RESEARCH · RESEARCH DESIGN

Field Experiments Versus Laboratory Experiments

How laboratory and field experiments differ, the trade-off between control and realism, the spectrum of designs between the two, the debate over whether laboratory findings generalise, and how to choose and combine the two in academic and commercial research.

ModQu research article · 2,877 words · 24 references (APA 7th edition)

Experiments can be run in two very different places. In a **laboratory experiment**, participants come to a controlled setting, often a room with computers, and take part in tasks designed by the researcher. They know they are being studied, and almost everything about the situation is chosen in advance. In a **field experiment**, the treatment is applied in the natural environment where the behaviour of interest occurs: in shops, workplaces, websites, schools or villages. Participants are often unaware that they are part of a study at all. Both are true experiments, because the researcher manipulates a treatment and assigns it at random, but they make different trade-offs between control and realism. This article explains those trade-offs, the range of designs between the two extremes, the long-running debate over whether laboratory results apply to the real world, and how researchers and organisations can choose between, and combine, the two approaches.

What the two settings share

Both laboratory and field experiments rest on the same logic. A researcher manipulates a treatment, randomly assigns units to treatment and control conditions, and compares outcomes. Random assignment makes the groups alike, on average, in every respect except the treatment, so a difference in outcomes can be attributed to the treatment (Shadish et al., 2002). The distinction between the two is not about this logic, which is identical, but about the setting in which it is applied, the people who take part, the tasks they perform and how much they know about the study.

The laboratory experiment

The laboratory offers the researcher maximum control. The environment is standardised, so every participant experiences the same instructions, the same timing and the same surroundings. Extraneous influences such as noise, distractions and other people are removed or held constant. Treatments can be designed with precision, varying one feature at a time while everything else stays fixed. Outcomes can be measured carefully, often with instruments that would be impractical elsewhere, such as reaction times or eye tracking.

This control brings three main benefits. First, laboratory experiments have high **internal validity**: when the treatment and control groups differ, there are few plausible explanations other than the treatment. Second, they are well suited to testing theory, because they can isolate the precise mechanism a theory proposes. Falk and Heckman (2009) argue that

this ability to vary one factor at a time, while holding everything else constant, is the laboratory's chief contribution, and that it makes laboratory experiments a major source of knowledge in the social sciences. Third, they are relatively cheap, quick and easy to replicate, so findings can be checked by other researchers using the same procedures.

The laboratory has well-known costs. Participants know they are being observed, and Orne (1962) showed that they often try to work out the purpose of a study and behave in the way they think the researcher expects. These **demand characteristics** can produce effects that would not occur outside the laboratory. The tasks are usually simplified, the stakes are usually small, and the time frame is usually short. And the participants are often unusual: many laboratory studies recruit university students, who differ from the wider population in age, education and experience. Henrich et al. (2010) showed that most published behavioural research draws on people from Western, educated, industrialised, rich and democratic societies, who are often outliers on measures such as fairness, cooperation and visual perception.

The field experiment

The field experiment gives up some control in exchange for realism. The treatment is delivered in the setting where the behaviour naturally occurs, to the people who naturally engage in it, often without their knowledge that a study is taking place. The outcome is real behaviour with real consequences: purchases, attendance, votes, energy use or job applications.

Bertrand and Mullainathan (2004) provide a well-known example. They sent about 5,000 fictitious résumés in response to job advertisements in Boston and Chicago, randomly assigning names that sounded typically white or typically African American to otherwise similar résumés. Résumés with white-sounding names received about 50% more callbacks. Because the employers did not know they were in a study, and the only difference between résumés was the name, the experiment measured discrimination in actual hiring decisions in a way that a survey of employers' attitudes, or a laboratory simulation, could not.

Gerber et al. (2008) offer another. In a 2006 primary election in Michigan, they randomly assigned about 180,000 households to receive one of four mailings or nothing. The most effective mailing told recipients that whether they voted was a matter of public record and listed the past voting of their neighbours; it raised turnout by about 8 percentage points compared with the control group. The study showed the power of social pressure on real behaviour, at a scale and with a realism no laboratory could match.

The strengths of field experiments mirror the weaknesses of the laboratory. Because participants usually do not know they are being studied, demand characteristics are reduced. Because the setting, the people and the stakes are real, results are more likely to apply to similar real situations. Gerber and Green (2012) argue that for questions about how people respond to actual interventions, such as policies, programmes and marketing campaigns, field experiments provide the most direct evidence available.

Field experiments also have costs. They are usually more expensive and slower to run. The researcher has less control over how the treatment is delivered: staff may not follow instructions, participants in the control group may encounter the treatment anyway, and other events may intervene. Outcomes are often limited to what can be observed in the field, which may not reveal why an effect occurred. And field experiments raise distinctive ethical questions, because participants may not consent to taking part.

A spectrum, not a dichotomy

The line between laboratory and field is not sharp. Harrison and List (2004) proposed a classification that places experiments along a spectrum according to who the participants are, what task they perform, what is at stake and whether they know they are in an experiment.

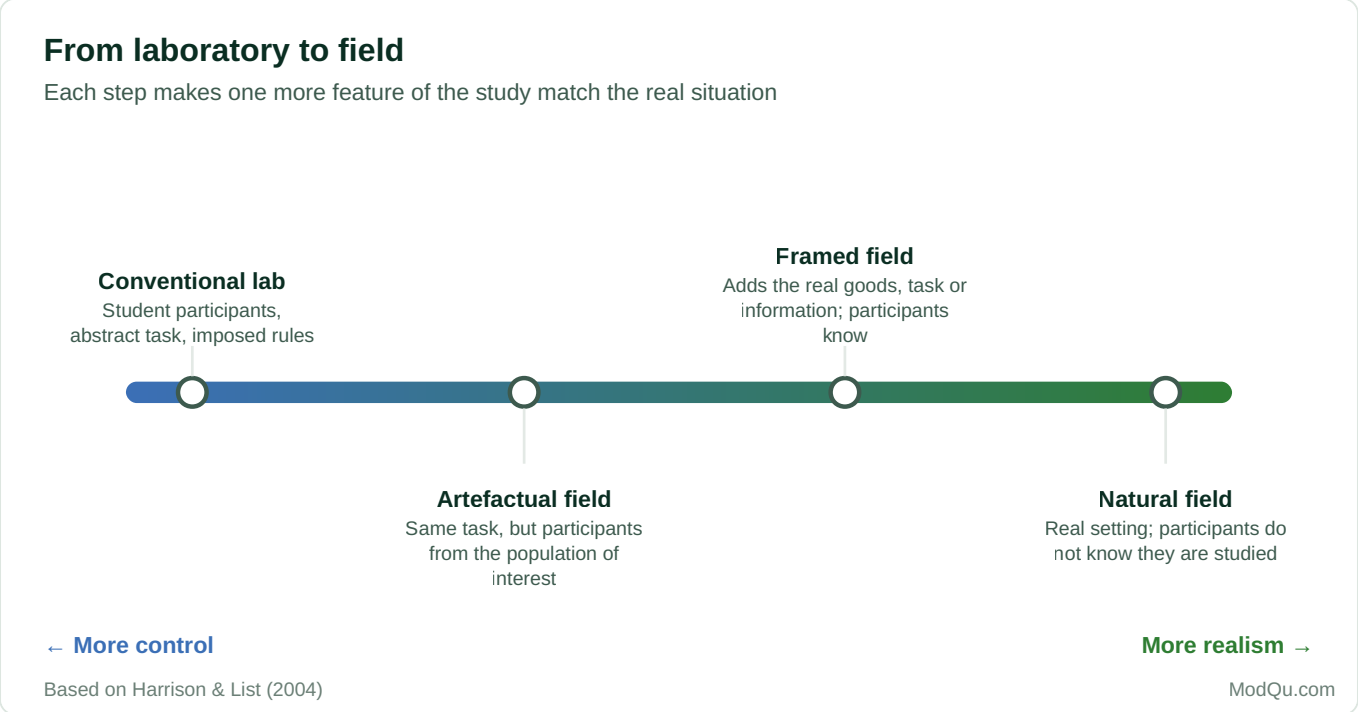


Figure 1. Harrison and List's (2004) classification of experiments. Moving from left to right, more features of the study match the real situation, while control over the setting declines. Diagram by ModQu.

- A **conventional laboratory experiment** uses a standard participant pool, usually students, an abstract task and imposed rules.
- An **artefactual field experiment** uses the same kind of task but recruits participants from the population of interest, such as traders, farmers or business owners.
- A **framed field experiment** adds the real context: the commodity, task, stakes or information that participants would normally use. Participants still know they are in an experiment.
- A **natural field experiment** takes place in the environment where participants naturally perform the task, and they do not know they are taking part.

This classification helps make clear which features of a study differ from the real situation of interest. A study can be realistic in its participants but artificial in its task, or realistic in its task but artificial in its awareness. Each departure is a possible reason why results might not generalise.

Does the laboratory generalise?

Whether laboratory findings hold in the real world has been debated for decades, and the answer depends on what kind of finding is being generalised.

The case for caution

Levitt and List (2007) argued that laboratory experiments measuring social preferences, such as fairness, altruism and trust, are likely to overstate how much these preferences matter outside the laboratory. They identified several reasons: participants know they are being watched, which makes them more generous; the stakes are small; participants are selected in ways that differ from the population; and the context of the experiment may suggest how participants are expected to behave. In laboratory games, participants often share money with strangers; in comparable field settings, where people do not know they are observed, generosity is often lower.

Mitchell (2012) compared effect sizes from laboratory and field studies in psychology across many topics. Some areas, such as industrial and organisational psychology, showed close agreement. Others, notably social psychology, showed weaker agreement, and a substantial number of findings reversed direction between the two settings.

The case for the laboratory

Defenders of laboratory research make several replies. Mook (1983) argued that much laboratory research is not intended to generalise directly. Its purpose is to test whether a theory's predictions hold under the conditions the theory specifies. If they do, the theory, not the specific result, is what generalises. Anderson et al. (1999) compared effect sizes for the same psychological phenomena studied in the laboratory and the field, such as the effects of leadership style or depression on behaviour, and found that they were generally similar. Camerer (2015) reviewed studies that tested the same economic behaviours in both settings and concluded that laboratory results generally predicted field behaviour well when the two were closely matched.

Aronson and Carlsmith (1968) drew a useful distinction between **mundane realism**, the extent to which an experiment resembles everyday life, and **experimental realism**, the extent to which participants are engaged and take the situation seriously. A laboratory experiment can lack mundane realism but still have experimental realism, and it is the latter that matters most for whether participants behave genuinely.

What the debate suggests

The two positions are less opposed than they first appear. Falk and Heckman (2009) point out that field experiments also have generalisation problems: a result from one country, one company or one time may not apply to another. Allcott (2015) gave a striking example. An energy company programme that sent households reports comparing their electricity use with their neighbours' was evaluated by randomised field trials at many sites. The first ten sites produced larger savings than later sites, because the utilities that adopted the programme first were those whose customers were most likely to respond. The field setting did not guarantee that results would generalise even to other field settings.

The lesson is that external validity is never established by setting alone. It depends on whether the people, treatments, outcomes and contexts in the study resemble those to which the results will be applied, a point Shadish et al. (2002) make in their account of generalised causal inference.

Comparing the two approaches

The main differences can be summarised along several dimensions.

Laboratory and field experiments compared

The same causal logic, applied with different trade-offs

Dimension	Laboratory	Field
Control	High: one factor varied, all else fixed	Lower: delivery and setting vary
Realism	Simplified tasks, small stakes	Real settings, people and stakes
Awareness	Participants know they are studied	Participants often unaware
Mechanism	Can isolate why an effect occurs	Usually shows whether it occurs
Cost and speed	Cheaper, faster, easy to replicate	Costlier, slower, needs partners
Consent	Informed consent as standard	Departures must be justified

Based on Harrison & List (2004); Falk & Heckman (2009); Gerber & Green (2012) ModQu.com

Figure 2. Laboratory and field experiments compared on six dimensions. Based on Harrison and List (2004), Falk and Heckman (2009) and Gerber and Green (2012); table by ModQu.

Control. The laboratory offers precise control over the treatment and environment; the field offers much less.

Realism. The field offers real settings, people and stakes; the laboratory offers simplified versions.

Awareness. Laboratory participants know they are in a study; field participants often do not.

Mechanism. The laboratory can isolate why an effect occurs; the field usually shows whether it occurs.

Cost and speed. Laboratory studies are typically cheaper and faster; field studies often require partnerships with organisations and longer timelines.

Ethics. Laboratory studies obtain informed consent as standard; field studies must justify any departure from it.

Examples of the two working together

Some of the most informative research programmes combine both settings, using each to address the weaknesses of the other.

Gneezy and Rustichini (2000) studied ten day-care centres in Haifa, Israel, where parents sometimes arrived late to collect their children. In six centres, the researchers introduced a small fine for late collection; four centres served as controls. Contrary to standard expectations, late pickups increased in the centres with fines, and remained higher after the fine was removed. The authors argued that the fine changed how parents understood the situation, turning a moral obligation into a price they could choose to pay. A finding this surprising, from a natural setting, invites laboratory work to test the proposed mechanism under controlled conditions.

The reverse sequence is also common. Ariely et al. (2003) showed in laboratory studies that people's willingness to pay for ordinary goods could be influenced by an arbitrary number, such as the last two digits of their social security number, that they had been asked to consider beforehand. Laboratory findings of this kind have prompted field tests of how anchors, defaults and reference prices affect real purchases.

Iyengar and Lepper (2000) conducted part of their influential study of choice in a supermarket, where shoppers were offered either six or twenty-four jams to taste. More shoppers stopped at the larger display, but a much higher proportion of those offered the smaller selection went on to buy. The finding prompted many follow-up studies, both in the laboratory and in the field, and later work found that the effect varies considerably depending on the setting and the people involved. The sequence illustrates how a single striking result, whether from the field or the laboratory, needs replication across settings before it can guide decisions.

Field experiments in business and policy

Field experiments have become a routine tool for many organisations. Online companies run large numbers of randomised experiments on their websites and apps, testing changes to design, pricing, recommendations and messages on real users. Kohavi et al. (2020) describe how these online controlled experiments, often called A/B tests, have become central to product development at large technology firms, and the safeguards needed to run them reliably.

Field experiments have also become prominent in development economics and public policy. Duflo et al. (2007) set out the practical methods for running randomised evaluations in developing countries, and Banerjee and Duflo (2009) described how the approach transformed the study of programmes in education, health and microfinance. Ayres et al. (2013) used two large field experiments with energy utilities to show that sending households comparisons of their own energy use with their neighbours' reduced consumption by about 1–2%.

Deaton and Cartwright (2018) caution that randomised field trials are not a gold standard that settles every question. A well-run trial shows that a treatment worked, on average, in the population and setting studied. Knowing whether it will work elsewhere requires understanding why it worked, which often calls for theory, other evidence and, in some cases, laboratory studies of the mechanism.

Choosing between them

The right choice depends on the question.

- **If the aim is to test a theory or isolate a mechanism**, the laboratory is usually the better starting point. It allows precise manipulation of the factor the theory identifies and measures intermediate outcomes, such as beliefs or attention, that show how the effect works.
- **If the aim is to estimate the effect of a real intervention**, such as a price change, a new service or a policy, a field experiment is usually better, because the result applies directly to the setting of interest.
- **If the behaviour is strongly affected by being observed**, such as honesty, generosity or discrimination, a natural field experiment is especially valuable.
- **If the treatment cannot ethically or practically be applied in the field**, a laboratory study or a framed field experiment may be the only option.
- **If resources are limited**, laboratory or online studies can screen ideas before a more costly field test.

List (2011) offers practical advice for those running field experiments: build relationships with partner organisations early, understand the setting and its constraints, design the study so that the partner benefits from the result, and plan for the things that will go wrong, such as staff not following assignment rules.

Implications for commercial research

For organisations, the distinction between laboratory and field corresponds closely to the distinction between asking customers what they would do and observing what they actually do. Concept tests, simulated shopping tasks and online survey experiments are, in effect, laboratory experiments. They are fast, cheap and controlled, and they are excellent for comparing options and understanding reactions. But their results reflect a situation in which respondents know they are being asked, the money is not real and the competing demands of everyday life are absent.

Field tests, such as in-market trials, A/B tests and staggered rollouts with random assignment, show what happens when real customers make real decisions. They are the stronger basis for a final decision about whether to launch, but they are slower and more expensive, and they show less about why customers respond as they do.

A sensible approach combines the two. Controlled studies can narrow a set of options and explain customer reasoning. Field tests can then confirm that the preferred option works in practice, at the scale and in the setting where it will be used. When the two agree, decision-makers can act with confidence. When they disagree, the disagreement itself is informative, often pointing to a feature of the real setting that the controlled study did not capture.

Conclusion

Laboratory and field experiments apply the same causal logic in different places. The laboratory offers control, precision and the ability to test mechanisms, at some cost in realism. The field offers real behaviour in real settings, at some cost in control and explanation. Neither setting guarantees that results will generalise; that depends on how closely the people, treatments and contexts studied resemble those of interest. The strongest evidence usually comes from programmes of research that use both, letting the laboratory explain what the field reveals and the field test what the laboratory predicts.

References

Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal of Economics, 130*(3), 1117–1165. https://doi.org/10.1093/qje/qjv015

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*(1), 3–9. https://doi.org/10.1111/1467-8721.00002

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics, 118*(1), 73–106. https://doi.org/10.1162/00335530360535153

Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., Vol. 2, pp. 1–79). Addison-Wesley.

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization, 29*(5), 992–1022. https://doi.org/10.1093/jleo/ews020

Banerjee, A. V., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics, 1*, 151–178. https://doi.org/10.1146/annurev.economics.050708.143235

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), 991–1013. https://doi.org/10.1257/0002828042002561

Camerer, C. F. (2015). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. In G. R. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology* (pp. 249–295). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328325.003.0016

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of development economics* (Vol. 4, pp. 3895–3962). Elsevier. https://doi.org/10.1016/S1573-4471(07)04061-2

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535–538. https://doi.org/10.1126/science.1168244

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. W. W. Norton.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review, 102*(1), 33–48. https://doi.org/10.1017/S000305540808009X

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies, 29*(1), 1–17. https://doi.org/10.1086/468061

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055. https://doi.org/10.1257/0022051043004577

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*(6), 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press. https://doi.org/10.1017/9781108653985

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

List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives, 25*(3), 3–16. https://doi.org/10.1257/jep.25.3.3

Mitchell, G. (2012). Revisiting truism: External validity of social psychological research. *Perspectives on Psychological Science, 7*(2), 109–117. https://doi.org/10.1177/1745691611432343

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*(4), 379–387. https://doi.org/10.1037/0003-066X.38.4.379

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776–783. https://doi.org/10.1037/h0043424

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.