UNITED STATES, WORLD

WHEN DO MONETARY INCENTIVES BACKFIRE?

APRIL 13, 2016 | STEPHAN MEIER & MATTHEW STEPHENSON | LEAVE A COMMENT

In economics "incentives matter", so much so that the bestselling book *Freakonomics* defines the discipline of economics as the study of incentives. Extrinsic incentives, like cash payments, are familiar forms of incentives and they are often proposed in order to motivate or alter the behavior of individuals. So, should students get financially rewarded for class attendance? Would a smoking cessation program with monetary incentives improve health outcomes? Or, does rewarding good behavior always motivate individuals to behave even more prosocially than without incentives?

The economic logic is clear: suppose, for example, if I asked a friend to help me with a chore, like moving furniture in my apartment. Most friends would be willing to help (otherwise, they wouldn't be your friends). However, suppose I added "and for your help I'd like to give you 80 cents." In theory my friend should be even happier, as he now gets 80 cents on top of whatever motivated him in the first place. But

contrary to the economic prediction, our intuition tells us that with a reward of 80 cents the friend is less likely to help. In fact, in psychology it has been known for quite a while that extrinsic rewards can crowd out intrinsic motivation. A book by Frey (1997) is an early example of work highlighting the way in which this idea from psychology has dramatic effects on economics.

The first field evidence on the detrimental effect of incentives (and also the best-known illustration) is by Gneezy and Rustichini (2000b). The authors looked at parents' picking up of children at an Israeli daycare center. Parents inevitably came to pick up their kids. A fine was imposed on parents who picked up their children too late and, according to economic theory, this should have made parents more likely to show up on time. But the opposite was true: parents showed up later even when the fine was imposed. In fact, after the fine was removed, parents showed up even later than they had before.

The evidence from the daycare (and elsewhere) showed that incentives did not always work as intended. Of course, if the incentives are large enough they likely will: if you paid your friend $1000 for moving the furniture, he would be more than happy to help. The empirical evidence bears this out as well: Gneezy and Rustichini (2000a) find that high school students collecting money for charity put forth more effort if they are paid more. However, they find that unpaid volunteers put forth more effort than those paid a small amount. That is, the added (small) incentives "crowded out" other motivations driving the students and led them to actually do less. The lesson is pithily summarized in the title of that Gneezy and Rustichini paper: "Pay enough– or don't pay at all."

Of course, sometimes even small incentives work. This leads us to the question: when do incentives work and when can detrimental effects be expected? Work on the mechanisms through which incentives can crowd out motivation sheds light on this question and in the following section we will review some of this work.

How can incentives backfire?

In economic interactions, incentives don't just fall from the sky. We typically assume that incentives are imposed for a reason, and our beliefs about this reason can be important. In such interactive settings, incentives can backfire for at least three reasons:

1. *Distrust and social norms.* Incentives can be seen as a signal of distrust. For instance, in a principal-agent relationship if the principal imposes an incentive scheme on the agent, that will be a signal to the agent about what the principal thinks about him or her. Specifically, the principal thinks the agent wouldn't work without the incentive and, depending on the task, this could be a signal of distrust.

To show that incentives can also signal distrust and affect social norms, Fehr and List (2004) conduct an experiment in which one player (player 1) can send money to a second player (player 2). If player 1 decides to send the money it is multiplied by a number greater than 1, then player 2 gets the new

amount—larger than what was sent—and has the option of returning some of it. Fehr and List give player 1 the option of fining player 2, thereby making it unprofitable for player 2 not to return some of the money. They observe that using this fine actually decreases the amount of money returned to them, perhaps because it signals distrust on the part of Player 1.

2. *Signaling about the task*. Offering incentives might provide information about the incentivized task itself, signaling that the task is either undesirable or difficult. Consider a question we discussed at the beginning of the paper: should we offer incentives for children to attend classes? Paying students may signal that attendance is difficult or unpleasant, or that those paying think the students are unreliable and need extra motivation.

In the daycare example described above, the small fine for picking up their children too late might have signaled that being late isn't as bad as parents assumed: "after all," they think, "it only costs this small fine." It could also signal to the parents that coming late was more common than assumed. Since the fine is being imposed, perhaps parents assumed that coming late was a social norm and thus not so bad. Especially pernicious in these cases is the fact that when the incentives (or the fine in the daycare situation) are subsequently removed, effort is often even lower than before incentives were introduced. While this may not be a problem if the incentives can be sustained, providing incentives for a short period and then removing them may be worse than never having had them in the first place.

3. *Reputation effects*. Incentives can also change behavior by reducing image motivation. For example, giving blood or recycling may be a way of showing people that you are "nice" just as cliff-diving may show people you are "brave." If you are incentivized to do it, the value of that image signal is weakened, since now you may just be doing it for the money. Ariely, Bracha, and Meier (2009) observe donations to charity, varying whether the donation is publically observable and whether the donation is incentivized. They find that, if the donations are not observed by the group, incentives work well to increase giving. However, when the donations are public, the incentives crowd out the image motivation and reduce donations. This research shows that depending on the condition incentives can work well or they can backfire.

Pay for Behavioral Change?

Incentives are often recommended for fostering behavioral change. As the discussion above shows, one has to be careful with incentives as they can may fail or even backfire. But here we explore what the empirical evidence shows in three areas: education, prosocial behavior, and health (for a more detailed discussion, see Gneezy et al. 2011).

Education: The economics literature has largely focused on two major types of educational incentives: incentives for attendance and incentives for performance. The evidence from field experiments appears to show that extrinsic incentives work well for increasing attendance and enrollment. However, the

evidence is mixed regarding effort and achievement, seeming to show that effectiveness depends on the subject and perhaps characteristics of the students. Finally, many of these experiments and studies are quite recent, so the question of how well these incentives do in the long-run largely remains to be seen.

Prosocial Behavior: In 1971, Richard Titmuss wrote that extrinsic incentives for blood donation actually caused them to decrease. This inspired much debate at the time, and the question of how extrinsic incentives affect prosocial behavior like donating blood continues to be explored. Evidence is mixed, but does seem to show that non-cash material incentives can induce more donating without crowding out intrinsic motivation. The effects of extrinsic incentives on volunteering, another example of a prosocial behavior, seem to depend on the level of the incentives as we discussed earlier. In terms of inducing volunteers to work, small payments are worse than no payments while suitably large payments are better than either.

Health: If you ask a doctor what simple steps many people could take to improve their health, "quitting smoking" and "exercising more" would top the list. Can incentives help us discourage bad habits like smoking or encourage good habits like exercising? Evidence on the long-term effects of paying people to participate in smoking-cessation programs is not promising. However, while incentives do not appear to work on smokers in the long-run, there are cases in which even the short-term benefits may be worth it (such as successfully incentivizing pregnant women to stop smoking.) The evidence is somewhat more promising for the good habit of exercise: incentives have been shown to not only help people exercise more while they receive incentives but to also sustain the habit after the incentives drop away.

Conclusion

That extrinsic incentives work to induce people to do what was incentivized is well documented and intuitive, but that is not all they do. They can also affect other intrinsic motivations such that the net effect on behavior is weakened or even the opposite of what was intended. Thus there is potential conflict between extrinsic incentives and intrinsic incentives. Because of this conflict, economists and practitioners need to be cautious and take a more comprehensive view of the things that motivate people to act. Attempts to incentivize studying, taking medications, giving blood and other desirable outcomes need to be carefully designed in light of the potential pitfalls of crowding out other motivations. It is very worthwhile to focus on incentives and to understand them, and behavioral economics is contributing to a richer picture of what those incentives might be and—crucially—how they interact.

References

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *The American Economic Review*, 544-555.

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *The American Economic Review*, 96(5), 1652-1678.

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3), 489-520.

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives—trust and trustworthiness among CEOs. *Journal of the European Economic Association*, 2(5), 743-771.

Frey, B. S. (1997). *Not just for the money: An economic theory of personal motivation*. Edward Elgar Pub.

Gneezy, U., & Rustichini, A. (2000a). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 791-810.

Gneezy, U., & Rustichini, A. (2000b). A fine is a price. *Journal of Legal Studies*, 29, 1.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 191-209.

Titmuss, R. (1971). The gift of blood. *Society*, 8(3), 18-26.

Author Biographies

Stephan Meier is an Associate Professor at Columbia Business School. His work has been published in the leading academic journals including the American Economic Review and Management Science, and has been profiled by the press such as The Economist, Wall Street Journal, Financial Times, New York Times, Los Angeles Times, and Neue Zuercher Zeitung.

Matthew Stephenson is a PhD candidate at Columbia Business School.