The Mathematics of Measuring Self-Delusion

Kristopher Tapp

Your brain is tricky. When you break up with your partner, you tell yourself that you never much liked him/her anyway. When you buy the more expensive house, you comfort yourself with an exaggerated memory of how inadequate the cheaper house was. These are examples of your tendency, perhaps unconscious, to devalue the things that you have rejected.

It might seem self-delusional, but psychologists consider it an important cognitive mechanism. It is one of the ways in which your brain rationalizes your past choices, even the irrational ones, perhaps in part to avoid regret. When you have a big new mortgage to pay and a new partner to find, you can't afford to waste mental energy rethinking your past.

Psychology experiments are also tricky. How could you design an experiment that measures this cognitive tendency to devalue already-rejected objects? How could you be sure that your experiment is measuring what you want it to measure?

One of the first such experiments was performed in 1956 by J. W. Brehm. In his experiment, an adult subject was asked to rank 10 household objects (a toaster, a hair dryer, etc.) from 1 = most desirable to 10 = least desirable. The subject was then shown two of these objects and asked to choose one of the pair that he or she would be allowed to take home. After making this selection, the subject was again asked to rank all 10 objects. In this second ranking, the subjects tended to upgrade their opinion of the object that they had selected to take home and downgrade their opinion of the object that they had elected not to take home. This seemed to support the hypothesis that adults devalue previously rejected objects.

To interpret these results, Brehm and others hypothesized that an uncomfortable psychological state (which they called cognitive dissonance) occurs when one's actions and choices are at odds with one's underlying cognitions (values,

beliefs, and attitudes). To reduce dissonance, one is motivated to shift one's cognitions to eliminate the inconsistency. Thus, Brehm believed that his subjects shifted their rankings to make these rankings more compatible with their decisions about which objects to take home, thereby reducing dissonance.

It was trickier to measure the rejection phenomenon in animals and young children, who don't have the ability to communicate a complete ranking of 10 objects. L. C. Egan, L. R. Santos, and P. Bloom claimed in 2007 to have verified that even capuchin monkeys have a cognitive tendency to devalue previously rejected objects. But the conclusions of their experiment were incorrect because of a simple logical flaw that was pointed out recently by Chen and Risen.

The Egan, Santos, and Bloom experiment went like this—see if you can spot the mistake. They used three colors of M&Ms that a capuchin monkey seemed to find about equally desirable. Let's call these colors red, blue, and green. The monkey was offered a first choice between two of these colors. After it made this choice, it was then offered a second choice between the color it had just rejected and the third color. For example, suppose its first

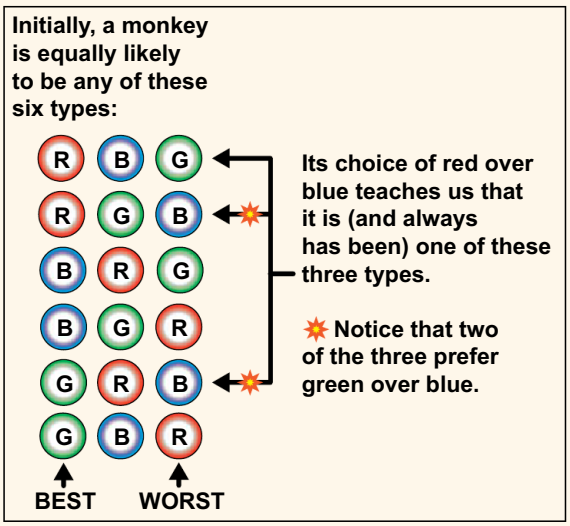
Photo by Lucas, flickr.com/photos/mrmighty/2074116664/

choice was between red and blue, and it chose red. Then its second choice would be between blue (the previously rejected color) and green (the new color). This process was repeated many times with many monkeys. About two-thirds of the time, the monkey preferred the new color over the previously rejected color. Since two-thirds is more than one-half, this was considered evidence that monkey brains are hardwired to devalue previously rejected objects and therefore re-reject them.

Even monkeys seem to rationalize their choices, the authors concluded, and according to a similar experiment with children, so do 4-year-olds. These claims shaped the discussion about how human adults rationalize their choices. Since this rejection phenomenon "occurs in creatures that lack language and complex senses of self," the authors concluded that "one must either accept that these processes are mechanistically simpler than previously thought or ascribe richer motivational complexity to monkeys and children."

Did you catch the mistake? As Chen and Risen pointed out, 2/3 is exactly the number that one should expect for mathematical reasons, without assuming that monkey brains are wired with any kind of re-rejection tendencies. Here is the correct way to think about it: Any individual monkey will have slight color preferences before the experiment begins. In fact, there are six possible preference rankings that a monkey might initially have: RBG, RGB, BRG, BGR, GRB, or GBR. It is reasonable to suppose that these six initial preference rankings are all equally likely. Suppose that, in the first stage of the experiment, the monkey chooses red over blue. This choice doesn't necessarily change the monkey's preferences, but it does give us new information about what type of monkey it is (and always has been). Of the six original types, we remove the types that prefer blue over red, and only three possibilities remain: RBG, RGB, and GRB. These three are all now equally likely. But of these three remaining possibilities, two prefer green (the new choice) over blue (the previously rejected choice). Thus, there is a 2/3 chance that the monkey will re-reject the blue. That's all there is to it. This discussion is summarized in the figure at left.

In summary, the conclusions of Egan, Santos, and Bloom were based on the unjustifiable assumption that each monkey began as a blank slate with absolutely no initial color preferences and therefore that the choices made in the second stage of the experiment could be attributed solely to attachments and aversions formed in the first stage of the experiment. Chen and Risen countered that each monkey may have had slight initial color preferences before the experiment began, and the choice that it made in the second stage of the experiment was exactly what one would expect from the type of monkey that it had re-



Chen and Risen's explanation of the 2/3 result.

vealed itself to be in the first stage of the experiment.

The experimenters were careful to use three colors that the capuchin monkey seemed to find about equally desirable. Nevertheless, the monkey still may have had slight preferences, and the experimenters' care may have insured only that the six possible rankings of its slight preference were about equally likely, not that it was completely preference free.

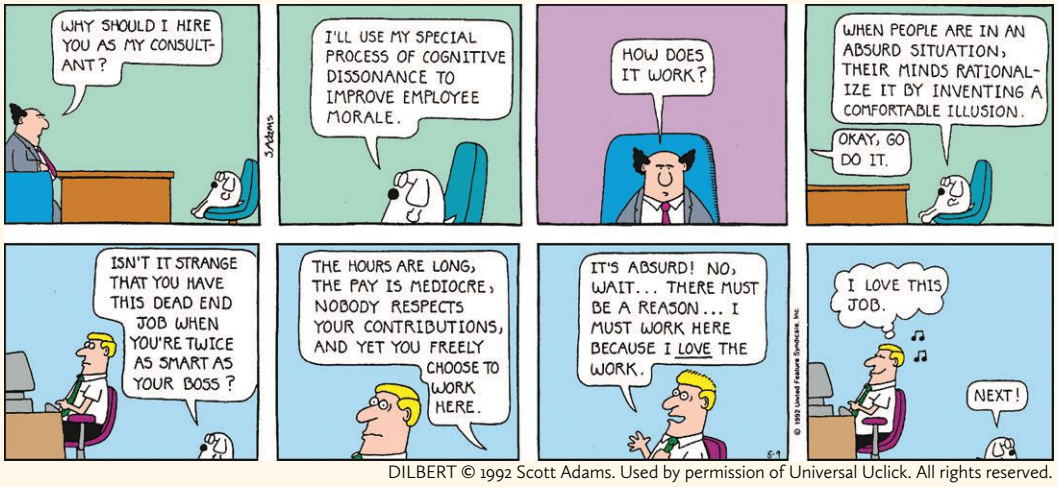
So perhaps psychologists don't yet understand whether monkeys or children rationalize their choices, but can we at least trust the experiments on adults? Unfortunately, no. Chen and Risen also took aim at Brehm's above-described experiment on adults, and many experiments like it. Brehm's experiment contains a more subtle version of the same essential logical flaw that's found in the monkey experiment.

Here is the problem. Suppose that an adult subject named Bob ranked the 10 household objects and was then given a choice between the object he ranked fourth best (a toaster) and the object he ranked seventh best (a hair dryer). Let's suppose that Bob chose to take home the hair dryer, even though he just gave it a worse rank (this is called a *reversal*). Imagine that Bob did what (according to Brehm's results) most people in Bob's situation would do: He constructed a second ranking with the hair dryer's status improved (maybe from seventh best to fifth best) and the toaster demoted (maybe from fourth best to sixth best).

Brehm would claim that selecting the hair dryer caused Bob to change his preferences. Rather than painfully dwell on the irrationality of his hair dryer choice, he unconsciously changed his preferences to become more compatible with this choice, thereby making his hair dryer selection seem perfectly rational. Maybe he even changed his hair to a style

that requires drying.

Chen and Risen, however, can explain Brehm's experimental results without assuming that the subjects ever change their minds. They constructed a mathematical model that assumes every subject has a single never-changing "true" ranking of the 10 objects throughout the experiment. At each of the three steps (the subject's first ranking, the object selection, and the second ranking), there is some random noise. Maybe the subject answers too quickly or becomes distracted or fails to take the task seriously, so each response has a certain probability of disagreeing at least slightly with the subject's true ranking. When Bob selects the hair dryer, this doesn't change his mind, but it does provide us with new probabilistic information about Bob's true ranking. If we were to guess Bob's true ranking at this point, we would of course take his hair dryer selection into account. His true feelings about hair dryers and toasters are most likely a sort of average of the feelings indicated by his first ranking and the feelings indicated by his selection of the hair dryer. Thus, based on all of our information at this point, we would guess that Bob truly likes hair dryers more (and toasters less) than his first ranking indicated. Chen and Risen proved that this additional probabilistic information is enough to account for Brehm's experimental result, without needing to assume that subjects ever change their minds.



DILBERT © 1992 Scott Adams. Used by permission of Universal Uclick. All rights reserved.

In other words, subjects' second rankings tended to shift in the direction of their object selection, not because the object selection changed their minds, but because the object selection indicated that their true feelings always were in this direction. Bob's second ranking is exactly what one would expect from the type of person he revealed himself to be when he chose the hair dryer.

The objections raised by Chen and Risen call into question the validity of a large body of social psychology literature related to choice rationalization and cognitive dissonance, including hundreds of research papers spanning five decades. After the dust settles, will there be any remaining evidence to support the rejection phenomenon? In a *New York Times* interview in 2008, Chen stated, "I don't know that there's clean evidence that merely being asked to choose between two objects will make you devalue what you didn't choose. I wouldn't be completely surprised if this effect exists, but I've never seen it measured correctly."

Is there any hope of measuring the effect correctly? Chen and Risen provided several suggestions for redesigning experiments that could correctly measure the rejection phenomenon. Implementing one of their own suggestions, they performed a modified version of Brehm's experiment, with the subjects divided between an experimental group and a control group. Just as in Brehm's study, each subject in the experimental group ranked a collection of objects, then chose between a pair of these objects, and finally ranked the same collection again. That is, members of the experimental group

Further Reading

Brehm's original experiment ranking household objects was reported in "Post-decision changes in desirability of alternatives," *Journal of Abnormal and Social Psychology* 52 (1956), 384–389.

Egan, Santos, and Bloom published the first round of their results on capuchin monkeys in "The origins of cognitive dissonance: Evidence from children and monkeys," *Psychological Science* 18 (2007), 978–983. Chen and Risen's objections first appeared in "How choice affects and reflects preferences: Revisiting the Free Choice Paradigm," *Journal of Personality and Social Psychology* 99 (2010), 573–594.

A complete bibliography is available from the author at *http://people.sju.edu/~ktapp/*

performed the three tasks in the "rank-choose-rank" order, abbreviated RCR. Members of the control group, however, performed the same three tasks in the RRC order; that is, they ranked the collection, then ranked again, and finally chose between a pair of them.

The experimenters calculated each subject's *spread*, which means the amount that the ranking of the selected object improved plus the amount that the ranking of the rejected object worsened (between the first and the second rankings). In past experiments, a positive spread has always been interpreted as evidence that, because of choosing the object, the subject shifted his/her ranking in order to reduce cognitive dissonance. But this is clearly not the correct interpretation of positive spread in a member of the RRC control group, who exhibited the positive spread before choosing the object.

As Chen and Risen predicted, members of the RRC control group had positive spread on average, presumably because their choices revealed information about their underlying preferences. The key question was whether the RCR experimental group had a significantly higher average positive spread than the RRC control group. If this were the case, then the RCR experimental group's "extra" amount of positive spread could be interpreted as evidence for Brehm's original hypothesis that they changed their preferences because the act of making the choice induced them to alter their underlying cognitions.

When the numbers were finally crunched, things looked bad for Brehm. The RCR experimental group and the RRC control group had about equal average positive spreads. Chen and Risen later performed a second experiment, with some details changed in a way that they hoped might give Brehm's hypothesis more of a fighting chance. In their second experiment, the RCR experimental group had nominally higher average positive spread than the RRC control group. But the score was close enough that the results provided only weak support for Brehm's thesis. ■

Kristopher Tapp enjoys teaching math and performing research in Riemannian Geometry at Saint Joseph's University in Philadelphia. He is the author of a textbook about Lie groups for math majors and a textbook about symmetry for all majors.

Email: ktapp@sju.edu

http://dx.doi.org/10.4169/mathhorizons.20.4.5

Concept Quiz

1. An experiment similar to Brehm's was performed in 2001 on adult amnesiacs with impaired short-term memories. When the subjects gave their second ranking, they no longer remembered their first ranking or their object selection. Nevertheless, they showed the same tendencies to upgrade their opinion of the selected object and downgrade the rejected object. How would you explain this? Does this result support Brehm's thesis or the Chen-Risen thesis?

2. Explain why positive spread is expected, on average, among people like Bob (whose object choice is a reversal of what one would predict based on his first ranking), even if nobody changes his or her mind. Explain why positive spread is also expected, on average, among people who do not exhibit such a reversal.

3. In the Egan-Santos-Bloom experiment, about two-thirds of the monkeys selected against the previously rejected color. Since their fraction was close to two-thirds (rather than being significantly more than two-thirds), does this provide evidence that monkeys do not have a tendency to devalue previously rejected objects?

4. To fix the problem with their monkey experiment, Egan, Santos, and Bloom in 2010 modified the experiment by making the monkey's first choice "blind." More precisely, a monkey watched two candies (say, red and blue) go into a box of wood shavings and was allowed to hunt until it found one of them, which it ate. Let's suppose that it found and ate the red candy. The monkey was then offered a second real choice between the candy that it didn't find (blue) and the third color (green). Egan, Santos, and Bloom hypothesized that, even though the monkey didn't really choose the red over the blue (and therefore didn't reveal any information about its underlying color preferences), the process of finding the red first might still induce the monkey to devalue the blue. Their initial results were inconclusive because they didn't use enough monkeys. Do you think that their modified approach successfully avoids the error of their original monkey experiment? If they re-perform this modified experiment with enough monkeys, what result would you expect?

http://dx.doi.org/10.4169/mathhorizons.20.4.9