Prepared for: *The American Journal of Psychology*
Running head: Song

The Sirens' Song of Incentives

Review of 'Mixed signals: How incentives really work'
by Uri Gneezy

Joachim I. Krueger, Brown University

2,429 words of text
excluding references

Correspondence:

Joachim I. Krueger
Department of Cognitive & Psychological Sciences
Brown University
190 Thayer St.
Providence, RI 02912
Phone: (401) 863-2503
Home page: http://research.brown.edu/research/profile.php?id=10378

Keywords: incentives, rewards, motivation, self-help, signaling

Note: This book is published by Yale University Press, New Haven, CT, 2023.
ISBN 978-0300255539, list price: $28.00, 320 pages, hardcover

Every individual behavior ultimately depends on incentives. – U. Gneezy

Working as an economist is rewarding. – Hoca Camide

The concept of incentives lies at the heart of the social and human sciences. Why do people do what they do? Let's look at the incentives of their behavior! How can we get people to do what we want them to do? Let's look for the incentives that will work! The latter question is harder than the former because it calls for behavior change. It moves us from explanation to prediction and control (Krueger, 2020). The behavior we are currently seeing is no longer the behavior we want to see. Incentives are already at work, but we need to find new ones to effectuate change, and these new incentives have to be stronger than the old ones.

Uri Gneezy, a behavioral economist at the University of California at San Diego, defines an incentive as "a tool used to motivate people to do something that they would otherwise not do" (p. 11). Gneezy is concerned with behavior change and how to scale it up to social engineering, much like Skinner (1948) did. Gneezy asks questions about the incentives driving past and current behavior, but his goal is to have us do better in the future.

Mixed signals: How incentives really work records an economist's effort to understand human motivation and to use this understanding for the common good. In conservative economics circles, behavioral economics is seen as the bastard child of the neo-classical paradigm. The smell of illegitimacy comes from its proximity to psychology (Gul & Pesendorfer, 2008), the whore among the social sciences. Puritanical economists are careful to avoid her in public but seek her out after dark (see Bruni, & Sugden, 2007, on the uneasy co-existence of these two disciplines). Psychological research, in the neo-classical view, is contaminated by the use of deception, limited by small samples, and undermined by cheap-talk data. Economists wish to do better by avoiding experimental work altogether, and when that fails, by using incentivized behavior as data, that is, by purporting to know what an incentive is, namely money. They beg, in short, the question of what an incentive is.

Gneezy asks questions psychologists have asked for a century. He does not, for the most part, acknowledge psychological research. We do not know whether he does not know this work, or whether he conceals his knowledge for fear of offending his economics brethren. They might sanction him, thereby giving Gneezy an incentive to stay in his silo. Still, *Mixed Signals* is an important book, written by someone who has done creative and original research, and it is presented in a way that the educated reader can learn significant lessons about human behavior and how to shape it when this can be morally or politically justified.

Having established that the concept of the incentive points to both the behavioral past and the behavioral future, we may note its linguistic prehistory. The Latin *incentivum* refers to something that incites, a word stemming from the verb *incantare*. Cantare means to sing (perhaps like a siren), and the prefix in- adds the notion of something being evoked or brought forth. Money talks, economists say, but underneath the speech lies poetry, charm, and – literally – enchantment. Being enchanted, a person does what they originally – or 'really' – did not want to do. They are being used and they are loving it.

Mixed Signals comprises seven parts and a total of 30 chapters, counting the introduction and the conclusion. The first five parts, including chapter 20, form a block dedicated to questions of signaling (i.e., a form of communication) and the experimental control of behavior. The second block with chapters 21 to 24 takes the reader into behavioral and cultural change in the wild. Here, behavioral economics meets anthropology. The final block with chapters 25 to 28 is a review of a few select principles of judgment and decision-making – a belated nod to psychological science.

Gneezy's first concern is signaling. To understand incentives properly, and to see how they can end up working in unintended ways, it is necessary to understand the psychology of the signal, its sender and its receiver. "Signals," Gneezy writes, "serve as a valuable tool for communicating private information that others possibly wouldn't believe if you simply told them" (p. 21). Signals must be honest, and honesty comes at a price (McAndrew, 2019). If incentives are to come into play with transactional behavior, someone will have to pay for them. Besides being costly, signals must be sufficiently visible if the message is to be persuasive and not just self-gratifying. If Hubert tattoos the name of his beloved Hilda on his hindquarter, this may not be much of a signal in his day-to-day life, and it might not sit well with Hilda's successor. Genital mutilation, notorious when perpetrated on girls (chapter 24), though accepted for billions of boys, is a troubling case of signaling. The signal can be observed only in intimate or abusive contexts. Its message may work as intended (raising a bride's social value), but it can backfire (signifying subjugation, pain, and resentment). Gneezy does not define what a signal is. We infer from his narrative that a signal is known by its effects, which involves the way "it" is received and interpreted by an audience. "It," in other words, is a psychological datum. The female bosom may indicate fecundity, but it is a signal only if the male of the species likes to look at it.

Gneezy stresses signaling because he construes incentives as signals. We must understand, he writes, that "incentives send signals" (p. 2). Behavioral engineers provide incentives and thus send signals to shape the behaviors of others. The paradigmatic case is a promise to pay an amount of X dollars for the performance of job Y. The worker believes the promise, performs the job, and gets paid. In the world of economics, the price of the behavior is at equilibrium; no one is exploited, and both parties are better off after the transaction. This is where the story might end, but things get complicated because of unintended consequences (Kamenica, 2012). Signals can turn out to be mixed (hence the book's title), and this is why economics must call upon psychology, even if her assistance is not openly acknowledged.

A critical way in which signaling can go awry is by making behavior too dependent on incentives. This may seem paradoxical if the point of incentives is to bring behavior under control. Psychologists are familiar with the phenomenon of *overjustification* (Lepper, Greene, & Nisbett, 1973) and its grounding in theories of self-perception (Bem, 1965) and cognitive dissonance (Festinger, 1957). When more than one credible explanation for a person's behavior is available, each one loses some of its explanatory force. This *discounting effect* can occur without homeostatic motives or self-representational concerns in play (Kelley, 1972; Krueger, 2009). The availability of monetary rewards for the performance of a behavior weakens any intrinsic interest in performing it. Psychologists may bemoan the erosion of personal autonomy, while economists may bemoan the paymaster's willingness to leave money on the table.

Gneezy spares the reader this history of psychology. Instead, he quotes Thomas Schelling who once sent Gneezy an email praising his experimental work. Schelling recalled when the President of the United States ordered overtime pay to be disbursed to him and his colleagues. Schelling and friends were not happy. "I think," writes Schelling, "two motives may be involved. One was that we were all so highly motivated about our work that the idea of getting paid took all the heroic excitement out of it. The other was that we wouldn't want someone to think we were eager for Saturday work because of the overtime pay" (p. 11). Schelling – and Gneezy – maintains the motivational view of discounting, which is, if not false, incomplete.

Most people wish they had a job that allows them to do what they love and also pays well. The discounting principle militates against the fulfillment of this wish, as do the employer's interests. Perhaps we should not be surprised that some bullshit jobs in the consulting industry or the college deanery pay better than jobs involving the care of the elderly or the education of the young (Graeber, 2018). We should, however, consider it anomalous when well-compensated bullshit artists claim high moral value for the "work" they do. We should consider it anomalous when certain offices of institutional support are incentivized to reproduce and multiply, to create extra work for the value-generating branches of the business, and to lay claim to being central to the company's mission. This dynamic is as familiar to academics as it is to workers in the private sector.

Gneezy reports that employers' incentive schemes often miss the mark, as when they erode quality by incentivizing quantity (chapter 4), jeopardize risk-taking and innovation by punishing failures (chapter 5), flirt with long-term inefficiency by rewarding short-term success (chapter 6), and choke the instinct for teamwork by pampering a few pet employees with awards (chapters 7 & 12). Other ironic failures are seen when fines, taxes, or punishments are meant to stimulate behavior change (chap 8). The rich can afford to treat a fine as a cost added to a behavior they relish. Engaging in this behavior (e.g., motoring their yacht through a no-wake zone) may not only feel more delicious in the face of the risk of getting caught, but the fine, if it must be paid, can serve to broadcast one's prowess and derring-do. How about then, we may ask, a little jail time? Would the psychology of spite still work for the rich?

Whether an incentive is seen as a realized gain or an averted loss is a matter of the perceiver's mindset (Litovsky, Loewenstein, Horn, & Olivola, 2022). Gneezy banks on the stylized fact of loss aversion to argue for the strategy of paying people first and then managing their behavior with threats of taking money back. The prospect of a loss will keep workers busy, so says the logic of the endowment effect. Examples of this working well can be adduced, but one wonders how far this tactic can be taken. Some students take the loss frame for granted. They seem to think that they already "have" a grade A waiting for them. If they do not receive it, they must have "lost" points somewhere. They are puzzled by the suggestion that they start with zero points and go up from there if they do the work. The idea that loss aversion is the best motivator begs the question of why many people (like these students) construe their work as an effort to avoid loss. Why do they choose a frame that implies aversion of a frame that implies gratification? Why not opt for a frame that casts one's accomplishments as gains? If losses were averted as hoped, everyone would end up with a trophy – except those who lose them.

Another curious scenario is the pay-to-quit scheme (chapter 15). Some adventurous employers incentivize workers to quit by offering them money. It is one thing to do this in order to nudge people into retirement; it is another thing to see who will reject the offer and then work harder for less. This latter tactic suggests that a cynical game is being played. The pay-to-quit tactic is a commitment-generating device that puts the logic of overjustification into reverse. Workers who reject the offer now find their efforts underjustified, and a dissonance-reductive increase in intrinsic motivation seems to be the only psychological way out. It is fair to say that this tactic is morally questionable. Economists not given to such qualms will only regret that they haven't made this move sooner.

Gneezy next turns into the heartland of general psychology: the acquisition (chapter 17) and the breaking of habits (chapter 18). Creating a good habit – or building a habitual stock, as Gneezy calls it – should recall the literatures on reinforcement schedules (Skinner, 1938), which shapes the behavior of organisms paying attention, the role of public commitment (Tedeschi, Schlenker, & Bonoma, 1971), which is a form of social signaling in today's parlance, and self-binding, which affords self-signaling (Elster, 1984). The corollary of binding is the idea that the removal of barriers (chapter 20) liberates a person to engage in a desired behavior, an idea traceable to Kurt Lewin (1931). There is no new theory here, and Gneezy's observation that "stable habits are developed from routines" (p. 182) ends where is begins – as does his observation that "the best predictor of whether you'll [do X] next week is you have [done X] this week" (p. 184). Moving on, the discussion of breaking habits comprises compelling evidence from recent experiments, although again, little new conceptual ground is broken.

Almost as an afterthought, when discussing the curbing of impulses, Gneezy raises the question of what an incentive actually is besides being a signal (chapter 19). An event is an incentive if it controls behavior. Thorndike (1998) elevated this idea to the status of a law (see also Schlosberg, 1937). Observing humans has taught us that most of us want things like sex, sleep, and food when we are deprived, and things like money and fame whenever we can. When these *a priori* incentives are exhausted, we can simply observe what people do when acting spontaneously, and bundle the availability of these activities with more necessary but less pleasant activities, such as work. Premack (1958) discovered this principle and got it named after himself (fame!). Gneezy rediscovers this principle when writing that "people are aware of their limited willpower and are willing to pay for an effective temptation bundle that forces them to commit to a "should" activity" (pp. 197-198). This trick helps get people into the gym.

Gneezy at last goes into the field, the Maasai country to be precise. Using some of the principles he has studied in the lab, he and his collaborators have run programs to break the local traditions of young men having to kill lions to become men and of girls to have their genitals mutilated to become women. Gneezy's success in breaking the hold of these abhorrent practices is truly impressive, and he does a good job relating these stories with tact and grace. Readers will be captivated and wish for more initiatives such as these. Alas, the final few chapters on judgment and decision making seem dated and dangling. They diminish an otherwise strong book. All told, this reviewer recommends *Mixed Signals*, merely wishing the author had given psychology her due.

References

Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology, 1*, 199-218.

Bruni, L., & Sugden, R. (2007) The road not taken: how psychology was removed from economics, and how it might be brought back. *The Economic Journal,* 117, 146–173.

Elster, J. (1984). Ulysses and the sirens: Studies in rationality and irrationality. Cambridge University Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. Row, Peterson.

Graeber, D. (2018). *Bullshit jobs*: A theory. Simon & Schuster.

Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin & A. Schotter (eds.), *The foundations of positive and normative economics* (pp. 3-41). Oxford University Press.

Kamenica, E. (2012). Behavioral economics and psychology of incentives. *Annual Review of Economics, 4,* 427-452.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–174). General Learning Press.

Krueger, J. I. (2009). A componential model of situation effects, person effects and situation-by-person interaction effects on social behavior. *Journal of Research in Personality, 43*, 127-136. doi:10.1016/j.jrp.2008.12.042

Krueger, J. I. (2020). Prediction and explanation in a postmodern world. *Frontiers in Psychology: Theoretical and philosophical psychology.* 11:597706. https://doi.org/10.3389/fpsyg.2020.597706

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*, 129–137.

Lewin, K. (1931). Environmental forces in child behavior and development. In C. Murchison (Ed.), *A handbook of child psychology* (pp. 94–127). Clark University Press.

Litovsky, Y., Loewenstein, G., Horn, S., & Olivola, C. Y. (2022). Loss aversion, the endowment effect, and gain-loss framing shape preferences for noninstrumental information. *Proceedings of the National Academy of Sciences,119*(34), e2202700119.

McAndrew, F. T. (2019). Costly signaling theory. In T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science*. Springer.

Premack, D. (1959). Toward empirical behavior laws: I. Positive reinforcement. *Psychological Review, 66,* 219-233.

Schlosberg, H. (1937). The relationship between success and the laws of conditioning. *Psychological Review, 44*, 379-394.

Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century-Crofts.

Skinner, B. F. (1948). *Walden two*. Macmillan.

Tedeschi, J. T., Schlenker, B. R., & Bonoma, T. V. (1971). Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist, 26*(8), 685–695.

Throndike, E. L. (1898). Animal intelligence: An experimental study of the associative process in animals. *Psychological Monographs, 2*, No. 8.