How Field Experiments in Economics Can Complement Psychological Research on Judgment Biases

John A. List

University of Chicago, ANU, and NBER

Prepared for *Current Directions in Psychological Science*

Abstract

This review summarizes results of field experiments examining individual behaviors across several market settings—from open-air markets to rideshare markets to tax-compliance markets—where people sort themselves into market roles wherein they make consequential decisions. Using three distinct examples from my own research on the endowment effect, left-digit bias, and omission bias, I showcase how field experiments can help researchers understand mediators, heterogeneity, and causal moderation involved in judgment biases in the field. In this manner, the review highlights that economic field experiments can serve an invaluable intellectual role alongside traditional laboratory research.

C76, D76

Keywords

economics, endowment effects, field experiments

Increasingly, economists have turned to the experimental model of the physical sciences as a method to understand human behavior. As in the field of psychology, much of this research has taken the form of laboratory experiments in which volunteers enter a research lab to make decisions in a controlled environment. Over the past few decades, however, economists have been leaving the reservation. They are recruiting subjects in the field rather than in the classroom, using field goods rather than induced valuations, and using field context rather than abstract terminology in instructions. In short, they are exploring behaviors of people who have sorted themselves into a market and a function within that market.

Harrison and I (Harrison & List, 2004) argued that there is something methodologically fundamental behind this trend, and Czibor et al. (2019) summarized how field experiments have deepened scientific knowledge in economics. In this review, I take a step back to explore what can be learned from field experiments in three distinct areas of decision making and bias. The three topics I focus on—the endowment effect, left-digit bias (LDB), and omission bias—are interesting areas of decision making that psychologists have investigated for years. My explorations span the spectrum of field experiments (see the next section) and provide a glimpse of behaviors in open-air market settings (markets that buyers and sellers engage in periodically), in Lyft ridesharing, and in response to messages from tax authorities in the United Kingdom and Dominican Republic.

This review yields several insights. First, decision-making biases identified in research conducted in psychology labs are alive and well in the marketplace. Second, such biases can even have first-order influence on key economic variables with which policymakers and economists concern themselves: prices, profits, and taxes paid. Third, the field-experiment

approach provides insights that complement the psychological research in (a) identifying important mediators, heterogeneity, and causal moderators; (b) showing that a reinterpretation of certain findings is warranted; and (c) demonstrating that market pressures at work do not automatically, or immediately, neutralize behavioral biases observed in the lab. Finally, I conclude with a methodological lesson concerning generalizability: Theory is the tool that permits researchers to take results from one environment to make predictions in another.

A Primer on Economic Field Experiments

As Levitt and I (Levitt & List, 2009) described, modern economics has been influenced by three distinct periods of field experimentation. First, Neyman and Fisher laid the experimental foundation in the 1920s and 1930s by conceptualizing randomization as an instrument to achieve identification of treatment effects via experimentation with agricultural plots (see, e.g., Splawa-Neyman et al, 1990, a translation of a portion of Neyman's 1923 work). One key example within psychology during this period is the landmark work of McCall (1923), who is credited as an early proponent of using randomization rather than matching as a means to exclude rival hypotheses. His work continues to influence field experiments conducted in education today.

The second period of field experimentation was in the mid-20th century, when large-scale social experiments conducted by government agencies moved the exploration to groups of individuals affected by government tax and spending programs (see Levitt & List, 2009, for examples).

More recently, in the third period, the nature and range of field experiments has expanded through diverse controlled experiments being completed outside of the typical laboratory environment.

This third period has produced four general categories of experimentation (Harrison & List,

2004). The first is laboratory experiments, which study university students as subjects in an artificial environment. The artificiality is a feature, not a bug, of lab experiments because key design considerations and the environment are controlled. This approach has been used for centuries by social scientists; within my field, Vernon Smith won the Nobel Prize in 2002 for pioneering its use to answer economic questions. The second type of economic experiment is the artifactual field experiment (AFE), also known as "lab in the field." AFEs share most of the characteristics of lab experiments (such as having an artificial environment) but use the relevant population of interest as subjects. For example, Haigh and I (Haigh & List, 2005) conducted an AFE at the Chicago Board of Trade to explore whether professional traders exhibit myopia and loss aversion in their trading behaviors. The third type of economic experiment is the framed field experiment (FFE). Similar to AFEs, FFEs use the relevant population as subjects, but they take place in a natural environment, such as a car market, school, or hospital, with natural tasks, stakes, context, and so forth. For example, Gosnell et al. (2020) investigated whether performance information, personal targets, and prosocial incentives would motivate airline captains, who were aware that they were taking part in an experiment, to save fuel during takeoff, in flight, and upon landing. This type of field experiment adds naturalness to the choice setting and is a logical step to tackle questions concerning the import of the representativeness of the situation.

Note that all three types of economic experiments described so far are overt: Subjects are aware that they are part of an experiment. In contrast, the fourth type of economic experiment, the natural field experiment (NFE), is covert. For example, I used a within-subjects design to explore sellers' discrimination against women in open-air markets (List 2004a). The within-sellers design not only reaped greater experimental power than a between-subjects design but went beyond

measuring independent effects of manipulated variables, making it possible to estimate the full joint distribution of outcomes. The data showed that more than half the sellers gave women inferior offers (whereas only a small fraction of sellers gave women better offers than men). Further, among the discriminators, those who engaged in the greatest amount of discrimination made offers that were several standard deviations worse than the offers of the least discriminating sellers. A key mediator was sellers' subjective beliefs about the difference between men's and women's willingness to pay for the good. A key feature of NFEs is that they take one unique aspect of the experimental method—randomization—and combine it with a key trait of traditional empirical exercises in economics—field realism—to estimate treatment effects. In doing so, NFEs measure economic and behavioral parameters different from those that the other three types of economic experiments measure (see Al-Ubaydli & List, 2013).

AFEs, FFEs, and NFEs have been used extensively in economics in the past few decades to lend insight into myriad economic issues (see Czibor et al., 2019). They not only yield keen insights into economic theory and provide useful guidance to policymakers, but also provide an empirical account of behavioral principles that are shared across different domains. In this sense, field experiments are important methodologically as a useful bridge between data generated in the lab and observational data obtained in empirical studies (see also Paluck, 2010, and Cialdini, 1980).

Three Behavioral Biases Meet the Market

The Endowment Effect

People demand much more in compensation to give up a good than they are willing to pay to acquire the same good. In economics, this bias, called the *endowment effect*, is fundamental because the basic independence assumption (that preferences are orthogonal to current

entitlements) is a core feature, a bedrock, of economic theory. In a seminal laboratory study, Kahneman et al. (1990), provided evidence of an endowment effect in eight experiments.[1] Their study had an additional important influence on the profession beyond showing evidence of the endowment effect, as they argued that the endowment effect is a fundamental characteristic of individual preferences that will not be eliminated by market experience. Of course, psychological science has done more than simply pile on to this initial work, as research for decades has revealed the import and boundary conditions of endowment effects (see, e.g., Florack et al., 2014; Maddux et al., 2010).

This is where my first example of a field experiment enters the picture (List, 2003; its contribution is described in more detail in Sutter, in press). Using AFEs, FFEs, and NFEs, I examined trading patterns of sports memorabilia on the floor of a sports-card show and trading patterns of collector pins in a market constructed by Walt Disney World at the Epcot Center in Orlando, Florida. The experimental design was similar to Knetsch's (1989) groundbreaking lab experimental design in that in one treatment, I endowed a subject with good A and she had the option to trade it for good B. In a second treatment, a different subject was endowed with good B and had the option to trade it for good A. Because subjects were allocated to one of the two treatments randomly, for preferences to be consistent with standard economic theory, the proportion of subjects who choose B over A should be equal to 1 minus the proportion who choose A over B. Such a comparison permits a strict test of endowment theory versus standard economic theory.

My main results suggest that Kahneman et al. (1990) were correct that endowment theory helps to organize some of the received data: There is a lot of market evidence in favor of endowment theory, especially for inexperienced subjects. Yet Kahneman et al. seemingly overinterpreted

their data in arguing that the endowment effect is a fundamental characteristic of preferences that will not be attenuated by market experience.

In my 2003 study and a subsequent one (List, 2003, 2004b), I found that trading experience mattered a lot: Whereas inexperienced people showed an endowment effect, people with vast trading experience behaved much less in accord with endowment theory and much more in line with standard economic theory. Yet whether such moderation was causal was difficult to establish in the data: Was the observed difference in behavior truly attributable to trading experience, or was it due to a different correlated variable or even, for example, a selection effect? In a later study (List, 2011), I effectively randomized variation in trading experience by giving one group of subjects an incentive to trade in the market to gain experience and giving another group no such incentive. I found that the treatment directly altered the endowment effect. The thrust of these results regarding market and trading experience has been broadly replicated in economics and psychology (see, e.g., Anagol et al., 2018; Kermer et al., 2006; Lindsay, 2019).

The scientific process did not stop there, however. A group of us (Tong et al., 2016) went back to the lab and dug a bit deeper into my results (List, 2003, 2004b, 2011). We conducted two functional MRI studies to investigate mediation by neural responses, that is, how trading experience changes neural correlates of the endowment effect. We found that experience matters in a manner similar to what I originally reported, but, more important, we found evidence that the influence of greater trading experience on the endowment effect is mediated, in part, by lower sensitivity of the right anterior insula during selling. This finding is consonant with the notion that frequent trading mitigates the endowment effect indirectly by modifying negative affective responses in the context of selling. Overall, these insights on mediation and moderation

naturally open an area of inquiry to which psychologists can lend deeper insights with future research.

Left-Digit Bias

The leftmost digit (or digits) of a number disproportionately affects people's choices. For example, when people use the Lyft app for a ride, a price difference between $7.99 and $8.00 has a much greater effect on choices than a price difference between $8.00 and $8.01, even though they are equivalent price differences. This cognitive bias has important pricing implications and holds import for how humans receive and process information. For economists, unlocking secrets of this bias holds promise to advance understanding of perhaps the most fundamental tenet in economics: The Law of Demand, which states that when price rises, quantity demanded falls, and when price falls, quantity demanded rises. This law is as close as economists come to emulating the laws of physics, chemistry, biology, and the earth sciences. Nevertheless, very little is known about its underpinnings, and whether psychological science can lend relevant insights.

Of course, the LDB has not evaded psychologists. Hinrichs et al. (1982) examined how subjects use place information in a number task and provided foundational evidence on how the number of digits influences speed and accuracy in understanding a number's value. Fraser-Mackenzie et al. (2015) combined loss aversion with the left-digit effect to develop a value function that predicts the key characteristics of LDB. They found support for their theory using a laboratory experiment and observational data. More recently, Tang (2020) found the LDB among a group of Chinese subjects residing in Hong Kong, where prices ending in the number 9, such as $2.99, do not occur in the local pricing system. This finding suggests that familiarity has little influence on

the LDB.

My colleagues and I recently worked in partnership with Lyft to explore the LDB (see List et al., 2020). Combining an NFE that included more than 21 million Lyft passengers with observational data from more than 600 million Lyft sessions, we reported several unique insights. First, the law of demand consistently held in the overall market data. Yet its underpinnings surprised us. Using a simple model of LDB, we estimated that the value of the parameter for the level of *inattention* to prices in this market is approximately 0.5. This parameter value means that passengers perceive a price that is lowered 1¢ below a dollar value as if it were lowered by 50¢. To put this another way, the contribution of the LDB to the law of demand is such that roughly 50% of the decline in quantity demanded occurs at dollar values, even though such changes are only1/100 of the overall price variation. Thus, observed customer choice revealed strong warpings, or nonlinearities, between actual and subjective prices, even in the rich decision-making setting of a rideshare app with highly competitive pricing.

Second, examining possible moderation of the LDB, we found a remarkable level of stability in the inattention parameter across passengers and ride types. Even though responsiveness to changes in price was very different for different types of rides (e.g., airport rides vs. non-airport rides), the level of inattention we observed was quite similar. The robustness of inattention we found across different types of people and situations sets this bias apart from the many other quite heterogeneous biases that are studied in behavioral economics: The endowment effect, present bias (tendency to give stronger weight to payoffs that are closer to the present time), omission bias (discussed in the next section), and others all have key moderators. Further research in this area is warranted, as the level of robustness of the LDB in our data merits serious consideration.

Finally, we also found that relative prices worked in a similar way. When a consumer was deciding whether to choose one of Lyft's two main products, Classic and Shared, LDB influenced which they chose, interestingly showing that cross-price elasticities are affected by this behavioral bias. To my best knowledge, this is the first illustration of such a comparison effect in pricing and LDB's role in that comparison. These results combined with those above suggest that Lyft could substantially increase its profitability by taking the LDB fully into account in its pricing should it choose. For behavioral scientists, this work suggests an overarching deeper message: Fundamental tenets in economics, such as The Law of Demand, might have deeper psychological underpinnings than most people believe.

Omission bias

The omission bias is the tendency to judge harmful actions as worse than equally harmful inactions. Looking no further than COVID-19 vaccination decisions will reveal the power of this bias. People who argue that they will not get vaccinated because of the potential chance of death from the vaccination, even though the probability of dying from the disease itself is much larger, are behaving consistently with omission bias. In some ways, such behavior makes sense; the "I don't want this weight on my head" mentality permeates many decisions, and even when the choice is stacked in favor of taking an action, some people might view inaction as appropriate, and even desirable (see Ritov & Baron, 1990).

To dig into the underpinnings of omission bias, Spranca et al. (1991) had experimental subjects read scenarios concerning pairs of options that differed in whether they involved omission or commission. In creating these scenarios, they attempted to control intentions, motives, and consequences. They found that subjects often rated harmful omissions as less immoral than

identical harmful commissions. Recently, Jamison et al. (2020) attempted to replicate these findings, examining their generalizability to a between-subjects design with extensions examining causality, intent, and regret. They found data patterns similar to those reported by Spranca et al. (1991): Subjects had a preference for harm caused through omission rather than through commission. Finally, Baron and Ritov (2004) provided a discussion that raises questions about the generality and applicability of omission bias and reported new data showing omission bias, but a substantial amount of heterogeneity.

I became interested in this area when I partnered with the United Kingdom's tax authority to approach citizens who had failed to repay overpaid government benefits, a potential act of omission. The key part of our NFE (reported in Hallsworth et al., 2015) augmented the government's traditional repayment letter by including the following message in our omission-to-commission condition: "Previously, we treated your lack of response as an oversight. Now, if you do not call [telephone number], we will treat this as an active choice."

Our NFE, which included more than 40,000 observations, showed that repayment rates were roughly doubled when the act was reframed as one of commission rather than omission. We estimated that this reframing of the perceived nature of the action generated more than $1.3 million of new yield. Our data suggest that citizens' response to the commission message was rational in the sense that they anticipated greater punishment if they did not act honestly, as indicated by their responses to subsequent surveys. This is at odds with the traditional omission-bias literature, which interprets omission as a bias rather than a strategic choice. An "omission strategy" can be part of how a person optimally deals with such decision environments. This key mediation result is consistent with insights from psychological work using AFEs (DeScioli et al., 2011).

To explore this insight in a different setting, my colleagues and I (Holz et al., 2020) conducted an NFE in partnership with the Dominican Republic's tax authority. Just before 84,000 subjects filed their taxes, we sent them messages concerning errors on their tax forms and found that the effect of framing errors as acts of commission depended on the perceived underlying punishment. A message that framed errors as an active choice and reminded recipients to pay their taxes or threatened public disclosure of tax evasion did not affect the amount of taxes paid. However, a message that framed errors as acts of commission and also highlighted the possibility of a prison sentence increased the amount subjects paid by $50 million relative to a message that only highlighted the possibility of a prison sentence and by $100 million relative to a message that only reminded recipients to pay their taxes.

Epilogue

The three examples I have described paint a consistent picture of how field experiments in economics can lend insights that complement received results from psychology. Given that these examples come from specific markets, the reader might wonder whether the results will generalize from these experimental settings to other target settings. When generalizing, one must keep in mind that initial thoughts are basically guesses about possible valid generalizations (see List, 2020). In the course of scientific discovery itself, researchers learn about the justification of possible generalizations through their accumulated experience, but a logical generalization is not deducible from the details of one experiment. Rather, researchers should use theory to guide their guesses about generalizability.

Economics provides a useful set of theoretical tools to help researchers think about external validity. In the economic framework, individual choices depend on preferences, beliefs, and

constraints. Observed differences in choice across settings can be rationalized by each of these three types of variables. That should be the starting point for determining whether, and to what extent, behavior will change from one setting to another. Results I have summarized here show that key pieces of behavioral biases do traverse settings. Yet situational features (preferences, beliefs, and constraints) will likely lead to difficulties in generalizing in some cases (Levitt & List, 2007). In the end, all empirical results are externally valid for some setting, and no empirical result will be externally valid for all settings.

Acknowledgments

I received incredible research support from Uditi Karna, Justin Holz, and David Novgorodsky. The Editor, Robert Goldstone, and 3 anonymous reporters provided excellent remarks. All errors remain my own.

Note

1. Note that Knetsch and his colleagues (see, e.g., Knetsch, 1989; Knetsch & Sinden, 1984) had previously published experimental evidence of the endowment effect, and Thaler (1980) originally coined the term.

References

Al-Ubaydli, O., & List, J. A. (2015). On the generalizability of experimental results in economics. In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology* (pp. 420–462). Oxford University Press.

Anagol, S., Balasubramaniam, V., & Ramadorai, T. (2018). Endowment effects in the field: Evidence from India's IPO lotteries. *The Review of Economic Studies, 85*(4), 1971–2004.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.

Cialdini, R. (1980). Full-cycle social psychology. *Applied Social Psychology Annual, 1*, 21–47.

Czibor, E., Jimenez-Gomez, D., & List, J. A. (2019). The dozen things experimental economists should do (more of). *Southern Economic Journal, 86*(2), 371–432. https://doi.org/10.1002/soej.12392

DeScioli, P., Christner, J., & Kurzban, R. (2011). The omission strategy. *Psychological Science, 22*(4), 442–446.

Florack, A., Kleber, J., Busch, R., & Stöhr, D. (2014). Detaching the ties of ownership: The effects of hand washing on the exchange of endowed products. *Journal of Consumer Psychology, 24*(2), 284–289.

Fraser-Mackenzie, P., Sung, M., & Johnson, J. E. V. (2015). The prospect of a perfect ending: Loss aversion and the round-number bias. *Organizational Behavior and Human Decision Processes, 131*, 67–80. https://doi.org/10.1016/j.obhdp.2015.08.004

Gosnell, G., List, J. A., & Metcalfe, R. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy, 128*(4), 1195–1233.

Haigh, M. S., & List, J. A. (2005). Do professional traders exhibit myopic loss aversion? An experimental analysis. *Journal of Finance, 60*(1), 523–534.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2015). *The making of homo honoratus: From omission to commission* (NBER Working Paper No. 21210). National Bureau of Economic Research. https://doi.org/10.3386/w21210

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.

Hinrichs, J. V., Berie, J. L., & Mosell, M. K. (1982). Place information in multidigit number comparison. *Memory & Cognition, 10*(5), 487–495.

Holz, J. E., List, J. A., Zentner, A., Cardoza, M., & Zentner, J. (2020). *The $100 million nudge: Increasing tax compliance of businesses and the self-employed using a natural field experiment* (NBER Working Paper No. 27666). National Bureau of Economic Research. https://doi.org/10.3386/w27666

Jamison, J., Yay, T., & Feldman, G. (2020). Action-inaction asymmetries in moral scenarios: Replication of the omission bias examining morality and blame with extensions linking to causality, intent, and regret. *Journal of Experimental Social Psychology, 89*, Article 103977. https://doi.org/10.1016/j.jesp.2020.103977

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325–1348.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science, 17*(8), 649–653.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review, 79*(5), 1277–1284.

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics, 99*(3), 507–521.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives, 21*(2), 153–174. https://doi.org/10.1257/jep.21.2.153

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review, 53*(1), 1–18.

Lindsay, L. (2019). Adaptive loss aversion and market experience. *Journal of Economic Behavior & Organization, 168*, 43–61. https://doi.org/10.1016/j.jebo.2019.09.023

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics, 118*(1), 41–71.

List, J. A. (2004a). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics, 119*(1), 49–89.

List, J. A. (2004b). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica, 72*(2), 615–625.

List, J. A. (2011). Does market experience eliminate market anomalies? The case of exogenous market experience. *American Economic Review, 101*(3), 313–317.

List, J. A. (2020). *Non est Disputandum de generalizability? A glimpse into the external validity trial* (NBER Working Paper No. 27535). National Bureau of Economic Research. https://doi.org/10.3386/w27535

List, J. A., Muir, I., Pope, D., & Sun, G. (2020). *The Case of Left-Digit Bias: A Behavioral Anomaly that Transports Across People and Places* [Unpublished manuscript]. Kenneth C.

Griffin Department of Economics, University of Chicago.

Maddux, W. W., Yang, H., Falk, C., Adam, H., Adair, W., Endo, Y., Carmon, Z., & Heine, S. J. (2010). For whom is parting with possessions more painful? *Psychological Science, 21*(12), 1910–1917.

McCall, W. A. (1923). *How to experiment in education*. Macmillan.

Paluck, E. L. (2010). The promising integration of qualitative methods and field experiments. *The ANNALS of the American Academy of Political and Social Science, 628*(1), 59–71. https://doi.org/10.1177/0002716209351510

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*(4), 263–277.

Splawa-Neyman, J., Dabrowska, D. M., & Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science, 5*(4), 465–472. https://doi.org/10.1214/ss/1177012031

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*(1), 76–105.

Sutter, M. (in press). Does market experience eliminate market anomalies? (by John A. List). In G. Charness & M. Pingle (Eds.), *The art of experimental economics: Twenty top papers reviewed*. Routledge.

Tang, J. (2020). The left-digit effect in price and date perception. *Human Behavior and Brain, 1*(3), 64–66. https://doi.org/10.37716/HBAB.2020010302

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Tong, L. C. P., Ye, K., J., Asai, K., Ertac, S., List, J. A., Nusbaum, H. C., & Hortaçsu, A. (2016).

Trading experience modulates anterior insula to reduce the endowment effect. *Proceedings of the National Academy of Sciences, USA, 113*(33), 9238–9243.

https://doi.org/10.1073/pnas.1519853113