional well-being] is lower (Csikszentmihalyi & Csikszentmihalyi 1988). Low

expectations of success have also been associated with high negative affect (Emmons 1986),

(Ryan & Deci, 2001)

Prefer intrinsic (vs. extrinsic) goals

Definition:

Psychologists make a distinction between two important kinds of goals—intrinsic and extrinsic. Intrinsic goals involve activities and projects that are personally rewarding and meaningful, and that satisfy people's basic needs for competence, relatedness, and autonomy (Kasser & Ryan, 1993, 1996; see Ryan & Deci, 2000, for a review). By contrast, extrinsic goals involve strivings for fame, money, or favorable outward appearances. Research suggests that positive events generated by the fulfillment of intrinsic goals (e.g., making purchases for others rather than yourself) produce more happiness than those generated by extrinsic goals (Dunn, Aknin & Norton, 2008; see also Kasser, 2002; cf. Dunn et al., 2011).

(Chancellor & Lyubomirsky, 2011)

Because high aspirations undermine the benefits of a positive change, are people simply better off with few goals and lowered aspirations? Not necessarily. Ambitious goals held before beginning a new venture motivate people to work harder on that venture and improve their overall performance (Heath, Larrick & Wu, 1999). Individuals would, however, be happier if they focused their monies and efforts on meaningful, intrinsic goals and abandoned extrinsic ones.

Extrinsic goals undermine well-being in several ways. First, by their very nature, extrinsic goals do not satisfy people's basic needs directly, if at all. Instead, much like an addiction (Koob & Le Moal, 2001), such goals lead to ever-increasing desires for psychologically unfulfilling commodities (Myers, 2000). Second, extrinsic goals appear to be incompatible with close, meaningful relationships. Those who pursue extrinsic goals report poorer relationships (Kasser & Ryan, 2001). Indeed, even being reminded of money, as Dunn and colleagues (2011) mention, can cause people to be less prosocial and less generous (Vohs, Mead & Goode, 2006), as well as to be perceived as less friendly and likable by others (Vohs, 2010).

... over-reliance on external contingencies such as becoming famous, wealthy, or attractive may lead to fragile self-worth (Sheldon, Ryan, Deci & Kasser, 2004). For example, a student seeking a law degree from a prestigious and pricey school with the aim of gaining peer respect might become hopelessly depressed if not admitted. Finally, due to limits of attention, time, and energy, extrinsic goals can lead to the neglect of intrinsic pursuits, which are associated with higher well-being (Vohs et al., 2006).

An entrepreneur investing in a new company with the aim of striking it rich might neglect his true interests and hobbies to invest all his energy into his business, and thus miss the need-satisfying personal growth, flow, and joy derived from his more authentic pursuits.

Fittingly expressing the futility and unhappiness wrapped up in pursuing extrinsic goals, a notorious New York tabloid editor confessed that he was “part of that

> strange race of people aptly described as spending their lives doing things they detest to make money they don't want to buy things they don't need to impress people they dislike” (Gauvreau, 1941). As Benjamin Franklin well knew, money is best directed to goals that directly satisfy personal needs such as affiliation, autonomy, and competence rather than expensive pursuits that are unfulfilling and distracting in the end.
>
> In contrast, intrinsic goals, such as building close relationships, making new self-discoveries, and investing in the community, directly activate feelings of satisfaction and con- tentment, which are more likely to be appreciated and less likely to be taken for granted. Dunn and colleagues (2011) rightfully emphasize the link between generosity and well- being, recommending that, to follow the example of Warren Buffett, people spend their money on others rather them themselves.
>
> Intrinsic goals can also trigger “upward spirals”— for example, streams of positive moods and prosocial behavior that gain momentum and reinforce one another as they unfold (Lyubomirsky, King, & Diener, 2005; Norton, Dunn, Aknin & Sandstrom, 2009; Otake, Shimai, Tanaka-Matsumi, Otsui & Fredrickson, 2006).

(Chancellor & Lyubomirsky, 2011)

MATERIALISM AND PURCHASING

The Hedonic Treadmill: We adapt to life changes. Many things that give pleasure will soon cease to do so, thereby driving us to seek more, and more…

> The “pursuit of happiness” is central to the U.S. worldview, yet the very expression also illustrates a paradox of that worldview: Perhaps when one [naively] pursues happiness too single-mindedly, one fails to notice and take advantage of what one already has. In other words, [naively] striving for ever greater happiness may set one on a hedonic treadmill to nowhere. (More on this below)

(Sheldon & Lyubomirsky, 2012)

Prefer experiential purchases; avoid materialistic goals. It is better to collect (positive) experiences than to collect things.

(But do not keep repeating the same positive experience, lest hedonic adaptation set in quicker. See "Stretching Happiness" farther down.)

> Experiential activities are inherently more social (Caprariello & Reis, 2010; Van Boven & Gilovich, 2003) and for this reason fulfill the psychological need for relatedness (Howell & Hill, 2009).

(Howell et al., 2012)

> Discretionary experiential purchases ostensibly foster more social contact than discretionary material purchases (Millar & Thomas 2009; Van Boven, 2005), which is a key component to happiness (Argyle, 2001).

> Research has demonstrated that people are happier with experiential purchases compared to material items.
>
> Experiential purchases are more central to positive self-identity than material purchases.
>
> Further, experiential purchases may satisfy the personal needs of development and growth more than material acquisitions (Kasser & Ryan, 1996).

(Thomas, 2010)

> ... we show that on average the most happiness obtained through purchasing is likely to be obtained through experiential purchases that turn out well.
>
> ... positive social interaction is a major source of happiness; many experiential purchases involve activities with other people, including family.

(Nicolao et al., 2009)

> Materialism might lead to lower well-being because materialistic people tend to downplay the importance of social relationships and to have a large gap between their incomes and material aspirations (Solberg, Diener, & Robinson, 2004).

(Diener & Seligman, 2004)

> Several studies have documented that a materialistic lifestyle is associated with diminished subjective well-being.
>
> ... consistent with previous research, we found that materialism is negatively correlated with life satisfaction (Belk 1984, 1985; Burroughs and Rindfleisch 2002; Christopher et al. 2007; Ryan and Dziurawiec 2001; Wright and Larsen 1993).
>
> ... in line with previous research (Christopher and Schlenker 2004; Chris- topher et al. 2009). High materialistic consumers experience negative emotions more frequently than low materialistic consumers.

(Hudders & Pandelaere, 2011)

> ... we administered three widely used measures of a materialistic value orientation to 92 business students in Singapore. As expected, those students who had strongly internalized materialistic values also reported lowered self-actualization, vitality and happiness, as well as increased anxiety, physical symptomatology, and unhappiness. (Kasser & Ahuvia, 2002)
>
> past research demonstrating that materialistic values are associated with experiences of general and existential insecurity (Pyszczynski et al., 1997; Rindfleisch et al., 2009).

(Howell et al., 2012)

> ... positive experiences not only live on in memories but also lend themselves to even more positive reinterpretations over time as the negative aspects of them fade

(Nicolao et al., 2009)

> ... when security needs are met, it may be more adaptive to broaden one's experience and acquire new knowledge, skills, and relationships that often accompany experiential purchases. These experiences, if they do not arouse competing security concerns, may then provide increased SWB with accompanying reductions in feelings of anxiety and insecurity, encouraging further experiential purchases, and resulting in the 'upward spiral' depicted in our model. In this way, the benefits of an experiential purchasing tendency may accrue over a lifetime and individuals may develop stable purchasing habits.

(Howell et al., 2012)

> In sum, evidence suggests that when looking to spend money, the most satisfying pursuits should involve learning new skills (e.g., mastering a new instrument or learning a foreign language), spending time with others (e.g., taking out one's family to dinner or having coffee with a friend), or doing something good for someone else (e.g., buying Christmas decorations for an elderly neighbor or sending a care package to a sick friend).

(Chancellor & Lyubomirsky, 2011)

> We provide evidence that this purchase type by valence interaction is driven by the fact that consumers adapt more slowly to experiential purchases than to material purchases, leading to both greater happiness and greater unhappiness for experiential purchases.
>
> adaptation happens more quickly for material purchases than for experiential purchases.

(Nicolao et al., 2009)

Don't engage in 'comparison shopping.' And don't place much stock in the happiness potential of any one positive change.

Comparison shopping makes us aware of previously unimportant differences and makes us forget the salient qualities of what we want.

> Sites like [bizrate.com] offer consumers the opportunity to search for everything [...], comparing a vast range of available options within a given category. [...] Recent research suggests that comparison shopping may sometimes come at a cost. By altering the psychological context in which decisions are made, comparison shopping may distract consumers from attributes of a product that will be important for their happiness, focusing their attention instead on attributes that distinguish the available options.
>
> Another problem with comparison shopping is that the comparisons we make when we are shopping are not the same comparisons we will make when we consume what we shopped for (Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2004).
>
> One of the dangers of comparison shopping, then, is that the options we don't choose typically recede into the past and are no longer used as standards for comparison.

> A similar process is likely to unfold in the real estate market. Before purchasing a home, people typically attend scores of open houses and viewings, scrutinizing spec sheets for information about each property's features. Through this process of comparison shopping, the features that distinguish one home from another may come to loom large, while their similarities fade into the background. As a result, home buyers might over- estimate the hedonic consequences of living in a big, beautiful house in a great location versus a more modest home, leading them to take out a larger loan than they can truly afford (potentially sowing the seeds for a nationwide financial crisis).
>
> This suggests that consumers who expect a single purchase to have a lasting impact on their happiness might make more realistic predictions if they simply thought about a typical day in their life.
>
> Conclusion When asked to take stock of their lives, people with more money report being a good deal more satisfied [eudamonic well being]. But when asked how happy they are at the moment, people with more money are barely different than those with less [hedonic well being] (Diener, Ng, Harter, & Arora, 2010). This suggests that our money provides us with satisfaction when we think about it, but not when we use it. That shouldn't happen. Money can buy many, if not most, if not all of the things that make people happy, and if it doesn't, then the fault is ours. We believe that psychologists can teach people to spend their money in ways that will indeed increase their happiness, and we hope we've done a bit of that here.

(Dunn et al., 2011)

> When people consider the impact of any single factor on their well-being—not only income—they are prone to exaggerate its importance. We refer to this tendency as the focusing illusion.

(Kahneman, 2006)

'Follow the herd.' (Dunn et al., 2011)

> Research suggests that the best way to predict how much we will enjoy an experience is to see how much someone else enjoyed it. In one study, Gilbert, Killingsworth, Eyre, and Wilson (2009) asked women to predict how much they would enjoy a speed date with a particular man. Some of the women were shown the man's photograph and autobiography, while others were shown only a rating of how much a previous woman had enjoyed a speed date with the same man a few minutes earlier. Although the vast majority of the participants expected that those who were shown the photograph and autobiography would make more accurate predictions than those who were shown the rating, precisely the opposite was the case. Indeed, relative to seeing the photograph and autobiography, seeing the rating reduced inaccuracy by about 50%. It appears that the 17th century writer François de La Rochefoucauld was correct when he wrote: “Before we set our hearts too much upon anything, let us first examine how happy those are who already possess it.”

(Dunn et al., 2011)

INTERPERSONAL

Socialize — with the right people.

> ... the effects of wealth are not large, and they are dwarfed by other influences, such as those of personality and social relationships.

(Diener & Seligman, 2004)

> ... it is not only whether you spend your time with others that influences your happiness, but also who you spend your time with. Interaction partners associated with the greatest happiness levels include friends, family, and significant others, whereas bosses and co-workers tend to be associated with the least happiness (Kahneman et al., 2004).
>
> social leisure activities contribute more to happiness than solitary ones (Reyes-Garcia et al., 2009).
>
> Furthermore, people who frequently engage in social activities experience higher levels of happiness than people who participate in social activities less often (Lloyd & Auld, 2002), and being with others typically improves the quality of an experience (whereas being alone makes most people sad, lonely, or both; Csikszentmihalyi & Larson, 1984; Lewinsohn, Sullivan, & Grosscup, 1982).

(Aaker et al., 2010)

> Compared with the less happy groups, the happiest respondents did not exercise significantly more, participate in religious activities significantly more, or experience more objectively defined good events. No variable was sufficient for happiness, but good social relations were necessary.
>
> Our findings suggest that very happy people have rich and satisfying social relationships and spend little time alone relative to average people. […] In contrast, unhappy people have social relationships that are significantly worse than average.

(Diener & Seligman, 2002)

> Income and education are more closely related to life evaluation, but health, care giving, loneliness, and smoking are relatively stronger predictors of daily emotions.

(Kahneman & Deaton, 2010)

Associate with happy people.

> ... research has also shown that our relationships with weak ties, and even strangers, can affect our happiness. Using a large-scale, longitudinal dataset, Fowler and Christakis [5] suggested that happiness spreads throughout social networks,extending up to three degrees of separation: a person becomes happier if their friend’s friend’s friend becomes happier, even if they don’t know that person.

(Aknin, Sandstrom, Dunn, & Norton, 2011b)

> People who are surrounded by many happy people and those who are central in the network are more likely to become happy in the future. Longitudinal statistical

> models suggest that clusters of happiness result from the spread of happiness and not just a tendency for people to associate with similar individuals. A friend who lives within a mile (about 1.6 km) and who becomes happy increases the probability that a person is happy by 25% (95% confidence interval 1% to 57%). Similar effects are seen in coresident spouses (8%, 0.2% to 16%), siblings who live within a mile (14%, 1% to 28%), and next door neighbours (34%, 7% to 70%). Effects are not seen between coworkers. The effect decays with time and with geographical separation.

(Fowler & Christakis, 2008)

Give the people around you opportunities to be generous. Ask them for favors.

You can possibly make people around you happier by allowing them to be kind and generous, and you want to surround yourself with happy people (see above). Aside from making them happier, you will also improve your relationship with them via the Benjamin Franklin effect, which — unintuitively — makes people like you more if you ask them for favors.

Be actively kind (and occasionaly reminisce about your recent acts of kindness).

> Subjective happiness was increased simply by counting one's own acts of kindness for one week.
>
> Happy people became more kind and grateful through the counting kindnesses intervention.
>
> Our results further suggest that a reciprocal relationship may exist between kindness and happiness, as has been shown for gratitude and happiness [see below].

(Otake et al., 2006)

STRETCHING HAPPINESS (fighting hedonic adaptation)

Hedonic adaptation — definition:

> The pleasure of success and the ignominy of failure abate with time. So does the thrill of a new sports car, the pain over a failed romance, the delight over a promotion, and the distress of a scary diagnosis. This phenomenon, known as hedonic adaptation (HA), has drawn increasing interest from both psychologists and economists (e.g., Diener, Lucas, & Scollon, 2006; Easterlin, 2006; Frederick & Loewenstein, 1999; Kahneman & Thaler, 2006; Lucas, 2007a; Lyubomirsky, 2011; Lyubomirsky, Sheldon, & Schkade, 2005; Wilson & Gilbert, 2008).

(Sheldon & Lyubomirsky, 2012)

Choose smaller, more frequent successes rather than larger ones.

Even big positive changes can get old fast, and soon stop bringing happiness.

> ...every one of the published studies evidences fairly rapid and apparently complete adaptation to positive changes. The most widely-cited study is that of Brickman and his colleagues (1978), who reported that lottery winners were no happier up to 18 months after the news than those who had experienced no windfall.

(Sheldon & Lyubomirsky, 2012)

Go for variety and surprise. Don't keep doing the same thing.

> ...variable stimuli resist adaptation more than do unchanging stimuli (see also Wilson & Gilbert, 2008).

(Chancellor & Lyubomirsky, 2011)

> ...these findings support the notion that variety and surprise spice up life in ways that sustain well-being (Sheldon et al., in press; Sheldon & Lyubomirsky, 2006, 2009; Wilson, Centerbar, Kermer, & Gilbert, 2005)"

(Sheldon & Lyubomirsky, 2012)

Savor the anticipation. Delay consumption. Actively anticipate good experiences.

> Research in the field of neuroscience has shown that the part of the brain responsible for feeling pleasure, the mesolimbic dopamine system, can be activated when merely thinking about something pleasurable, such as drinking one's favorite brand of beer (McClure, Li, Tomlin, Cypert, Montague, & Montague, 2004) or driving one's favorite type of sports car (Erk, Spitzer, Wunderlich, Galley, & Walter, 2002).
>
> ... the brain sometimes enjoys anticipating a reward more than receiving the reward (Loewenstein, 1987; Berns, McClure, Pagnoni, & Montague, 2001).
>
> ... the pleasure derived from window shopping for a dress may exceed the pleasure from actually acquiring the dress.

(Aaker et al., 2010)

(Perhaps the above can inform the discourse on the [ir]rationality of lotteries.)

Divide positive experiences into smaller pleasures, if possible.

> Dividing consumption into smaller doses and separating it out over time can multiply [the pleasure of] "first bites," and subsequently, the enjoyment. Savoring a chocolate bar could be as simple as dividing it into squares and eating one piece per day, instead of devouring it all in a single sitting. Research supports the idea that breaks are beneficial for positive experiences, such as enjoying a television program, but detrimental for negative experiences, such as enduring a dental drill (Nelson, Meyvis & Galak, 2009).

(Chancellor & Lyubomirsky, 2011)

Dividing into smaller doses also increases the amount of pleasurable anticipation. See previous subsection.

Corollary: Conclude negative experiences as soon as possible.

Don’t "think about it tomorrow.” Prolongation increases the effect of both negative and positive experiences, and bad is stronger than good:

> Although the same hedonic adaptation process is involved in both positive and negative experiences, an important asymme- try exists between the two that further complicates efforts to remain happy, especially if a positive change comes at a high financial cost. To sum up almost two decades of research, bad is stronger than good (Baumeister, Bratslavsky, Finkenauer & Vohs, 2001; see also Taylor, 1991), or as Einstein quipped, “Put your hand on a hot stove for a minute, and it seems like an hour. Sit with a pretty girl for an hour, and it seems like a minute.” [...] positive changes are weaker than negative changes, and that their effects also evaporate more quickly (e.g., Nezlek & Gable, 2001; Sheldon, Ryan & Reis, 1996; see also Oishi, Diener, Choi, Kim-Prieto & Choi, 2007).

(Chancellor & Lyubomirsky, 2011)

Make a point of avoiding experiences that make you feel bad.

> Well-being is about more than just frequently feeling good—it is also about infrequently feeling bad (Diener, Suh, Lucas & Smith, 1999).
>
> All else being equal, the elimination of negative experiences could provide a three- to five-fold hedonic return on investment over the creation of positive experiences, due to positive/negative asymmetry (e.g., David, Green, Martin & Suls, 1997; Fredrickson & Losada, 2005; Gottman, 1994).

(Chancellor & Lyubomirsky, 2011)

APPRECIATION

Be grateful. Count your blessings (literally). Recycle happiness. Reminisce about good experiences.

> A number of experiments have demonstrated that the regular practice of gratitude—a practice closely related to and often indistinguishable from appreciation and savoring—brings about significant increases in well-being when performed over the course of 1 to 12 consecutive weeks. For example, relative to performing neutral activities, the intentional and effortful practice of “counting one's blessings” once a week (Emmons & McCullough, 2003; Froh, Sefick & Emmons, 2008; Lyubo- mirsky, Sheldon, & Schkade, 2005) or penning appreciation letters to individuals who have been kind and meaningful (Boehm, Lyubomirsky, & Sheldon, in press; Lyubomirsky, Dickerhoof, Boehm, & Sheldon, in press; Seligman, Steen, Park & Peterson, 2005) has been shown to produce increases in happiness for as long as 6 months.

(Chancellor & Lyubomirsky, 2011)

Think of counterfactuals. (“If I didn’t have this, what do I lose?”)

> Another cognitive exercise that directs attention toward existing positive changes or events is counterfactual thinking. This strategy involves mentally subtracting a purchased positive experience from ever having taken place, and enumerating all the subsequent blessings that also would have disappeared (Koo, Algoe, Wilson, & Gilbert, 2008).

(Chancellor & Lyubomirsky, 2011)

Breathe deeply. Expand your time — by slowing down.

> [People feel less rushed and hurried when they] simply breathe more deeply. In one study, subjects who were instructed to take long and slow breaths (vs. short and quick ones) for 5 minutes not only felt there was more time available to get things done, but also perceived their day to be longer.

(Aaker et al., 2010)

Stay in the present.

> One possible benefit of being present-focused is that thinking about the present moment (vs. the future) slows down the perceived passage of time, allowing people to feel less rushed and hurried (Rudd & Aaker, 2010).

(Aaker et al., 2010)

OPTIMAL HAPPIFICATION

Actively *want* to be happier. Motivation and investment matter.

> First, and most important, we found that to become happier, people need both a will and a proper way. The will can come from motivation, expectations, and diligence. The proper way comes from performing the “right” activity, not merely a placebo. Accordingly, we found that motivation and investment in becoming a happier person matters. That is, expressing gratitude and optimism did not generally increase well-being unless a person was truly cognizant of the exercises’ purpose and motivated to improve his or her happiness. Second, effortful pursuit of happiness activities was found to be important to improving and maintaining well- being.
>
> ... happiness interventions are more than just placebos, but […] they are most successful when participants know about, endorse, and commit to the intervention.
>
> According to our model of well-being change (Lyubomirsky, Sheldon, et al., 2005; Sheldon & Lyubomirsky, 2004), sustainable increases in happiness are possible, but only if pursued under optimal conditions, such as when people are motivated to perform a positive activity, when they bring to bear effort and persistence, and when the activity is a legitimately efficacious one.

(Lyubomirsky et al., 2011)

Learn about the science of happiness. Internalize the recommendations in this article and in here.

> ... people often hold incorrect intuitive theories about the determinants of happiness. For instance, they overestimate the impact of specific life events on their experienced well-being with regard to intensity, as well as with regard to duration. (see also Comparison Shopping above)
>
> ... four major sources for systematic over- and undervaluation of choice options that can be distinguished: (i) the underestimation of adaptation, (ii) distorted memory of past experiences, (iii) the rationalization of decisions, and (iv) false intuitive theories about the sources of future utility.

(Quoidbach et al., 2010)

> Money is an opportunity for happiness, but it is an opportunity that people routinely squander because the things they think will make them happy often don't.
>
> It is not surprising when wealthy people who know nothing about wine end up with cellars that aren't that much better stocked than their neighbors', and it should not be surprising when wealthy people who know nothing about happiness end up with lives that aren't that much happier than anyone else's.

(Dunn et al., 2011)

EXTRA CONSIDERATIONS

> Happiness predicts [future] income.

(Diener & Seligman, 2004)

(^But try not to think of it that way!)

> ... we found that with all but one specification, initial happiness levels were positively and significantly correlated with future earnings. [...] An additional finding is that the effects of initial period happiness on future income and on future happiness seem to be more consistent across all income groups than are the effects of initial period income on either future income and future happiness. The effects of initial period income seem more important for those at higher levels of income.
>
> The studies by psychologists that find that happiness has positive effects on future income also find that these effects are stronger at the higher end of the income scale. See Diener and Biswas-Diener (1999).

(Graham, Eggers & Sukhtankar, 2004)

> ... higher cheerfulness in the first year of college correlated with higher income 19 years or so later, when respondents reached their late 30s; this effect was greatest for those who came from the most affluent families

(Diener & Seligman, 2004)

It will be easier to stay happy when you become happier

> ... our findings also dovetail with those of Cohn and Fredrickson (in press) by demonstrating that initial happiness gains can cause a happiness intervention to become self-reinforcing.

(Aknin, Dunn, & Norton, 2011a)

Happiness Interventions Work!

> Fordyce (1977, 1983) created an intervention program based on the idea that people's subjective well-being can be increased if they learn to imitate the traits of happy people, such as being organized, keeping busy, spending more time socializing, developing a positive outlook, and working on a healthy personality. Fordyce found that the program produced increases in happiness compared to a placebo control, as ell as compared to participants in conditions receiving only partial information. Most impressive, he found lasting effects of the intervention in follow-up evaluation 9-28 months after the study.

(Diener et al., 2009)

> Recently, a number of additional effective interventions on happiness have been reported, ranging from the kindness interventions (Otake, Shimai, Tanaca-Matumi, Otsui, & Fredrickson, 2006) and gratitude interventions (Emmons & McCulough, 2003) to variants of the writing intervention (King, 2001; Lyubomirsky, Sousa, & Dickerhoof, 2006). Recent intervention studies are clearly promising. However, more diverse dependent variables and measuring instruments would be desirabe, as well as explorations of which interventions are most beneficial, and why.

(Diener et al., 2009)

Extra extra: Cultural Differences

> Veenhoven (1999) found that among poor countries, individualism was negatively associated with happiness; whereas among richer countries, individualism was positively associated with happiness. This suggests that economic growth is part of a complex system of modernization that needs to be seen holistically. Collectivism may exist in poorer countries because it is highly functional in that environment, but it may give way to more individualism as societies modernize and the needs of those societies change. Overall, individualism/collectivism stands out as an extremely promising construct for explaining differences in national average levels of SWB, when investigated holistically as part of the larger social system (Cummins, 1998; Myers and Diener, 1995).

> ... economic development increases SWB by creating a cultural environment where individuals make choices to maximize their happiness rather than meet social obligations (Coleman, 1990; Galbraith, 1992; Triandis, 1989; Triandis et al., 1990; Veenhoven, 1999; Watkins and Liu, 1996). This cultural transformation away from obligation and toward the pursuit of happiness is part of a broader transition away from collectivism and toward individualist cultural values and forms of social organization.

> Cross-cultural research shows that values like “enjoying life” and leading “an exciting life” are stronger in individualist societies, whereas “social recognition,” “preserving my public image,” being “humble,” and “honoring parents and elders” are particularly strong in collectivist societies (Triandis et al., 1990, p. 1015). There is no more reason to think that people seek social recognition with the ultimate goal of personal happiness, than there is to think that people seek happiness with the ultimate goal of getting others to think well of them for having such a pleasant affect.

(Ahuvia, 2002)

References

Aaker, J. L., Rudd, M., & Mogilner, C. (2010). *If Money Doesn’t Make You Happy, Consider Time*. Journal of Consumer Psychology, 2011.

Ahuvia, A. C. (2002). *Individualism/collectivism and cultures of happiness: A theoretical conjecture on the relationship between consumption, culture and subjective well-being at the national level*. Journal of Happiness Studies, 3(1), 23–36. Springer.

Aknin, L. B., Dunn, E. W., & Norton, M. I. (2011a). *Happiness Runs in a Circular Motion: Evidence for a Positive Feedback Loop between Prosocial Spending and Happiness*. Journal of Happiness Studies, 13(2), 347–355. doi:10.1007/s10902-011-9267-5

Aknin, L. B., Sandstrom, G. M., Dunn, E. W., & Norton, M. I. (2011b). *It's the Recipient That Counts: Spending Money on Strong Social Ties Leads to Greater Happiness than Spending on Weak Social Ties*. (M. Perc, Ed.)PLoS ONE, 6(2), e17018. doi:10.1371/journal.pone.0017018

Aknin, L. B., Barrington-Leigh, C. P., Dunn, E. W., Helliwell, J. F., Biswas-Diener, R., Kemeza, I., Nyende, P., et al. (2010). *Prosocial spending and well-being: cross-cultural evidence for a psychological universal*. National Bureau of Economic Research.

Block, J., & Koellinger, P. (2009). *I Can't Get No Satisfaction—Necessity Entrepreneurship and Procedural Utility*. Kyklos, 62(2), 191–209. doi:10.1111/j.1467-6435.2009.00431.x

Chancellor, J., & Lyubomirsky, S. (2011). *Happiness and thrift: When (spending) less is (hedonically) more*. Journal of Consumer Psychology, 21(2), 131.

DeVoe, S. E., & House, J. (2012). *Time, money, and happiness: How does putting a price on time affect our ability to smell the roses?* Journal of Experimental Social Psychology, 48(2), 466–474. doi:10.1016/j.jesp.2011.11.012

Diener, E., Oishi, S., Lucas, R.E. (2009). *Subjective Well-Being: The science of happiness and life satisfaction*. Oxford Handbook of Positive Psychology, 187-194.

Diener, E., & Seligman, M. P. (2002). *Very happy people*. Psychological Science, 13(1), 81–84.

Diener, E., & Seligman, M. P. (2004). *Beyond money*. Psychological science in the public interest, 5(1), 1–31.

Drnovšek, M., Örtqvist, D., & Wincent, J. (2010). *The effectiveness of coping strategies used by entrepreneurs and their impact on personal well-being and venture performance*. Journal of Economics and Business, 28, 193–220.

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). *Spending Money on Others Promotes Happiness*. Science, 319(5870), 1687–1688. doi:10.1126/science.1150952

Dunn, E. W., Ashton-James, C. E., Hanson, M. D., & Aknin, L. B. (2010). *On the Costs of Self-interested Economic Behavior: How Does Stinginess Get Under the Skin?* Journal of Health Psychology, 15(4), 627–633. doi:10.1177/1359105309356366

Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). *If money doesn“t make you happy, then you probably aren't spending it right*. Journal of Consumer Psychology, 21(2), 115.

Fowler, J. H., & Christakis, N. A. (2008). *The dynamic spread of happiness in a large social network*. BMJ: British medical journal, 337, a2338.

Frey, B. S., & Stutzer, A. (2002). *The economics of happiness*. World Economics, 3(1), 1–17.

Graham, C., Eggers, A., & Sukhtankar, S. (2004). *Does happiness pay?: An exploration based on panel data from Russia*. Journal of Economic Behavior & Organization, 55(3), 319–342. Elsevier.

Howell, R. T., Pchelin, P., & Iyer, R. (2012). *The preference for experiences over possessions: Measurement and construct validation of the Experiential Buying Tendency Scale*. The Journal of Positive Psychology, 7(1), 57–71.

Hudders, L., & Pandelaere, M. (2011). *The Silver Lining of Materialism: The Impact of Luxury Consumption on Subjective Well-Being*. Journal of Happiness Studies. doi:10.1007/s10902-011-9271-9

Kahneman, D. (2006). *Would You Be Happier If You Were Richer? A Focusing Illusion*. Science, 312(5782), 1908–1910. doi:10.1126/science.1129688

Kahneman, D., & Deaton, A. (2010). *High income improves evaluation of life but not emotional well-being*. Proceedings of the National Academy of Sciences, 107(38), 16489–16493. doi:10.1073/pnas.1011492107

Kasser, T., & Ahuvia, A. (2002). *Materialistic values and well-being in business students*. European Journal of Social Psychology, 32(1), 137–146. doi:10.1002/ejsp.85

Lyubomirsky, S., Dickerhoof, R., Boehm, J. K., & Sheldon, K. M. (2011). *Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being*. Emotion, 11(2), 391.

Mogilner, C. (2010). *The Pursuit of Happiness: Time, Money, and Social Connection*. Psychological Science, 21(9), 1348–1354. doi:10.1177/0956797610380696

Nickerson, C., Schwarz, N., Diener, E., & Kahneman, D. (2003). *Zeroing in on the Dark Side of the American Dream A Closer Look at the Negative Consequences of the Goal for Financial Success*. Psychological Science, 14(6), 531–536.

Nicolao, L., Irwin, J., & Goodman, J. (2009). *Happiness for Sale: Do Experiential Purchases Make Consumers Happier than Material Purchases?* Journal of Consumer Research, 36(2), 188–198. doi:10.1086/597049

Otake, K., Shimai, S., Tanaka-Matsumi, J., Otsui, K., & Fredrickson, B. L. (2006). *Happy People Become Happier through Kindness: A Counting Kindnesses Intervention*. Journal of Happiness Studies, 7(3), 361–375. doi:10.1007/s10902-005-3650-z

Philippe, F. L., Vallerand, R. J., & Lavigne, G. L. (2009). *Passion does make a difference in people's lives: A look at well-being in passionate and non-passionate individuals*. Applied Psychology: Health and Well-Being, 1(1), 3–22.

Quoidbach, J., Dunn, E. W., Petrides, K. V, & Mikolajczak, M. (2010). *Money Giveth, Money Taketh Away*. Psychological Science, 21(6), 759.

Ryan, R. M., & Deci, E. L. (2001). *On happiness and human potentials: A review of research on hedonic and eudaimonic well-being*. Annual Review of Psychology, 52(1), 141–166.

Sheldon, K. M., Lyubomirsky, S. (2012).*The Challenge of Staying Happier: Testing the Hedonic Adaptation Prevention Model*. Personality and Social Psychology Bulletin. doi:10.1177/0146167212436400

Stutzer, A., & Frey, B. S. (2010). *Recent advances in the economics of individual subjective well-being*. Social Research: An International Quarterly, 77(2), 679–714.

Thomas, R. L. (2010). *Mediating and moderating variables between discretionary purchases and happiness*. UNLV Theses/Dissertations/Professional Papers/Capstones. Paper 889.

How can we get more and better LW contrarians?

I'm worried that LW doesn't have enough good contrarians and skeptics, people who disagree with us or like to find fault in every idea they see, but do so in a way that is often right and can change our minds when they are. I fear that when contrarians/skeptics join us but aren't "good enough", we tend to drive them away instead of improving them.

For example, I know a couple of people who occasionally had interesting ideas that were contrary to the local LW consensus, but were (or appeared to be) too confident in their ideas, both good and bad. Both people ended up being repeatedly downvoted and left our community a few months after they arrived. This must have happened more often than I have noticed (partly evidenced by the large number of comments/posts now marked as written by **[deleted]**, sometimes with whole threads written entirely by deleted accounts). I feel that this is a waste that we should try to prevent (or at least think about how we might). So here are some ideas:

- Try to "fix" them by telling them that they are overconfident and give them hints about how to get LW to take their ideas seriously. Unfortunately, from their perspective such advice must appear to come from someone who is themselves overconfident and wrong, so they're not likely to be very inclined to accept the advice.
- Create a separate section with different social norms, where people are not expected to maintain the "proper" level of confidence and niceness (on pain of being downvoted), and direct overconfident newcomers to it. Perhaps through no-holds-barred debate we can convince them that we're not as crazy and wrong as they thought, and *then* give them the above-mentioned advice and move them to the main sections.
- Give newcomers some sort of honeymoon period (marked by color-coding of their usernames or something like that), where we ignore their overconfidence and associated social transgressions (or just be extra nice and tolerant towards them), and take their ideas on their own merits. Maybe if they see us take their ideas seriously, that will cause them to reciprocate and take us more seriously when we point out that they may be wrong or overconfident.

I guess these ideas sounded better in my head than written down, but maybe they'll inspire other people to think of better ones. And it might help a bit just to keep this issue in the back of one's mind and occasionally think strategically about how to improve the person you're arguing against, instead of only trying to win the particular argument at hand or downvoting them into leaving.
P.S., after writing most of the above, I saw [this post]:

> OTOH, I don’t think group think is a big problem. Criticism by folks like Will Newsome, Vladimir Slepnev and especially Wei Dai is often upvoted. (I upvote almost every comment of Dai or Newsome if I don’t forget it. Dai makes always very good points and Newsome is often wrong but also hilariously funny or just brilliant and right.) Of course, folks like this Dymytry guy are often downvoted, but IMO with good reason.

To be clear, I don't think "group think" is the problem. In other words, it's not that we're refusing to accept valid criticisms, but more like our group dynamics (and other factors) cause there to be fewer good contrarians in our community than is optimal. Of course what is optimal might be open to debate, but from my perspective, it can't be right that my own criticisms are valued so highly (especially since I've been moving closer to the SingInst "inner circle" and my critical tendencies have been decreasing). In the spirit of making oneself redundant, I'd feel much better if my occasional voice of dissent is just considered one amongst many.

How accurate is the quantum physics sequence?

Prompted by Mitchell Porter, I asked on Physics StackExchange about the accuracy of the physics in the Quantum Physics sequence:

> What errors would one learn from Eliezer Yudkowsky's introduction to quantum physics?
>
> Eliezer Yudkowsky wrote an introduction to quantum physics from a strictly realist standpoint. However, he has no qualifications in the subject and it is not his specialty. Does it paint an accurate picture overall? What mistaken ideas about QM might someone who read only this introduction come away with?

I've had some interesting answers so far, including one from a friend that seems to point up a definite error, though AFAICT not a very consequential one: in Configurations and Amplitude, a multiplication factor of *i* is used for the mirrors where -1 is correct.

Physics StackExchange: What errors would one learn from Eliezer Yudkowsky's introduction to quantum physics?

Crowdsourcing the availability heuristic

There are goals which can be achieved only by personal exertion and hard work – finishing a university degree, learning a language, mastering a martial art… But there is also a plethora of smaller goals, where small differences in approach and resources can make a huge difference. I'm going to examine how one particular cognitive bias affects execution of small-to-midrange goals, why this bias cannot be realistically overcome on a personal level, and how it can be effectively short-circuited simply by involving other minds.

Do note: the point I'm making may seem obvious; in my personal experience, and from observation, it is one of those things that are obvious once you know the answer (and one still needs occasional reminders). The solution to the problem is high-impact and available to practically everyone, but remains vastly underused.

The bias in question is *availability heuristic*. LW has a decent definition:

> The availability heuristic judges the probability of events by the ease with which examples come to mind. Sometimes this heuristic serves us well, but the map is not the territory; the frequency with which concepts occur in your thoughts need not reflect the frequency with which they occur in reality. Undue salience, selective reporting, even subtle features of how the human brain stores and recalls memories can distort our perceptions about the probability of events. Because it's easier to recall words by their first letter, people judge words that begin with the letter r to be more frequent than words with r as their third letter, even though in fact, the latter is more frequent. Or selective reporting by the media of dramatic tragedies makes them seem more frequent than more threatening albeit mundane risks.

This topic has been talked about many times on LW, and there is a great deal of academic research as well [1], including seminal texts many here will be familiar with [2]. It's all interesting, but there are two critical points I want to pull up to the forefront.

I - availability heuristic is commonly treated as a simple perceptual bias. In actuality, it is also a *choice bias* - if you fail to perceive an option, you cannot choose to pursue it. If you do pursue an option, you are likely going to focus on attaining it with cognitively available resources, while missing much better resources that are sitting idle. Similarly, you may waste resources (sometimes to the point of simply giving up) while pursuing a sub-optimal but cognitively available path towards your goal – completely oblivious to much easier roads which are actually available to you.

II - availability heuristic is not really "curable." Sure, it's good to be aware of it. But we all have limited information about the world. Even if we objectively write down all known relevant factors for some observation or decision, our sample is still going to be at least somewhat biased, and certainly very narrow. Outside our direct areas of expertise, the amount of information we can include into any decision is quite limited; and the number of items our working memory holds while the decision is made is limited yet further.

One obvious mitigation strategy is probably apparent: simply research your desired goal by consulting experts (or the Googlian Oracle). How did other people achieve the goal? What preparations did they undertake? What strategies are recommended?

Good idea, which comes with several limitations.

Much of the information out there is written by people who are unaware of the existence of availability heuristic. The recommendations, however refined, are still usually descriptions of single strategies which happened to work optimally for the person who wrote the article. They could work (and often will), but they may not be the optimal solution to your particular problem set. Furthermore, as I will illustrate soon, the most common strategies you will find are the ones *most cognitively available* to the most people; an Internet search will, in effect, *potentiate* the availability heuristic even further, hiding less obvious strategies even further.

But by far the most significant limitation comes from the fact that knowledge does not equal resources. Even if you research an optimal strategy, you may still be unaware of the full scope of resources that are available to you. To avoid abstraction, let's take a specific example.

Example: crowdsourcing adventure opportunities

A friend of mine has an interesting strategy for increasing the overall awesomeness of her life. Every January 1st, she comes up with a general rule, which she then follows until the end of the year. The rule is modified by common sense (you don't follow it if it will get you into an extraordinarily dangerous situation, or if following it is otherwise prohibitively expensive), but other than that, it *has* to be followed.

This year, the rule is "when you think or hear of something that causes you to be afraid, go ahead and do it." She's afraid of heights, so the obvious "go skydiving" is on the list from the start. But then, someone mentions flying in an acrobatic aircraft. That gets added to the list. I'm sure you see the general principle.

Seems a bit cheesy, at first glance. But... within the first four months of this year, she went flying in the aforementioned acrobatic aircraft (and a helicopter), learned how to ski, and even *rode a damn ostrich*. All within *four months.* She went caving this past weekend. Hang-gliding is firmly scheduled in a few months. And there are other, less glamorous experiences as well, but it's quite a list.[3, 4]

Now, let's say I wanted to do one of those things – say, an acrobatic airplane ride. What comes to mind? I look up places that offer such rides. I would need to travel there (and given the distances involved, get a hotel room as well). Pay the fee. Get to spend about 15-20 minutes being flown around. It's a lot of effort, and the payoff doesn't seem really worth it. Hey, let's see what other people did! Google, google... a bunch of testimonials about people having a great/awful time taking the aforementioned touristy rides (most significant finding: corkscrews often cause explosive nausea). So I give up.

How did my friend do it? She talks to people about it; asks them for ideas. And soon enough, someone says "yeah, I know a guy who owns an acrobatic plane, wanna ride with him?" And lo and behold, a free ride of much higher quality than touristy nonsense one pays for, plus it's very close to home.

My approach above is the cognitively available one. I'm proceeding towards the goal in accordance with the patterns I've followed previously, when achieving similar goals.

I'm thinking about resources that are available to me, personally (my money, my time, etc.). I end up with a suboptimal plan.

Her approach is to crowdsource: throw the desire into the world, and see what others come up with. Many people, with many different resources, ideas, and further links to even more people out there.

Once I started thinking about this, I decided to test this concept. I tried throwing out the acrobatic ride idea to my friends - and lo and behold, a friend of a friend of a friend is going to be flying in this summer. In his acrobatic aircraft. And now, when he gets here, I'm likely to get an hour or so of riding time with him, for free. Just because I asked.

It's a somewhat silly example, of course, but I think it illustrates the point. This friend (of a friend)^2 is a *resource.* I was unaware I had this resource, until I asked for it. Finding an acrobatic aircraft owner through personal connections is a strategy that would have never occurred to me (since my availability heuristic informs me that such people are exceedingly rare).

Does this still seem obvious? Many articles were written about the “breakthrough” design of the Apple headquarters – a building made to force people together, to produce conversations between workers in different areas, and interactions between people who think in very different ways. In MIT lore, legends are written about Building 20, a “magical incubator” that has produced an incredible amount of breakthrough technologies and world-class thinkers. One of the main reasons given for this productivity is that many disparate small groups of researchers in a wide range of areas were thrown together in a small space – where they had to interact and talk to each other.

In other words, there is a certain kind of “magic”... in places that force people to simply utilize each other’s cognitive resources, to seek out different ways of thinking, and to avoid falling into cognitively available approaches to the problems they are trying to solve. And the point I’m trying to make is that one doesn’t need to work in Building 20 – just to intentionally maximize the utilization of their own social network (and work on diversifying it as much as possible).

There are people who already utilize their social network to the utmost, and who expand it strategically, adding people just to enhance the diversity of available viewpoints. But I will take a chance, and state that most of us probably don’t. And as a result, we aren’t able to recognize all of the resources available to us, to optimally use those we do recognize, or to realize optimal strategies for approaching our goals. Chances are that most of us could improve the strategy and execution of *any* given midrange goal – simply by asking around.

(EDIT) Addendum: help-seeking, status, etc.

There is a bit of discussion in the comments regarding some important questions - how and when does seeking help affect status within the group, would seeking help on a regular basis cause people to become uninterested in helping, etc. These may become a basis for a different text in the near future.

But these questions miss an important point here. Sure, asking for help can be a part of the strategy I discuss above, in some cases. In most cases, however, *you should not*

be seeking help. The point of the article here is to simply *talk* about your goal and your strategy with others. "Involving other minds" does not necessarily require them to take an active helper role in the achievement of your goal.

In the specific example given in the text, I didn't go around saying "hey, I'm looking for help in finding an acrobatic airplane ride." Instead, I would say something like "Riding an acrobatic plane seems like an interesting thing; I'm trying to look into finding an opportunity to do so in the near future." Thoughts, offers and the eventual connection grew organically from the discussions that followed. Sure, the pilot himself is going to be doing me a favor (which I'll eventually repay), but the people who made the connection for me were just having a conversation.

To illustrate further on an example that popped up in the comments: a CEO of a company that always asks for help in making decisions will rapidly lose status (and therefore become ineffective at his or her job). This much is true. But most effective CEOs will organize their companies so that people of varying backgrounds will have to talk (at some point or other) about current company projects and strategies. The CEO doesn't ask for help in making the decision: she *requires* that her underlings produce ideas and overviews, which then become a basis for making optimal decisions.

References:

[1] A few recent examples: Hayibor, S., Wasieleski, D.M. (2009). "Effects of the use of availability" Journal of Business Ethics 84: 151–165. Also, Klinger, D., Kudryavtsev, A. (2010). "The availability heuristic and investors' reactions to company-specific events" The Journal of Behavioral Finance 11 (50-65).

[2] Tversky, A., Kahneman, D. (1973) "Availability: A heuristic for judging frequency and probability" Cognitive Psychology 5 (1): 207–233.

[3] She's blogging her progress through this year, and the whole thing is highly recommended; for sheer hilarity as much as for some very interesting insights.

[4] One could write an excellent text on goal-setting strategies around this example. The rule is simple and absolute ("never have any chips, cookies or other snacks available at home" is easier to follow and will in most cases lead to a greater weight loss than an intricate diet), and it is overarching (applies everywhere in life, not only to some particular times and places, making lawyering around the rule much more difficult). If you are going to set rules for yourself, this is the way to do it. But since I'm writing a loose set of texts on a completely different topic, this footnote will be all I have to say on that topic.

SotW: Be Specific

(The Exercise Prize series of posts is the Center for Applied Rationality asking for help inventing exercises that can teach cognitive skills. The difficulty is coming up with exercises interesting enough, with a high enough hedonic return, that people actually do them and remember them; this often involves standing up and performing actions, or interacting with other people, not just working alone with an exercise booklet and a pencil. We offer prizes of $50 for any suggestion we decide to test, and $500 for any suggestion we decide to adopt. This prize also extends to LW meetup activities and good ideas for verifying that a skill has been acquired. See here for details.)

Exercise Prize: Be Specific

During YCombinator's Startup School 2011, Paul Graham and Harj Tagger did "office hours" onstage. One pair of entrepreneurs were doing a matchmaking (dating) startup, and Paul and Harj were trying to figure out what their startup *did,* exactly - for example, what their startup could do that the existing low-tech solution couldn't. (Video.)

> Harj: Low-tech like, you know, just like word of mouth, telling someone "hey, you should like, meet up with my friend" or "we're getting drinks, why don't you come along?" Like, what can the software do that's specifically better than that?
>
> Entrepreneur: I think that our software specifically is providing the better connections for people, um...
>
> Paul: Providing the better connections for people...?
>
> Entrepreneur: I mean, one way you can think about it, I don't know if this is the right answer, but... there's a lot of things that are happening in real life that they're trying to mimic online, maybe that's not the correct way to... Look at it like this: to give them an online tool to also do this, like they're already doing in real life, maybe they could reach, uh expand their reach through the online website.

This had been happening with *most* of the startups Paul and Harj were interrogating - they just *could not* seem to provide a customer use-case - and I couldn't *stand* it any more; which is why at this point I whispered audibly enough for a few nearby people to hear, "Be specific! Be specific!"

A moment later, on stage:

> Paul: Hm. Not very specific.

I got some strange looks from the people sitting next to me.

I hope this provides some background for my guess that around half of Paul Graham's advantage is based on years of incubator experience, and the other half is unusual rationality skills of the sort that the Center for Modern Rationality is trying to figure out how to teach. Obviously this is only a very rough conjecture. But you can see the basis for the hope that - after a fair amount more work - we'll be able to offer a 2-day

course for YCombinator entrepreneurs that eliminates 50% of the overhead from their conversations with Paul Graham.

(Also, note how this post starts off with a specific example - an instance of the *concrete-abstract* writing pattern in which you state the example first and the generalization afterward. This is one of the most common bits of nonfiction writing advice I dispense: "Open with the concrete example, not the abstract explanation!")

Theoretical background:

S. I. Hayakawa once gave this illustration of the "ladder of abstraction", and in particular, the difference between going *up* or *down:*

> "What is meant by the word red?"
> "It's a color."
> "What's a color?"
> "Why, it's a quality things have."
> "What's a quality?"

vs.

> "What is meant by the word red?"
> "Well, the next time you see some cars stopped at an intersection, look at the traffic light facing them. Also, you might go to the fire department and see how their trucks are painted."

"Red is a color" is moving *up* the ladder; "color" is a supercategory of red. All things which are red, have colors; but not all things which have colors, are red. And similarly, if you look at a specific firetruck, that firetruck is a red thing, but there are also many other red things which are not that firetruck.

What is true of one apple may not be true of another apple; suppose apple_1 weighs 100 grams and is slightly green in some places, and apple_2 weighs 200 grams and is entirely dark-red. You can say more truths about apple_2, like "apple_2 is dark red", then you can say that is true of *all* apples. (For more on this point see The Virtue of Narrowness.)

Thus, it may be easier to mentally picture "a firetruck" than "something red" - "firetruck" describes a narrower section of Thingspace, so you're less likely to get lost along the way.

S. I. Hayakawa called this the ladder of abstraction. I'm not sure if understanding the following section will really help with the skill of Being Specific, or help anyone construct exercises for the skill of being specific. But a better theoretical understanding does sometimes prove useful. So I will now digress to explain that abstraction isn't really a ladder, but a *lattice.*

Let's illustrate this using a classic example from the field of machine learning. Suppose that Days have three properties:

- Weather: {Sunny, Cloudy, Rainy}
- Temperature: {Cool, Hot}
- Timing: {Weekday, Weekend}

And suppose that we've been given some examples of Days on which it was good, or alternatively bad, to play tennis. For example, the Day {Sunny, Cool, Weekend} was good for playing tennis, but the day {Rainy, Hot, Weekday} was bad for playing tennis. A classic task in machine learning is to induct, from a set of pre-classified examples like these, a *rule* describing when it is good to play tennis.

Any proposed rule which can classify all days as good or bad is a *concept,* in the lingo of machine learning. "Sunny Days" is a concept; likewise "Sunny Cool Days", and "Days which are either Cool or Sunny". Each of these is a concept which classifies all 12 possible days either positively or negatively - instances or non-instances of the concept.

There are 2^{12} possible concepts over the 12 possible Days. Why so many? Because - for example - there's a concept which only includes the two Days {Sunny+Cool+Weekday} and {Cloudy+Cool+Weekend}}, but classifies all other Days as noninstances. This is a way of classifying all Days into instances or noninstances, hence a possible concept. It's not a *compact* concept, but it's a concept. Each Day can be classified either positively or negatively - one binary decision per Day - so 2^{12} possible concepts. (That's why induction is a difficult problem in machine learning.)

The concept "Sunny" is a superconcept of "Sunny and Cool"; it lies above it in the lattice of abstraction, since all days which are "Sunny and Cool" are "Sunny". "Sunny or Hot" is a supercategory of "Sunny". "Weekend" is neither a superconcept nor a subconcept of "Sunny".

Concepts form a directed lattice from *most general* to *most specific,* with "all Days" at the top (every Day classified as an instance) and "no Days" at the bottom (the concept which classifies every Day as a noninstance).

If you now go back to the problem of telling someone what "red" means, when you say "red is a color", then, even if the listener does happen to know what "color" means, you're still moving *upward in the lattice of abstraction.* When you said "color", you were talking about a concept that included all red things, but also many other things that were not red.

"Our software is providing the better connections for people" - the entrepreneur who said that might have had something specific in mind, or they might have just been bluffing or succumbing to wishful thinking. But they described it using an abstract statement so broad that it included Facebook, or Western Union back when they were sending telegrams. They might - though this is somewhat optimistic - they might have known themselves what they had in mind; *they* didn't think of Facebook; so they didn't realize how many other possibilities fit their words. This is a classic manifestation of the Illusion of Transparency, and it's why we have to keep telling people to navigate the lattice *downward.*

The skill of Being Specific is the skill of understanding how to *navigate the lattice of abstraction.* You can see why this would be a key element of cognition on a par with Bayes's Theorem or consequentialism.

And this is true in practice as well as theory. When I'm talking to anyone outside the local LW community, I find that a very large amount of my conversation involves repeatedly asking them to be more specific - and if you think that's just me being annoying, watch Paul Graham in the video.

A closely related skill is **concreteness,** which has to do with *nearness-to-sensory-experience* or *actionability.*

According to David Allen's "Getting Things Done", for your brain to stop thinking about an unfinished task, you must (1) know and trust that an external system will remind you to perform that task when it is time to perform it, and (2) have chosen the *next action taken* at a sufficiently concrete level that your brain is no longer trying to plan it out in the background. "Contact Luke about dispersing prize awards" is not a sufficiently concrete to-do; it leaves open the question of whether to phone or email, and what exactly to say. "Read through the comments, gather the LessWrong usernames of everyone who made a suggestion we tried or adopted, and email the list to Luke" is an action item I know how to perform straightforwardly, without my brain trying to plan it in the background. When you have a *trustworthy* external system to remind you of what to do, at the time you need to do it - so that the back of your mind isn't worrying about remembering to check the to-do list - and *all* to-do items have been concretized to the point of being executable without further background planning - then you have, in GTD parlance, "gotten to zero", a state of pure mental blissfulness in which your brain is not worrying about *anything* except what you're doing *right now.*

Similarly, for a statement like "Wulky Wilkinsen is a post-utopian" or "Earth gravity pulls at 9.8 meters per second squared" to be *falsifiable,* it must be *concretized* - rendered near-to-experience - to a sufficient degree that you can potentially *see* something and say "Oh, guess the hypothesis was wrong"; you must be able to have an experience which the concretized statement *constrains,* and which falsifies the theory if the experience is out-of-bounds.

Theoretically: If you imagine the universe as a huge directed graph of causes and effects - the Great Web of Causality - then "concreteness" is being near enough in the Web to either your *sensory inputs* or *motor outputs* that you can directly see the prediction unfold, or directly implement the plan, without much further thought.

"Be Specific" and "Be Concrete" could easily end up being the same unit - they're closely related - and we're happy to entertain exercises for Being Concrete, as well as Being Specific. Visualizing what your customer literally *sees* or *does* after navigating to your site, would've been a good first step toward being able to answer many of Paul Graham's questions.

A possible success criterion:

One question that we spent a lot of time discussing at CMR, was translating our sense of "specific enough" or "concrete enough" into a describable criterion. (Instead of just a wordless intuition for when something is "too abstract".)

There was an exchange in Paul Graham's office hours that went like this, while interviewing a startup that did metrics - analyzing pageviews, roughly - and the entrepreneur was having great trouble describing what they did that MixPanel didn't. It went on for a while. It was painful to watch.

> Paul: I don't get what the difference is. I *still* don't get what the difference is. What's the difference between you and MixPanel?

Entrepreneur: The difference is - when you have to supplement - they're a view company and we're a platform. That's what it comes down to. They're like a view, a reporting company. If you need something they don't have, a feature -

Harj: So what's an example of somewhere you'd use your thing over MixPanel? Can you give a use-case?

Entrepreneur: Yeah, I mean, we had revenue on day zero. There's a good reason for um... it's a start up, it's a series A company in the daily deals space. One we've signed a social game company to -

Harj: And why do they prefer your thing?

Paul: That wasn't what Harj was asking.

The problem (from the perspective of our present discussion) is that the Entrepreneur did not understand that Paul and Harj were repeatedly asking him to move downward on the ladder of abstraction. When the Entrepreneur said "We had revenue on day zero", he was trying to offer *confirmation* of the abstract statement "We can do things MixPanel can't", but Paul and Harj still had no idea what his startup *actually did.*[1]

A quick bit of theoretical background: There's an important difference, in the field of mathematical logic, between *models* and *axioms.* An axiom is something like "All kittens are cute", i.e. "All x: kitten(x)->cute(x)". A *model* is a particular universe of objects that includes {Obj #19834, kitten: T, cute: T, color: grey} and {Obj #19835, kitten: F, cute: F, color: striped}, and so on.

Correspondingly, in logical inference, there's a distinction between *model-checking* and *deduction.* Suppose you want to know whether it's true that all positive integers less than 5, when multiplied by 7, are less than 50. If you prove the general truth that all integers less than 5, times 7, are less than 35, by manipulating the axioms of multiplication and inequality, that's deduction. If you notice that the only positive integers less than 5 are just {1, 2, 3, 4} and enumerate their products {7, 14, 21, 28}, which are all less than 50, that's model-checking.

My hypothesis about what it means to be "specific enough" or "concrete enough" is that the picture painted is detailed enough to use in *model-checking* whatever points are being debated. Paul and Harj don't want to *trust* you when you state the abstract generalization, "We're better than MixPanel". They aren't even content with deducing support for this generalization from the further generalization, "We already have customers." They want a *picture* of something you do that MixPanel doesn't, which is detailed enough that they can *model-check* whether you have a competitive advantage.

Not to mention that Paul Graham is probably thinking about a number of other questions:

- How much would I pay for this product?
- Is this startup exciting enough that I would tweet about using it?
- How much resources will it take to develop these features further?

Paul Graham doesn't want you to say, "$50, yes, and twenty engineer-months". He wants a sufficiently specific picture of (a customer using) your product that he can arrive at his own answers by model-checking.

If Paul Graham is reading this, he's welcome to contradict my interpretation of what was going on in that particular session - but it did seem like a very nice concrete illustration.

That's my guess for what often constitutes "specific enough" - though I'm not sure that's the *only* thing that ever determines specific-enoughness.

[1]: The strange part was, near the end of that session, it started to look like this might be an interesting startup; that the Entrepreneur wasn't just bluffing. Their actual use-case was to let customers easily roll their own code to measure, e.g., the page-viewing behavior of only customers who'd bought more than $200 worth of stuff, which allegedly MixPanel wouldn't let you do. Which would've been a perfectly good answer if the Entrepreneur had given it *at the start of the session,* instead of the whole session being about Paul and Harj trying to get at that information.

Five-second-level skill:

The 5SL skill for this problem requires:

- Trigger: Recognizing when your words or thoughts are *too abstract.*
- Action: Moving downward in the abstraction lattice, or moving nearer to sense input or motor output; being able to render your thoughts more specific or more concrete.

Both of these are targetable for exercises.

Pain points & Pluses:

• You want Paul Graham to believe your startup is better than MixPanel. So you say, "My startup is better than MixPanel" - just produce the pure abstract conclusion you want Paul Graham to arrive at. You keep trying to convince Paul Graham of this statement, saying that you have customers or that you have venture capital, but never actually move *downward* to the level where Paul Graham could arrive at this conclusion by model-checking.

• You want to describe what your software does, so you say it makes connections between people. *You* have something specific in mind, but the words coming out of your mouth are so general that - although *you're* not thinking of those other cases - they could apply equally well to Facebook or telegraph lines. Paul Graham has no idea at all what you're trying to describe and is giving you blank looks.

• The worse version - and the reason why Paul Graham doesn't just trust you, even if he thinks you're honest - is the case where you *yourself* want to believe your startup is better than Facebook, but you can't think of any *specific* thing your startup does better than Facebook, so you think of other abstract generalizations that seem to support the conclusion, like "We have smarter people" or "We got more funding earlier." Where fuzzy thinking is motivated, overly abstract thinking is motivated.

• Abstract words can also avoid *emotion*. George Orwell: "Defenceless villages are bombarded from the air, the inhabitants driven out into the countryside, the cattle machine-gunned, the huts set on fire with incendiary bullets: this is called *pacification*." Or contrast "Humanity is awful, it'd be better for the planet if we all died" to "Everyone including my little sister is awful, we'd be better off if everyone

died including her." To feel sympathy, we need enough concrete detail that our emotions can model-check the picture and be activated.

• Cognitive-behavioral therapy is the big *experimentally supported* version of therapy, for anyone not aware of this, bearing very little resemblance to anything Freudian. CBT talks about using requests for specific details to interrupt thoughts looping around vague but affectively laden centers, like "I am a good husband", "I am a bad husband", or "my roommate is a slob". How are you a good husband? How are you a bad husband? Which specific feature of your roommate are you objecting to? Taboo the emotionally valent *word* at the center, like "slob", and replace it with something that's specific enough to be testable, or concrete enough to be acted upon.

•• Contrast also "It bothers me when you leave soda cans on the table" vs. "You're such a slob, stop being such a slob." Or contrast: "I'm upset" -> "I'm upset because I think the other person is looking down on me" -> "I'm upset because the person's tone of voice sounds like people who looked down on me in high school". This is related to the incredibly important skill, *search for the historical causes of your thoughts, rather than their justifications.*

• Focusing on the specific details of a concrete example, instead of repeating a word or arguing about a category, can interrupt Sneaking in Connotations and Arguing By Definition.

• All the failures of *concreteness* warned against in the Mysterious Answers sequence, where you go on and on about how Wulky Wilkinsen is a post-utopian without ever once asking or imagining how the world ought to look, and what you yourself should experience, if that were true or alternatively false.

• Visualizing specific examples often improves quality of thought in general - we're often smarter when we're using both model-checking and deduction, visualizing a picture of what we're supposed to be reasoning about, constantly checking our deductive steps against some specific model those deductions are supposed to be true about. Saith Richard Feynman:

> I had a scheme, which I still use today when somebody is explaining something that I'm trying to understand: I keep making up examples. For instance, the mathematicians would come in with a terrific theorem, and they're all excited. As they're telling me the conditions of the theorem, I construct something which fits all the conditions. You know, you have a set (one ball) - disjoint (two halls). Then the balls turn colors, grow hairs, or whatever, in my head as they put more conditions on. Finally they state the theorem, which is some dumb thing about the ball which isn't true for my hairy green ball thing, so I say, "False!"
>
> If it's true, they get all excited, and I let them go on for a while. Then I point out my counterexample.
>
> "Oh. We forgot to tell you that it's Class 2 Hausdorff homomorphic."
>
> "Well, then," I say, "It's trivial! It's trivial!"

• Being specific helps notice and call bluffs, should you be mischievously inclined.

> "Beware, demon!" he intoned hollowly. "I am not without defenses."
> "Oh yeah? Name three."
> -- Robert Asprin, Another Fine Myth

> Wannabe executive: "I will improve communications between employees and management."
> Me: "Can you give me a specific example of how you would do that?"

Known exercises for this skill:

- Rationalist Taboo

In our previous Rationality Camps, Anna found that her attempt to teach a unit on "Being Specific" didn't seem to work. Her central exercise was picking a category and asking people to name examples.

This isn't to say that the Camps were unsuccessful at teaching the skill. Attendees picked it up, not from the explicit unit, but from all the instructors having to repeatedly ask the attendees to be more specific, and then having to ask them again, while being specific themselves, until the attendees picked up the rhythm by example and feedback.

Given our present teaching technology, this skill seems *transmissible* from master to apprentice, but not yet *replicable* by exercises. That's why we're turning it over to you.

LessWrong downtime 2012-03-26, and site speed

Our investigation into last week's LW downtime is complete: here (Google Docs).

Executive summary:

We failed to update our AWS configuration after changes at Amazon, which caused a cycle of servers being spawned then killed before they could properly boot. Our automated testing should have notified us of this failure immediately, but included a predictable failure mode (identified by us last year but not fixed). We became aware of the downtime when I checked my email and worked on it until it was resolved.

I personally feel very bad about our multiple failures leading to this incident.

ref. the last time I did this to you: http://lesswrong.com/lw/29v/lesswrong_downtime_20100511_and_other_recent/

Actions:

1. We have reconfigured AWS and the tools we use to communicate with it to avoid this failure in the future.
2. Improvements to our automated site testing system (Nagios) are underway (expected to be live before 2012-04-13 - these tests will detect greater-than-X-failures-from-Y-trials, rather than the current detect zero-successes-from-Z-trials).
3. We have changed our staffing in part in recognition that some systems (including this one) had been allowed to fall out of date, and allocated a developer to review our system administration project planning.

Further actions - site speed:

We're unhappy with the site's speed. We plan on spending some time next week doing what we can to improve it.

(If you upvote this post, please downvote my "Karma sink" comment below - I would prefer not to earn karma from an event like this.)

Cryonics without freezers: resurrection possibilities in a Big World

And fear not lest Existence closing your
Account, should lose, or know the type no more;
The Eternal Saki from the Bowl has pour'd
Millions of Bubbles like us, and will pour.

When You and I behind the Veil are past,
Oh, but the long long while the World shall last,
Which of our Coming and Departure heeds
As much as Ocean of a pebble-cast.

-- Omar Khayyam, Rubaiyat

A CONSEQUENTIALIST VIEW OF IDENTITY

The typical argument for cryonics says that if we can preserve brain data, one day we may be able to recreate a functioning brain and bring the dead back to life.

The typical argument against cryonics says that even if we could do that, the recreation wouldn't be "you". It would be someone who thinks and acts exactly like you.

The typical response to the typical argument against cryonics says that identity isn't in specific atoms, so it's probably in algorithms, and the recreation would have the same mental algorithms as you and so be you. The gap in consciousness of however many centuries is no more significant than the gap in consciousness between going to bed at night and waking up in the morning, or the gap between going into a coma and coming out of one.

We can call this a "consequentialist" view of identity, because it's a lot like the consequentialist views of morality. Whether a person is "me" isn't a function of how we got to that person, but only of where that person is right now: that is, how similar that person's thoughts and actions are to my own. It doesn't matter if we got to him by having me go to sleep and wake up as him, or got to him by having aliens disassemble my brain and then simulate it on a cellular automaton. If he thinks like me, he's me.

A corollary of the consequentialist view of identity says that if someone wants to create fifty perfect copies of me, all fifty will "be me" in whatever sense that means something.

GRADATIONS OF IDENTITY

An argument against cryonics I have never heard, but which must exist somewhere, says that even the best human technology is imperfect, and likely a few atoms here

and there - or even a few entire neurons - will end up out of place. Therefore, the recreation will not be you, but someone very very similar to you.

And the response to this argument is "Who cares?" If by "me" you mean Yvain as of 10:20 PM 4th April 2012, then even Yvain as of 10:30 is going to have some serious differences at the atomic scale. Since I don't consider myself a different person every ten minutes, I shouldn't consider myself a different person if the resurrection-machine misplaces a few cells here or there.

But this is a slippery slope. If my recreation is exactly like me except for one neuron, is he the same person? Signs point to yes. What about five neurons? Five million? Or on a functional level, what if he blinked at exactly one point where I would not have done so? What if he prefers a different flavor of ice cream? What if he has exactly the same memories as I do, except for the outcome of one first-grade spelling bee I haven't thought about in years anyway? What if he is a Hindu fundamentalist?

If we're going to take a consequentialist view of identity, then my continued ability to identify with myself even if I naturally switch ice cream preferences suggests I should identify with a botched resurrection who also switches ice cream preferences. The only solution here that really makes sense is to view identity in shades of gray instead of black-and-white. An exact clone is *more* me than a clone with different ice cream preferences, who is *more* me than a clone who is a Hindu fundamentalist, who is *more* me than LeBron James is.

BIG WORLDS

There are various theories lumped together under the title "big world".

The simplest is the theory that the universe (or multiverse) is Very Very Big. Although the universe is probably only 15 billion years old, which means the *visible* universe is only 30 billion light years in size, inflation allows the *entire* universe to get around the speed of light restriction; it could be very large or possibly infinite. I don't have the numbers available, but I remember a back of the envelope calculation being posted on Less Wrong once about exactly how big the universe would have to be to contain repeating patches of about the size of the Earth. That is, just as the first ten digits of pi, 3141592653, must repeat somewhere else in pi because pi is infinite and patternless, and just as I would believe this with high probability even if pi were not infinite but just very very large, so the arrangement of atoms that make up Earth would recur in an infinite or very very large universe. This arrangement would obviously include you, exactly as you are now. A much larger class of Earth-sized patches would include slightly different versions of you like the one with different ice cream preferences. This would also work, as Omar Khayyam mentioned in the quote at the top, if the universe were to last forever or a very very long time.

The second type of "big world" is the one posited by the Many Worlds theory of quantum mechanics, in which each quantum event causes the Universe to split into several branches. Because quantum events determine larger-level events, and because each branch continues branching, some these branches could be similar to our universe but with observable macro-scale differences. For example, there might be a branch in which you are the President of the United States, or the Pope, or died as an infant. Although this sounds like a silly popular science version of the principle, I don't *think* it's unfair or incorrect.

The third type of "big world" is modal realism: the belief that all possible worlds exist,

maybe in proportion to their simplicity (whatever that means). We notice the existence of our own world only for indexical reasons: that is, just as there are many countries, but when I look around me I only see my own; so there are many possibilities, but when I look around me I only see my own. If this is true, it is not only possible but certain that there is a world where I am Pope and so on.

There are other types of "big worlds" that I won't get into here, but if any type at all is correct, then there should be very many copies of me or people very much like me running around.

CRYONICS WITHOUT FREEZERS

Cryonicists say that if you freeze your brain, you may experience "waking up" a few centuries later when someone uses the brain to create a perfect copy of you.

But whether or not you freeze your brain, a Big World is creating perfect copies of you all the time. The consequentialist view of identity says that your causal connection with these copies is unnecessary for them to be you. So why should a copy of you created by a far-future cryonicist with access to your brain be better able to "resurrect" you than a copy of you that comes to exist for some other reason?

For example, suppose I choose not to sign up for cryonics, have a sudden heart attack, and die in my sleep. Somewhere in a Big World, there is someone exactly like me except that they didn't have the heart attack and they wake up healthy the next morning.

The cryonicists believe that having a healthy copy of you come into existence after you die is sufficient for you to "wake up" as that copy. So why wouldn't I "wake up" as the healthy, heart-attack-free version of me in the universe next door?

Or: suppose that a Friendly AI fills a human-sized three-dimensional grid with atoms, using a quantum dice to determine which atom occupies each "pixel" in the grid. This splits the universe into as many branches as there are possible permutations of the grid (presumably *a lot*) and in one of those branches, the AI's experiment creates a perfect copy of me at the moment of my death, except healthy. If creating a perfect copy of me causes my "resurrection", then that AI has just resurrected me as surely as cryonics would have.

The only downside I can see here is that I have less measure (meaning I exist in a lower proportion of worlds) than if I had signed up for cryonics directly. This might be a problem if I think that my existence benefits others - but I don't think I should be concerned for my own sake. Right now I don't go to bed at night weeping that my father only met my mother through a series of unlikely events and so most universes probably don't contain me; I'm not sure why I should do so after having been resurrected in the far future.

RESURRECTION AS SOMEONE ELSE

What if the speculative theories involved in Big Worlds all turn out to be false? All hope is still not lost.

Above I wrote:

> An exact clone is *more* me than a clone with different ice cream preferences, who is *more* me than a clone who is a Hindu fundamentalist, who is *more* me than LeBron James is.

I used LeBron James because from what I know about him, he's quite different from me. But what if I had used someone else? One thing I learned upon discovering Less Wrong is that I had previously underestimated just how many people out there are *really similar to me*, even down to weird interests, personality quirks, and sense of humor. So let's take the person living in 2050 who is most similar to me now. I can think of several people on this site alone who would make a pretty impressive lower bound on how similar the most similar person to me would have to be.

In what way is this person waking up on the morning of January 1 2050 equivalent to me being sort of resurrected? What if this person is more similar to Yvain(2012) than Yvain(1995) is? What if I signed up for cryonics, died tomorrow, and was resurrected in 2050 by a process about as lossy as the difference between me and this person?

SUMMARY

Personal identity remains confusing. But some of the assumptions cryonicists make are, in certain situations, sufficient to guarantee personal survival after death without cryonics.

Hofstadter's Superrationality

Possibly the main and original inspiration for Yudkowsky's various musings on what advanced game theories should do (eg. cooperate in the Prisoner's Dilemma) is a set of essays penned by Douglas Hofstadter (of *Godel, Escher, Bach*) 1983. Unfortunately, they were not online and only available as part of a dead-tree collection. This is unfortunate. Fortunately the collection is available through the usual pirates as a scan, and I took the liberty of transcribing by hand the relevant essays with images, correcting errors, annotating with links, etc: http://www.gwern.net/docs/1985-hofstadter

The 3 essays:

1. discuss the Prisoner's dilemma, the misfortune of defection, what sort of cooperative reasoning would maximize returns in a souped-up Prisoner's dilemma, and then offers a public contest
2. then we learn the results of the contest, and a discussion of ecology and the tragedy of the commons
3. finally, Hofstadter gives an extended parable about cooperation in the face of nuclear warfare; it is fortunate for us that it applies to most existential threats as well

I hope you find them educational. I am not 100% confident of the math transcriptions since the original ebook messed some of them up; if you find any apparent mistakes or typos, please leave comments.

Why I Moved from AI to Neuroscience, or: Uploading Worms

This post is shameless self-promotion, but I'm told that's probably okay in the Discussion section. For context, as some of you are aware, I'm aiming to model C. elegans based on systematic high-throughput experiments - that is, to upload a worm. I'm still working on course requirements and lab training at Harvard's Biophysics Ph.D. program, but this remains the plan for my thesis.

Last semester I gave this lecture to Marvin Minsky's AI class, because Marvin professes disdain for everything neuroscience, and I wanted to give his students—and him—a fair perspective of how basic neuroscience might be changing for the better, and seems a particularly exciting field to be in right about now. The lecture is about 22 minutes long, followed by over an hour of questions and answers, which cover a lot of the memespace that surrounds this concept. Afterward, several students reported to me that their understanding of neuroscience was transformed.

I only just now got to encoding and uploading this recording; I believe that many of the topics covered could be of interest to the LW community (especially those with a background in AI and an interest in brains), perhaps worthy of discussion, and I hope you agree.

Let's create a market for cryonics

My uncle works in insurance. I recently mentioned that I'm planning to sign up for cryonics.

"That's amazing," he said. "Convincing a young person to buy life insurance? That has to be the greatest scam ever."

I took the comment lightly, not caring to argue about it. But it got me thinking - *couldn't* cryonics be a great opportunity for insurance companies to make a bunch of money?

Consider:

1. Were there a much stronger *demand* for cryonics, cryonics organizations would flourish through competition, outside investment, and internal reinvestment. Costs would likely fall, and this would be good for cryonicists in general.
2. If cryonics organizations flourish, this increases the probability of cryonics working. I can think of a bunch of ways in which this could happen; perhaps, for example, it would encourage the creation of safety nets whereby the failure of individual companies doesn't result in anyone getting thawed. It would increase R&D on both perfusion and revivification, encourage entrepreneurs to explore new related business models, etcetera.
3. Increasing the demand for cryonics increases the demand for life insurance policies; thus insurance companies have a strong incentive to increase the demand for cryonics. Many large insurance companies would like nothing more than to usher in a generation of young people that want to buy life insurance.[1]
4. The demand for cryonics could be increased by an insightful marketing campaign by an excellent marketing agency with an enormous budget... like those used by big insurance companies.[2] A quick Googling says that ad spending by insurance companies exceeded $4.15 billion in 2009.

Almost a year ago, Strange7 suggested that cryonics organizations could run this kind of marketing campaign. I think he's wrong - there's no way CI or Alcor have the money. But the biggest insurance companies *do* have the money, and I'd be *shocked* if these companies or their agencies aren't already dumping all kinds of money into market research.

What would doing this require?

1. That an open-minded person in the insurance industry who is in the position to direct this kind of funding *exists*. I don't have a sense of how likely this is.
2. That we can locate/get an audience with the person from step 1. I think research and networking could get this done, especially if the higher-status among us are interested.
3. That we can find someone who is capable and willing to explain this clearly and convincingly to the person from step 1. I'm not sure it would be *that* difficult. In the startup world, strangers convince strangers to speculatively spend millions of dollars every week. Hell, I'll do it.

I want to live in a world where cryonics ads air on TV just as often as ads for everything else people spend money on. I really can see an insurance company

owning this project - if they can a) successfully revamp the image of cryonics and b) become known as the household name for it when the market gets big, they will make lots of money.

What do you think? Where has my reasoning failed? Does anyone here know anyone powerful in insurance?

Lastly, taking a cue from ciphergoth: this is not the place to rehash all the old arguments about cryonics. I'm asking about a very specific idea about marketing and life insurance, not requesting commentary on cryonics itself. Thanks!

---

[1] Perhaps modeling the potential size of the market would offer insight here. If it turns out that this idea is not insane, I'll find a way to make it happen. I could use your help.

[2] Consider what happened with diamonds in the 1900s:

> ... N. W. Ayer suggested that through a well-orchestrated advertising and public-relations campaign it could have a significant impact on the "social attitudes of the public at large and thereby channel American spending toward larger and more expensive diamonds instead of "competitive luxuries." Specifically, the Ayer study stressed the need to strengthen the association in the public's mind of diamonds with romance. Since "young men buy over 90% of all engagement rings" it would be crucial to inculcate in them the idea that diamonds were a gift of love: the larger and finer the diamond, the greater the expression of love. Similarly, young women had to be encouraged to view diamonds as an integral part of any romantic courtship.

Peter Singer and Tyler Cowen transcript

In March 2009, Tyler Cowen (blog) interviewed Peter Singer about morality, giving, and how we can most improve the world. They are both thinkers I respect a lot, and I was excited to read their debate. Unfortunately the interview was available only as a video. I wanted a transcript, so I made one:

Cowen: This is Tyler Cowen of George Mason University. I'm doing a BloggingHeads with Peter Singer, the world-famous philosopher from Princeton. This is a forum on Peter's latest book, which he'll start off by telling you a bit about.

Singer: Hi. The book's called "The Life You Can Save: Acting Now To End World Poverty". It begins with an argument that I've used many times in articles about a child drowning in a pond, and suggests that if you saw a child drowning in a pond that you would jump in and save that child, and you think that is what you ought to do, even if it meant that you ruined an expensive pair of shoes that you were wearing.

From there I pull back to saying "what does this mean about the problem of world poverty, given that there are, according to Unicef, ten million children dying of avoidable poverty-related causes every year?" We could save some of them, and probably it wouldn't cost us much more than the cost of an expensive pair of shoes if we find an effective aid agency that is doing something to combat the causes of world poverty, or perhaps to combat the deaths of children from simple conditions like diarrhea or measles, conditions that are not that hard to prevent or to cure. We could probably save a life for the cost of a pair of shoes. So why don't we? What's the problem here? Why do we think it's ok to live a comfortable, even luxurious, life while children are dying? In the book I explore various objections to that view, I don't find any of them really convincing. I look at some of the psychological barriers to giving, and I acknowledge that they are problems. And I consider also some of the objections to aid and questions raised by economists as to whether aid really works. In the end I come to a proposal by which I want to change the culture of giving.

The aim of the book in a sense is to get us to internalize the view that not to do anything for those living in poverty, when we are living in luxury and abundance, is ethically wrong, that it's not just not a nice thing to do but that a part of living an ethically decent life is at least to do something significant for the poor. The book ends with a chapter in which I propose a realistic standard, which I think most people in the affluent world could meet without great hardship. It involves giving 1% of your income if you're in the bottom 90% of US taxpayers, scaling up through 5% and 10% and even more as you get into the top 10%, the top 5%, the top 1% of US taxpayers. But at no point is the scale I'm proposing what I believe is an excessively burdensome one. I've set up a website, thelifeyoucansave.com that people can go to in order to publicly pledge that they will meet this scale, because I think if people will do it publicly, that in itself will encourage other people to do it and, hopefully, the idea will spread.

Cowen: Thank you, Peter. Let me first stress: I agree with most of what's in your book; I think we all could give more and should give more. It would be good for other people and it would be good for ourselves. But let me start off the dialogue by mentioning a few points where I don't completely agree with you. One thing that struck me about the book was some of the omissions.

Immigration as an Anti-Poverty Program

Cowen: For instance, in my view, what is by far the best anti-poverty program, the only one that's really been shown to work, and that's what's called "immigration". I don't even see the word "immigration" in your book's index. So why don't we spend a lot more resources allowing immigration, supporting immigration, lobbying for immigration? This raises people's incomes very dramatically, it's sustainable, for the most part it's also good for us. Why not make that the centerpiece of an anti-poverty platform?

Singer: That's an interesting point, Tyler. I suppose, one question I'd like to ask is: is it sustainable? Isn't it the case that if we take, as immigrants, the people who are the most enterprising, perhaps, of the poor countries that we're still going to leave those countries in poverty, and their populations may continue to rise, and eventually, even if we keep taking immigrants, we will reach a capacity where we're starting to strain our own country?

Cowen: There's two separate issues: one is "brain drain" from the third world. I think here's a lot of research by [Michael Clemens], showing that it's not a problem, that third world countries that have even somewhat functional institutions tend to benefit by sending people to other countries. India's a good example: a lot of Indians return to India and start businesses, or they send money back home. Mexico is another example. Maybe North Korea is somewhat different, but for the most part immigration seems to benefit both countries.

I don't think we could have open borders; I don't think we could have unlimited immigration, but we're both sitting here in the United States and it hardly seems to me that we're at the breaking point. Immigrants would benefit much more: their wages would rise by a factor of twenty or more, and there would be perhaps some costs to us, but in a cost-benefit sense it seems far, far more effective than sending them money. Do you agree?

Singer: I must admit that I haven't thought a lot about immigration as a way of dealing with world poverty. Obviously, from what you're saying, I should be thinking more about it, but I can't really say whether I agree until I have thought more about it.

Changing Institutions: Greater Tax Break for True Charity

Cowen: Let me try another question along related lines. I think one general way in which I think about your book differently than you do, is that you think more about giving. I'm a big advocate of giving, I've written a whole book myself on philanthropy, but I think somewhat more in terms of changing institutions.

So another thing we might consider doing, along the lines of what you advocate, is to increase the tax benefits of giving. Right now, if you're itemizing deductions and you give $1, you deduct $1 from your taxes. But it wouldn't be very difficult to make it the case that for certain kinds of giving you could deduct $1.10 from your taxes or $1.20. Would you favor this kind of reform?

Singer: I might favor that, if giving were defined more narrowly than we do, in the US anyway, because I know I can deduct $1 from my taxes whether I give to Oxfam America, which I think is an effective organization fighting world poverty, or if I give to the Met so they can buy yet another painting to add to the already super-abundant collection of paintings they have. I don't see why the taxpayer should subsidize me if I decide I want to give to the Met but sure, if I'm giving to Oxfam I think that would be good.

Cowen: So, in other words, you favor a kind of tax cut as a way to help the world's poor. That, in this country, if targeted properly, tax policy, in essence cutting the taxes of rich people, is one of the very best ways to help the world's poor. Would you sign on to that?

Singer: I'm not quite sure why it is ... you seem to have leapt a little from what I was saying and I haven't followed the leap as to why cutting taxes for the rich would be one of the most effective ways of helping the poor. Can you explain that a little more?

Cowen: If we give a greater tax break to charitable donations, and here I mean only true charity, not say a fancy art museum, disproportionately this will benefit wealthy people. Wealthy people have a lot of money. In essence you're cutting their taxes. They're giving more, they may not have a higher level of consumption, but would you be willing to raise your hand and say "I, Peter Singer, think that cutting taxes on the US wealthy is in fact one of the very best things we could do for the world's poor, if we do it the right way"? Yes or no?

Singer: Yes, if the tax break only goes to those of the wealthy who are giving to organizations that are effectively helping the poor, I'll raise my hand to that.

Cowen: OK; I'm glad to hear that.

Millennium Villages Skepticism

Cowen: Let me focus on another point of difference between us in the book. I think when it comes to the effectiveness of aid, I'm not a total skeptic on the effectiveness of aid but I think I'm more skeptical than you are. In a number of places you site the work of Jeffrey Sachs. Now my view of Sachs is that his projects are actually doing individual human beings a lot of good but the return on investment I don't think is that high. I think he's improving the health of a lot of people but I don't think he's going to raise any villages, much less countries or continents, out of poverty. Given that my view is that the rate of return on this investment is much lower, and I think that the economics profession as a whole agrees with me, not with Sachs, this to me suggests that to really make a dent in world poverty we would have to give much more than 5% of our incomes, even more than 10%, that we're simply at a point where we can do some good, but that to abolish poverty we would have to engage in a very dramatic redistribution. What's your view on this?

Singer: Firstly, I think, as for whether Sachs is really going to succeed in raising

villages out of poverty, I think the data isn't in yet. The Millennium Villages project which he's working on has only been going a few years, I think we need to give it maybe another five years to see whether it's working. That's more or less what he's said. He hopes that the aid will be short term, that the villages will become self-sustaining, the improvements will last, they'll be out of the poverty trap. If that hasn't happened at the end of another five years I'm going to agree with you that we're going to need more, but I think it's really too early to call the result on that one.

Cowen: Take the overall opinion of economists, which is again that Sachs's projects can do good, people in those villages might be better off, but if you're in the middle of, say, a totally corrupt African country which is not democratic, which maybe has been fighting wars, which has an absolutely horrible infrastructure, which has a bureaucracy, a kleptocracy, massive problems, lack of literacy, that maybe you could eliminate infectious diseases or malaria within that village. People will be better off, it's worth doing, but at the end of the day is there really any reason to think, given the last 300 years of thinking and writing on development and economic history that this will at all cure poverty? Doesn't it just mean you'll have poor people without malaria, which is better than poor people with malaria, but they're still essentially poor people?

Singer: If the governments and the situation is as bad as you describe, you're probably right, but of course not all countries ... you describe pretty much a worst case scenario. I think there are a lot of countries where there are poor people which do not have governments which are as bad as you painted. I think in those countries we can hope that people actually will lift themselves out of poverty and I think that's what we need to try to do. Now, you may be right that that's still going to leave poor people in countries that are as bad as you describe, and there is a real question then as to how much we can do to help them, whether giving more will really be enough to help them or because of those governments in those situations there's really nothing much we can do. That will be the dilemma. But I don't think we've got to that point yet because we've not really worked out what we can do for people in the countries where the governments aren't so bad.

Chinese Reforms

Cowen: I think you and I are both looking for what are the most highly leveraged ways we can reduce poverty in this world.

Singer: Uh huh

Cowen: If I ask myself, historically, what has been the most successful anti-poverty program in the last century, I look at Communist China, and I would say that the reforms, starting in the late 1970s, have taken at least 300M-400M people, and probably more, and taken them from extreme poverty, perhaps starvation, to a situation where a lot of them live quite well or at least have some kind of tolerable lower middle class existence. I think that property rights and institutional reforms are they key to fighting poverty. China during that period, the aid it received didn't matter much. It doesn't mean we shouldn't give aid, I'm all for aid, but isn't the big leveraged investment here changing and improving institutions and not giving money?

Singer: I do that that's a really important thing when we can do it. The question is,

can we do it? Obviously the Chinese reforms that you refer to really were internally driven, I don't think they were a result of things the West did, unless you talk about the entire global economic system, which China clearly wanted to participate in. So the question is how can we be effective in producing those sorts of changes? In some countries we can come in and help, say countries recovering from civil war, and give some help in establishing good institutions, but I'm not sure what ideas you have about what's a good way to bring about that kind of reform in these countries that will lead everywhere to the sorts of benefits that you refer to in China.

Cowen: In countries like China in a way it's internally driven. It's not that anyone successfully pressured them, but in another way I think it's highly externally driven, that the Chinese, Taiwanese, Koreans, other countries followed the example of Japan They saw that Japan worked. They saw that an Asian country could rise to moderate wealth or even riches and at some point they decided to copy this in their own way. If we look at Japan, Japan copied the west, so maybe one of the very best most important things we can do is just ourselves be a beacon of progress: be humane, be tolerant, respect others, be wealthy and just show that it's possible. We shouldn't think of that as a substitute for aid, but maybe that's actually our number one priority. Does that make sense to you?

Singer: That makes sense. I don't know that we have to strive to be more wealthy than we are--well, maybe just right at this moment we need to strive to get back to being as wealthy as we were a year ago perhaps. But I think we are setting that example, undoubtedly. We are showing countries what can be done with reasonably good government, open economies, and I do hope that other countries will follow that. But maybe not all countries can do it. I think that Paul Collier argues in his book that it's going to be difficult for some African countries to get into this game now. There are reasons why it's going to be hard for them to compete with countries that have established positions, have developed markets, have low labor costs. It's not clear to me that this is going to be a path that every poor country can follow.

Military Intervention

Cowen: You mentioned Paul Collier. I found his book very interesting. One argument he makes--I would say I'm not, myself, convinced but I'm curious to hear what you think--is that we could do the world a great deal of good by selective military interventions. So take the case of Darfur. A large number of people are suffering, dying. Collier says, or implies, or at least opens the possibility, that we, the United States, the UN, whoever, should just move in and in military terms do something about this. It is again a topic that is not prominent in your book, but it seems that if it can work it's highly leveraged, more leveraged than giving away money. I'm curious as to your views on that.

Singer: I did discuss humanitarian intervention in my earlier book One World and I do support it under the right circumstances. I think, though, we do have to be pretty clear about defining it properly and trying to get support for it. Maybe it would work in Darfur. I think Darfur is quite a large area, relatively thinly populated, and it might take a lot of resources to really protect the people in Darfur. There are underlying issues, too, perhaps about climate change, even, that are causing scarcity in Darfur. But isn't possible, I mean I think

that Zimbabwe would be another possibility, though maybe just now with changes in the political system you wouldn't want to do it just now, you'd want to see how that played out for a while. But certainly a year ago you might well have thought that if the South Africans could be persuaded to move in and remove Mugabe that would be a good thing to do. That would have been better, I think, than having a white former colonial power come in, that obviously would have evoked a lot of echos of returning to a past that Zimbabweans don't want. But I'm not, in principle opposed to military intervention, I just think we have to be very very careful about the circumstances in which we do it, because obviously it can trigger a lot of violence and bloodshed and produce results that are the opposite of what you and I would both want.

Colonialism

Cowen: Do you think the end of colonialism was a good thing or a bad thing for Africa?

Singer: That's a really difficult question. I think, clearly, there were lots of bad things about colonialism, but you would have to say that some countries were definitely better administered and that some people's lives, although they may have had some sort of humiliation, perhaps through not being independent, being ruled by people of a different race, in some ways they were better. It's hard, really, to draw that balance sheet. Independence has certainly not been the unmitigated blessing that people thought it would be at the time.

Cowen: Let's say we have the premise, that with colonialism there would not have been wars between African nations. It's not the case that a British ruled colony would have attacked a French colony, for instance. It's highly unlikely. So given just that millions have perished from wars alone, wouldn't the Utilitarian view, if you're going to take one, suggest that colonialism was essentially a good idea for Africa, it was a shame that we got rid of it, and that the continent would have been better off under foreign rule, European foreign rule.

Singer: I don't think we can be so sure that it would have continued to be peaceful. After all we did have militant resistance movements, we had the Mau Mau in Kenya, for example. We had other militant resistance movements. It may simply have been that the fact of white rule would have provoked not one colony going to war against another but civil war within some of those countries. If what you're asking is would colonialism, had it been accepted by the people there, without military conflict, would that have been better than some of the consequences we've had in some of these countries, you would have to say undoubtedly yes. But we can't go back and wind back the clock and say "how would it have been if" because we don't really know whether that relative stability and peace would have lasted.

Cowen: If we compare the Mau Mau, say, to the wars in Kenya and Rwanda, it seems unlikely that rebellions against colonial governments would have reached that scope, especially if England, France, other countries, would have been willing to spend more money to create some tolerable form of order. My guess is you would have had a fair number of rebellions but it's highly highly

unlikely it would compare to the kind of virtual holocausts we've had in Africa as it stands.

Aid without stable government

Singer: I certainly agree that if you look at what's been happening in the Congo, just as one example, or countries like Sierra Leone or Liberia, yes, you could certainly think that it might have been better for those countries.

Cowen: Would you say that Zimbabwe is one example of a country where just giving it money through aid is unlikely to work?

Singer: At present, unless the government changes quite dramatically. Again, as you were saying before, there might be specific things we can do: we may be able to help particular people who have disease or are hungry, but I agree, in the present conditions it's unlikely to lift people out of poverty on any kind of large scale.

Cowen: Let's take a country like Madagascar, which as recently as two or three years ago was touted by the Bush administration, and I don't just mean Republicans, it was touted by many people, as being a kind of model for Africa. Here's a country were we could give a lot of aid, the aid would go to some good purpose, we're making progress, and now Madagascar seems to be in the midst of a civil war and the polity is collapsing, the economy is doing very poorly. How many countries in Africa do you think are there where aid works? Where do you draw the line? What in your opinion is the marginal country that is hopeless?

Singer: Look, I haven't got a list of African countries like that, I must admit. I think there are some countries where things seem to work, and that's not to say I could name a country and say, well, Mozambique, that aid programs have made a positive difference, or Sierra Leone. Maybe in a month there'll be a coup and you'll be able to tell me that I was wrong. I can't see the future. But there are countries where I think aid has worked, ones where it hasn't worked. I haven't got a rank ordering and I don't have a cutoff line where that is, I'm sorry I'm just not sufficiently expert on African politics and conditions to do that.

Genetically modifying ourselves to be more moral

Cowen: Let's try some philosophical questions. You're a philosopher, and I've been very influenced by your writings on personal obligation. Apart from the practical issue that we can give some money and have it do good, there's a deeper philosophical question of how far those obligations extend, to give money to other people. Is it a nice thing we could do, or are we actually morally required to do so? What I see in your book is a tendency to say something like "people, whether we like it or not, will be more committed to their own life projects than to giving money to others and we need to work within that constraint". I think we would both agree with that, but when we get to the deeper human nature, or do you feel it represents a human imperfection? If we could somehow question of "do we in fact like that fact?", is that a fact you're comfortable with about human nature? If we could

imagine an alternative world, where people were, say, only 30% as committed to their personal projects as are the people we know, say the world is more like, in some ways, an ant colony, people are committed to the greater good of the species. Would that be a positive change in human nature or a negative change?

Singer: Of course, if you have the image of an ant colony everyone's going to say "that's horrible, that's negative", but I think that's a pejorative image for what you're really asking ...

Cowen: No, no, I don't mean a colony in a negative sense. People would cooperate more, ants aren't very bright, we would do an ant colony much better than the ants do. ...

Singer: But we'd also be thinking differently, right? What people don't like about ant colonies is ants don't think for themselves. What I would like is a society in which people thought for themselves and voluntarily decided that one of the most satisfying and fulfilling things they could do would be to put more of their effort and more of their energy into helping people elsewhere in need. If that's the question you're asking, then yes, I think it would be a better world if people were readier to make those concerns their own projects.

Cowen: Let's say genetic engineering is possible, which is now not so far off on the major scale, and your daughter were having a daughter, and she asked you "daddy, should I program my daughter so that she's willing to sell her baby and take the money and send it to Haitians to save ten babies in Haiti". Would you recommend to her "yes, you should program the genes of your baby so she's that way"?

Singer: So she's going to sell her baby? What's going to happen to the baby?

Cowen: She's going to sell it to some wealth white couple that's infertile, they live in the Pacific Northwest, they'll take fine care of it, she'll receive $1M and save, say, 30 lives in Haiti. You've recommended that your granddaughter be programmed to act this way. Would you recommend that?

Singer: And so she's going to be happy with that? She's not going to suffer as current people would the pangs of separation from their daughter or the agonies of not knowing what's happened to my daughter? She's going to feel perfectly comfortable with that, and she's going to feel good about the fact that she's helped 30 babies in Haiti to have a decent life? Is that the assumption?

Cowen: We can do it that way, but keep in mind that even if she's unhappy that's outweighed by the 30 Haitian lives which are saved. Either way you want.

Singer: Right, but you're asking me and I'm like normal human beings, I haven't been reprogrammed, so I care about my daughter or my granddaughter, or whoever this is.

Cowen: Ok, she'll be happy.

Singer: Ok, good. Then I think I'm on board with your program.

Cowen: So you would want people to be much more cooperative in this way, if we could manage it in some way that won't wreck their psyches.

Singer: That's right.

Cowen: Do you think people would have a moral obligation to genetically reprogram themselves, or it would just be a nice thing they could do if they felt so inclined?

Singer: I think if we really had a system that was as good as you're saying, would lead to as good consequences, and would leave people happy, that's something they ought to do. Because that would really be a way of making a

huge difference to the world. They would be wrong not to take advantage of this, given the benefits it involves and the absence, it seems, as described, of any major drawbacks.

Problem areas in Utilitarianism

Cowen: What do you think is the biggest problem area in Utilitarian moral theory?

Singer: The biggest problem area? One problem people are talking about that's relevant to what I'm talking about is that Utilitarian moral theory leads to highly demanding consequences that people reject. So that's one problem. The second problem, of course, is that it requires very complex calculations because we don't have a set of simple moral rules that say "don't do this, do that". We have to work out what the consequences of our actions are. As in this area we're talking about, what kind of aid is effective, what will overcome world poverty, it's very difficult to work out what the consequences are, and it's sometimes very difficult to know what's the right thing to do.

Cowen: But you think we nevertheless should do what we think is best, no matter how imperfect that guess may be?

Singer: Yeah, I don't really see what else we're supposed to do. It would seem to me to be wrong to say "because I can't calculate the consequences I'm just going to follow this simple set of rules". Because I can't calculate the consequences. But why follow this simple set of rules? Where do they come from? I don't believe that we have any god-given rules. I don't think that our moral intuitions are a good source of rules, because that's the product of our evolutionary history, which may not be appropriate for the moment that we're in. So, despite the difficulty, I don't really see what the alternative is, to trying our best to figure out what the expected utility is.

Is Utilitarianism independent?

Cowen: Let me toss up a classic criticism of Utilitarianism. I'm curious to see what you say. The criticism is this, that neither pain nor pleasure is a homogeneous thing. There are many different kinds of pains and pleasures and they're not strictly commensurable in terms of any natural unit. So when we're comparing pain and pleasure that's a fine thing to do, but in fact we're calling upon other values. So Utilitarianism is in this sense parasitic upon some deeper sense of philosophic pluralism, and we're not pure utilitarians at all. But that being the case, why don't we sometimes just allow an intuitive sense of right or wrong to override what would otherwise be the Utilitarian conclusion, since Utilitarianism itself cannot avoid value judgments?

Singer: I think the form of Utilitarianism that you're describing is Hedonistic Utilitarianism because you were talking about pleasure and pain and you were suggesting that pleasure is a whole range of different things. The form that I hold is Preference Utilitarianism which looks at people's preferences and tries to asses the importance of the preference for them. Now this is still not an easy thing to know, in fact in some ways you might say it's harder than getting measures of pleasure and pain, but I think it already embraces the pluralism that you're talking about in terms of people's preferences,

people's understanding of what it is they're choosing and why. And so I don't think it's up to us to go back and try to pull in other kinds of values that we intuitively hold over the top of people's preferences. We can do it for ourselves, each of us can say "what are my preferences", "I value this", "I value the autonomous life over the happy life, and so that's what I'm going to choose". Of course, when I weigh out your preferences I should say "well here we give weight for the preference for an autonomous life and here we give weight to the preference for the pleasant life" but in making the final judgment, in which we take everyone's preferences into account, it would be wrong for us to just pull out some intuitive values and somehow give them weight in the overall calculation because then we're giving more weight to our preferences than we're giving to those of others.

Cowen: But doesn't preference utilitarianism itself require some means of aggregation? The means we use for weighing different clashing preferences, can require some kind of value judgments above and beyond Utilitarianism?

Singer: I don't quite see why that should be so. While acknowledging the practical differences of actually weighing up and calculating all the preferences, I fail to see why it involves other values apart from the preferences themselves.

Peter Singer: Jewish Moralist

Cowen: Let me try giving you my reading of Peter Singer, which is highly speculative, and I'm not even saying it's true, it's just what I think when I read you, especially the later Peter Singer, and I'm just curious to hear your reaction to it. My reading is this: that Peter Singer stands in a long and great tradition of what I would call "Jewish moralists" who draw upon Jewish moral teachings in somehow asking for or demanding a better world. Someone who stands in the Jewish moralist tradition can nonetheless be quite a secular thinker, but your later works tend more and more to me to reflect this initial upbringing. You're a kind of secular Talmudic scholar of Utilitarianism, trying to do Mishna on the classic notion of human well being and bring to the world this kind of idea that we all have obligations to do things that make other people better off, that you're very much out of the classic European, Austrian, Viennese, ultimately Biblical tradition about our obligations to the world. What do you say?

Singer: I'm amused, I have to say. I think it's interesting. You're right that I come from a Jewish family. It was a pretty secular Jewish family, so I never got as a child, actually, a lot of Jewish teaching, never went to Jewish Sunday school, I never learned Hebrew, I never had a Bar Mitzvah, I never read the Torah. So if I had got some of that it must have come kind of at a distance through, sort of, osmosis, as you say this vaguely Jewish Viennese culture that certainly was part of my family background but was very much secularized. The interesting thing to speculate is whether I'm doing something that, say, someone out of the British Utilitarian tradition, the tradition of Bentham and Mill and Sidgwick could not have done. What are the distinctive features of my version of Utilitarianism that they would have rejected? And if there is something, it probably is attributable to that background you mention. But I'd be interested in your answer, what do you think that there is in my view that Bentham or Mill or Sidgwick could not have whole-heartedly endorsed?

Cowen: I'm not sure if there's anything, but I think the mere fact that it is you who is

doing it nonetheless reflects something about this. I think of you as one of the worlds greatest theologians, in a way, having this understanding of the quality of mercy, which is put into a secular framework, but what the intuitions really consist of, I think none of us really ever know where our moral intuitions come from.

Singer: Ok. Well, look. It's a possible view, as I think you said introducing it, you don't know whether it's true but it's an interesting view of me and where I come from. You've put it out there. I find it hard to look internally, so I'll leave it to others to judge which of the elements of my background they see having formed me most strongly.

What charities does Peter Singer give to?

Cowen: Let me try a personal question but feel free to pass on this one. Let's say someone has read your book and they say "I'm on board, Peter, please tell me what charities you give to." You mentioned Oxfam, but would you have anything specific you'd like to say? And why?

Singer: I do support Oxfam substantially, I've got a long relationship with, different Oxfams. They're actually autonomous national groups that work together, so when I first became interested in this issue as a graduate student, way back in Oxford in the '70s, I was living in Oxford and that's the headquarters of the original Oxfam, Oxfam UK, so I got in touch with them and remain connected with their office. Then I went to Australia and was involved with Oxfam Australia, now I'm involved with Oxfam America. I like what they do good grassroots work, I've seen some of that, helping the most underprivileged people, plus they're not afraid to be a real advocate for the poor, to tackle big mining companies that are pushing the poor off their land, tackle the US government and its agricultural subsidies. That's one reason that I like them. But there are many good organizations around. I've recently started supporting GiveWell, you can find them on givewell.net, because they're doing something that I'm sure you would support: they're trying to get aid organizations to demonstrate their efficacy, to be more transparent about why they support some projects rather than others, and to show how much it costs for them to achieve their goals, whether those goals are saving lives or lifting people out of poverty. And so it's kind of at a meta level, saying I want to improve aid by helping organizations that are trying to do that. I think that's a really highly leveraged way of making an impact on what's going to happen in aid over the next couple of decades.

Zero-Overhead Giving

Cowen: I'm a big fan of what I call zero overhead giving, that is I send monetary transfers to poor people, maybe I've met them on my travels, by Western Union. I don't follow up, I don't monitor, there's no tax deduction, there's no overhead, it's just money from me to them. What do you think of that as a way of giving?

Singer: Interesting. I suppose I would like to have some followup. I would worry that I was getting conned. Now, you may have a good sense of who's genuine and who's not, but we all know there's con artists working here in New York city,

in other cities in America, who could tell you wonderful stories about how they just need the bus fare home and then they'll be fine, and you give them the bus fare home, and you believe them and then next month they come up to you in the same spot and tell you the same story. So I would like some kind of auditing, but let me just say for people who do want to give direct I think not with zero overhead but I think with 10% overhead, if you go to Kiva, kiva.org, you can give a microloan to someone who is online, tells you what they want. You'll eventually, mostly, get your money back and you can lend it to someone else. I think that's quite an effective way of helping people too.

Cowen: How do you know a good charity when you see one? Is low overhead really a good measure? Those numbers are very easily manipulated.

Singer: I agree. No, "low overhead" is not the right measure. Firstly, as you say, the numbers are manipulated. Second, look, you could cut your overhead by cutting your evaluation, exactly what we were talking about. You could say "look, I'm not going to do any followup or evaluation I'm just going to hand out, basically what you said. I'm going to hand out money to poor people." That way you can get your overhead down, but are you actually doing the most good? I think you don't know that until you do have some people in the field who are in touch with what's been happening and do follow up. So I'm looking for the kind of demonstrated effectiveness that you can find in the reports from GiveWell at givewell.net rather than just checking how much of it goes to overheads and administration and how much of it doesn't.

Cowen: Keep in mind, Utilitarian calculations are very difficult, as we discussed a few minutes ago, but you don't have to listen to con stories from con men. Just fly to an Indian village, ask for people's names, get the village phone number, pick names of people who appear to be poor, they're not expecting you to show up, and send them some money. It seems to me if there's anything were you would think the chance of this doing good is really quite high it would be just sending money, and even well run charities have pretty high overhead, and you can give the money directly. Western Union has a bit of overhead, but it's relatively low. Why not have this method replace a lot of charitable giving? Because we know there's massive poverty, we know there's people who need to eat, and if someone needs to eat and you give them money, they're going to spend it on food, no?

Singer: But you can't say there's no overhead if you, say, fly to an Indian village. There's a lot of overhead, unless you're a very wealthy person. The cost of your trip, not to mention your time, is a very substantial overhead on the amount that you're giving.

Cowen: But say you're traveling anyway. You take trips, as it is, right? You go to poor countries, for other reasons. You could do a side trip to a poorer part of an urban city, in Calcutta, it would take you an hour, maybe, it wouldn't take much time. I would think at the margin there's a way for it to be quite cheap.

Singer: It may be, and the other thing you have to consider is whether putting money directly in the hands of people, say, is better than bringing in a drill to provide water for an entire village where presently they have to walk two hours to carry water from a river and that water's polluted. Maybe some sort of structural changes like that are going to help them more than just putting money in the hands of individuals.

Cowen: Keep in mind, you're a Preference Utilitarian. That doesn't mean public goods can't be more valuable, but the tendency of a Preference Utilitiarian should be to just give people resources and let them do what they want, no?

Singer: I think that's an empirical question. As you say it will depend whether they

will actually satisfy their preferences more by individual action or whether there's a kind of cooperative dilemma situation here, that actually they could achieve more good by cooperating, but maybe their culture is such that they don't cooperate unless there's some outside stimulus to get them to do so.

Moral Intuitions

Cowen: Here's a philosophical question again: do you trust your own moral intuitions?

Singer: No, not really. Over along time period, I guess, I've thought about them and reflected on them, and I've dropped some or they've faded so maybe now I'm somewhat more comfortable with them, but no, I couldn't really say that I trust them as a whole.

Cowen: What's the moral intuition that you have which you trust least?

Singer: That's a good question. I suppose, the intuitions that you have are that you ... I have intuitions about equality and fairness that make me want to go for more egalitarian solutions and, yet, I'm not sure whether they are really the right thing to do, so I'm somewhat critical of them but I'm still drawn by them to some extent. Obviously things about equality can have Utilitarian benefits if we accept laws of diminishing marginal utility and so on, and I would like to say that's the only sense in which I support equality, but I'm not sure that my intuitions are not actually more egalitarian than I should be as a utilitarian.

Improving the world through commerce

Cowen: Let's say I'm an 18 year old and I'm in college, and I've read your book and I'm more or less convinced by it, and I say to you "well what I've decided to do is I'm going to have a career in the cell phone industry because I see that cell phones are revolutionizing Africa and making many people much better off. I'm not going to give a dime to poverty but I'm going to work my hardest to become a millionaire by making cheaper and better cell phones." What do you say to me?

Singer: Well, making cheaper and better cell phones may be great for Africa, and while you're building up your business, of course, you want to reinvest your capital and make the business bigger, but are you going to get to a point, at some stage in your life, where you'll have a lot of money, where you've done your work of providing the cheap cell phones, what are you going to do with that money? I think that's still, for a Utilitarian, a relevant question. It's the kind of question that Warren Buffet asked himself. He accumulated a lot of money and said "look, I can make this money earn money faster than anyone else, so I'm going to wait until I'm old before giving it away." And that was a good thing, I guess, although now we might wish he'd given it away last year rather than this year.

Cowen: That's right, but let's say I never give a dime, I've accumulated a fortune of $200M, I've done a lot for the cell phone industry. Am I a better person than someone who's earned $40K/year and every year given 15% of it away to the poor in India?

Singer: Well I'm not sure that you're a ... "better person" asks for a judgment about the character of the agent. I think it's quite possible you've done more good

for the world, and you should be congratulated on the good that you've done for the world. We do tend to judge people by their intentions, and your intentions are a little suspect because, although you've done a lot of good for the cellphone industry and maybe for Africans you've still got this $200M. Would you really be a lot happier with $200M than with $100M, $10M say. And if not, then why not, in addition to the benefits you've conferred on people also use that $190M for something that will help people?

Cowen: If you're a Utilitarian, isn't it a little irrational to judge people by their intentions? You're retreating to this "we". "We" judge people by their intentions. You're not willing to say "I do" because that would make you inconsistent. Why not just say "Utility is what matters, I'm a Utilitarian, this person did more for the poor, this person is a better person than the one who gave a lot to charity". It's not my personal view, I'm less of a strict Utilitarian, but why not indeed embrace that conclusion rather than distance yourself from it?

Singer: Because, as a Utilitarian, praise and blame have a function, to encourage people to do good and not to do things that are bad---

Cowen: --This isn't social, this is your true view, all things considered, it's not what you say publicly to incentivise people. It's the "what you really think" question. Like, all the viewers need to turn off their BloggingHeads TV, and then you can tell me what you really think and then turn it back on again.

Singer: --but we are on BloggingHeads TV, they haven't turned it off--

Cowen: --what would you say?

Singer: Look, if I'm talking to the 18 year old, and the 18 year old is saying "look, I have these two career options. One is I do this, I confer all these benefits by developing cell phones but the end I end up pretty with my $200M and I don't give it away, and the other is I earn the $40K/year and give away whatever the percentage was", and we assume, as you said, the benefits are much less. So I'm going to tell the 18 year old to do the thing that will produce the greatest benefits. That's true. Even when he gets to 60 and he has the $200M I'm still going to think, privately, that I gave him the right advice, that was the right thing to do, I'm glad he did it. So if that's what you're asking me, that will be my judgment, and in that sense he's a better person than he would have been if he had just earned the $40K and given a lot of it away.

Cowen: You think a Utilitarian has to be a kind of Straussian and embrace certain kinds of public lies to incentivise people?

Singer: I think that's a really interesting issue. Yeah, I would say he has to be a Sidgwickian. I prefer being a Sidgwickian to a Straussian, just because Straussians have a rather bad flavor to it after they were used in the Bush administration. You could say that the Iraq War conspiracy was kind of Straussian. But, of course, Henry Sidgwick talked about that, he said that for a Utilitarian it is sometimes going to be the case that you should do good, but you need to do it secretly because if you talk publicly about what you're doing this would set an example that would be misleading to others and would lead to bad consequences. I think that's true, and I think for a Utilitarian it's inevitable that there will sometimes be circumstances in which that's the case.

What makes Peter Singer happy?

Cowen: Let me try another personal question, again feel free to pass. If you just ask yourself, "what are the things in life that just make me, Peter Singer, happy", what would you say they are? What's your own self account of what makes you happy?

Singer: It's mixed. For example, I've been touring, talking about this book. I think the book has the potential to do good in the world. I'm happy when I see that people are responding to the book. Somebody told me last night at a dinner that they'd read the book and they'd told an aid organization that they support to find a village where they could support the drilling of a well to provide water and they were going to give whatever it took to drill that well. That makes me happy, that I had this impact. Obviously I've had an impact on people changing their diet too, I have people coming up to me all the time saying "I read Animal Liberation and I became a vegetarian or a vegan and I've been working for animal groups". That makes me happy too, that my work has had that effect, which I think it a beneficial effect. But I don't want to pretend to you or to the BloggingHeads viewers that I'm a saint. I can be happy when I'm on vacation, I can be hiking in the mountains. I love mountain scenery, I had a vacation in Glacier National Park a year or so ago, which was gorgeous. That sort of thing makes me happy, and I admit that it's probably not doing as much good for the world as I could have done if instead of spending the money on that vacation I had given it to Oxfam.

Human and animal pleasures

Cowen: I sometimes ask myself, I struggle with this question, I ask "are my own deepest pleasures actually quite primeval ones", basically food, sleep, and sex. In your own writings you've emphasized, correctly, the ties between human beings and non-human animals, and it seems that for other animals these are almost always, maybe always the deepest pleasures. So I tend to think that for human beings, including ourselves, they're the deepest pleasures as well, and the higher pleasures are worth something, but actually they're somewhat of an epiphenomenon and what makes us happy are similar to what make non-humans happy. Do you have a view on this?

Singer: I'm not sure I think that the things you mention ... food and sex are obviously important, actually sleep doesn't make me particularly happy. It's something I need to do or I feel bad, but It doesn't make me happy. But yeah, food and sex are important pleasures in life. Are they more important in my life than the things that I do in work? I wouldn't really say that. I think that food and sex are the kind of desires that get satisfied: I eat a good meal, I enjoy it but I don't want to eat again for a few hours, and even sex has its limits in how much you can do at any particular time and still want more of it. Whereas the kind of things we're talking about, you can call them "higher pleasures" or "more purposive, fulfillment sort of activities" you can just go on and find that it's better and better. So I think there's a difference in that.

Cowen: So what makes you happy is pretty different from what makes a non-human animal happy, you would say?

Singer: Yes, that's true, I think higher cognitive capacities to make a difference there.

Pescatarianism

Cowen: Let me ask you a question about animal welfare. I have been very influenced by a lot of what you've written, but I'm also not a pure vegetarian by any means, and when it comes to morality, for instance, my view is that it's perfectly fine to eat fish. There may be practical reasons, like depleting the oceans, that are an issue, but the mere act of killing and eating a fish I don't find anything wrong with. Do you have a view on this?

Singer: There's certainly, as you say, the environmental aspect, which is getting pretty serious with a lot of fish stocks, but the other thing is there's no humane killing of fish, right? If we buy commercially killed fish they have died pretty horrible deaths. They've suffocated in nets or on the decks of ships, or if they're deep sea fish pulled up by nets they've died of decompression, basically their internal organs exploding as they're pulled up. I would really ... I don't need to eat fish that badly that I need to do that to fish. If I was hungry and nothing else to eat I would, perhaps, do it but not given the choices I have.

Cowen: But now you're being much more the Jewish Moralist and less the Utilitarian. Because the Utilitarian would look at the marginal impact and say "most fish die horrible deaths anyway, of malnutrition or they're eaten or something else terrible happens to them". The marginal impact of us killing them to me seems to be basically zero. I'm not even sure a fish's life is happy, and why not just say "it's fine to eat fish"? Should it matter that *we* make them suffer? It's a very non-Utilitarian way of thinking about it, a very moralizing approach.

Singer: You would need to convince me that in fact they're going to die just as horrible deaths in nature, and I'm not sure that that's true. Probably many of them would get gobbled up by some other fish, and that's probably a lot quicker than what we are doing to them.

Cowen: You have some good arguments against Malthusianism for human beings in your book. My tendency is to think that fish are ruled by a Malthusian model, and being eaten by another fish has to be painful. Maybe it's over quickly, but having your organs burst as you're pulled up out of the water is probably also pretty quick. I would again think that in marginal terms it doesn't matter, but I'm more struck by the fact that it's not your first instinct to view the question in marginal terms. You view us as active agents and ask "are we behaving in some manner which is moral, and you're imposing a non-Utilitarian theory on our behavior. Is that something you're willing to embrace, or something that was just a mistake?

Singer: Look, I think economists tend to think more in terms of marginal impact than I do and you may be right that is something I may need to think about more. Look, Tyler, I have to finish unfortunately, I've got another interview I've got to go to, so it's been great talking to you, but I think we're going to have to leave it at that point.

Cowen: Ok, thank you very much.

Singer: Thanks a lot.

Cowen: I've enjoyed it a great deal. Bye!

Singer: Ok, bye!

(I also posted this on my blog.)

The Quick Bayes Table

This is an effort to make Bayes' Theorem available to people without heavy math skills. It is possible that this has already been invented, because it is just a direct result of expanding something I read at Yudkowsky's Intuitive Explanation of Bayes Theorem. In that case, excuse me for reinventing the wheel. Also, English is my second language.

When I read Yudkowsky's Intuitive Explanation of Bayes Theorem, the notion of using decibels to measure the likelihood ratio of additional evidence struck me as extremely intuitive. But in the article, the notion was just a little footnote, and I wanted to check if this could be used to simplify the theorem.

It is harder to use logarithms than just using the Bayes Theorem the normal way, but I remembered that before modern calculators were made, mathematics carried around small tables of base 10 logarithms that saved them work in laborious multiplications and divisions, and I wondered if we could use the same in order to get quick approximations to Bayes' Theorem.

I calculated some numbers and produced this table in order to test my idea:

Decibels	Probability	Odds
-30	0.1%	1:1000
-24	0.4%	1:251
-20	1%	1:100
-18	1,5%	1:63
-15	3%	1:32
-12	6%	1:16
-11	7%	1:12.6
-10	9%	1:10
-9	11%	1:7.9
-8	14%	1:6.3
-7	17%	1:5
-6	20%	1:4
-5	24%	1:3.2
-4	28%	1:2.5
-3	33%	1:2
-2	38%	1:1.6
-1	44%	1:1.3
0	50%	1:1
+1	56%	1.3:1
+2	62%	1.6:1
+3	67%	2:1
+4	72%	2.5:1
+5	76%	3.2:1
+6	80%	4:1
+7	83%	5:1
+8	86%	6.3:1
+9	89%	7.9:1
+10	91%	10:1
+11	93%	12.6:1
+12	94%	16:1

+15	97%	32:1
+18	98.5%	63:1
+20	99%	100:1
+24	99.6%	251:1
+30	99.9%	1000:1

This table's values are approximate for easier use. The odds approximately double every 3 dB (The real odds are 1.995:1 in 3 dB) and are multiplied by 10 every 10 dB exactly.

In order to use this table, you must add the decibels results from the prior probability (Using the probability column) and the likelihood ratio (Using the ratio column) in order to get the approximated answer (Probability column of the decibel result). In case of doubt between two rows, choose the closest to 0.

For example, let's try to solve the problem in Yudkowsky’s article:

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammographies. 9.6% of women without breast cancer will also get positive mammographies. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

1% prior gets us -20 dB in the table. For the likelihood ratio, 80% true positive versus 9.6% false positive is about a 8:1 ratio, +9 dB in the table. Adding both results, -20 dB + 9 dB = -11dB, and that translates into a 7% as the answer. The true answer is 7.9%, so this method managed to get close to the real answer with just a simple addition.

--

Yudkowsky says that the likelihood ratio doesn't tell the whole story about the possible results of a test, but I think we can use this method to get the rest of the story.

If you can get the positive likelihood ratio as the meaning of a positive result, then you can use the negative likelihood ratio as the meaning of the negative result just reworking the problem.

I'll use Yudkowsky's problem in order to explain myself. If 80% of women with breast cancer get positive mammographies, then 20% of them will get negative mammographies, and they will be false negatives. If 9.6% of women without breast cancer get positive mammographies, then 90.4% of them will get negative mammographies, true negatives.

The ratio between those two values will get us the meaning of a negative result: 20% false negative versus 90.4% true negative is between 1:4 and 1:5 ratio. We get the decibel value closest to 0, -6 dB. -20 dB - 6 dB = -26 dB. This value is between -24 dB and -30 dB, so the answer will be between 0.1% and 0.4%. The true answer is 0.2%, so it also works this way.

--

The positive likelihood ratio and the negative likelihood ratio are a good way of describing how a certain test adds additional data. We could describe the

mammography test as a +9dB/-6dB test, and with only this information we know everything we need to know about the test. If the result is positive, it adds 9dB to the evidence, and if it is negative, it subtracts 6dB to it.

Simple and intuitive.

By the way, as decibels are used to measure physical quantities, not probabilities, I believe that renaming the unit would be appropriate in this case. What about DeciBayes?

A question about Eliezer

I blew through all of MoR in about 48 hours, and in an attempt to learn more about the science and philosophy that Harry espouses, I've been reading the sequences and Eliezer's posts on Less Wrong. Eliezer has written extensively about AI, rationality, quantum physics, singularity research, etc. I have a question: how correct has he been? Has his interpretation of quantum physics predicted any subsequently-observed phenomena? Has his understanding of cognitive science and technology allowed him to successfully anticipate the progress of AI research, or has he made any significant advances himself? Is he on the record predicting anything, either right or wrong?

Why is this important: when I read something written by Paul Krugman, I know that he has a Nobel Prize in economics, and I know that he has the best track record of any top pundit in the US in terms of making accurate predictions. Meanwhile, I know that Thomas Friedman is an idiot. Based on this track record, I believe things written by Krugman much more than I believe things written by Friedman. But if I hadn't read Friedman's writing from 2002-2006, then I wouldn't know how terribly wrong he has been, and I would be too credulous about his claims.

Similarly, reading Mike Darwin's predictions about the future of medicine was very enlightening. He was wrong about nearly everything. So now I know to distrust claims that he makes about the pace or extent of subsequent medical research.

Has Eliezer offered anything falsifiable, or put his reputation on the line in any way? "If X and Y don't happen by Z, then I have vastly overestimated the pace of AI research, or I don't understand quantum physics as well as I think I do," etc etc.