

# The relationship between internal and external validity

Clarissa Busch\*

---

## **Abstract**

The connection between internal and external validity is a matter of some debate in the philosophy of science. Some authors claim that there is a tradeoff between internal and external validity; some argue that internal validity is a prerequisite for external validity. This paper presents an attempt at resolving the tension between these two views. The main contention is that the two views refer to different components of internal and external validity. The tradeoff is reframed as a tradeoff between the degree of confidence that a researcher can attain about the internal validity of her result and the degree of confidence she can attain regarding its external validity. This interpretation of the tradeoff is compatible with the

---

\*Thanks to Donal Khosrowi for his supervision of this paper and to Johanna Thoma for her comments on an earlier version.

prerequisite view. An important implication of this reinterpretation is that the tradeoff relationship places a constraint on experimental design, while the prerequisite relation does not.

---

## 1 Introduction

What is the relationship between internal and external validity? This question is a matter of some debate in the philosophy of science literature. There are two co-existing yet apparently contradictory views of the relation between the two. One is that internal validity is a prerequisite for external validity (see e.g. Hogarth 2005). The other is that there is a tradeoff between the two (see e.g. Campbell 1957). Both views are encountered frequently in the literature on the methodology of applied econometrics; at times both views are even supported by the same author (e.g. Guala 2003). While there is widespread agreement that there is some tension between these views (e.g. Persson and Wallin 2015), there have been few explicit attempts at resolving it. One such attempt is by Jimenez-Buedo and Miller (2010), who conclude that researchers are simply mistaking the difficulty of making predictions about the world based on experiments for a tradeoff between internal and external validity. However, since the tradeoff view is so widespread in the debate about the validity of experimental results, I propose an alternative

attempt to resolve the tension between these two views.

To this end, I begin by disambiguating the relevant concepts, namely internal validity and external validity, in Section 1. Specifically, I will rely on the definition of internal validity given by Guala (2005) and distinguish between an epistemic and an ontic element of internal validity. In Section 2 introduce the two conflicting views, namely the tradeoff and prerequisite view. I present Jimenez-Buedo and Miller's argument in favor of abandoning the tradeoff to solve the apparent conflict in Section 3. In Section 4, I present my own way of resolving the tension between the two views, using the epistemic/ontic distinction. To illustrate my reading of the tradeoff, I apply it to an example. I conclude in section 5.

## **2 The concepts**

### **2.1 Internal validity**

Before I begin, let me briefly lay out what I understand by internal and external validity for the purposes of the subsequent discussion.

It is useful to begin with Guala's (2005) account, who offers the following definition:

“The result of experiment E is internally valid if the experimenter attributes the production of an effect Y to a

factor (or set of factors) X, and X really is a cause of Y in E. Furthermore, it is externally valid if X causes Y not only in E but also in a set of other circumstances of interest F, G, H”

Guala’s definition highlights the importance of the researcher’s knowledge of the causality of the relationship between two factors. In fact, his definition seems to suggest that internal validity is a two-part concept, with an epistemic and an ontic component. On this reading, for an experimental result to be internally valid, it must fulfill an epistemic requirement, i.e. the experimenter must identify a certain factor X as a causal factor of Y, and an ontic requirement, i.e. this factor X must really be a cause of Y. The epistemic component of internal validity is the part that is accessible to the researcher. Thus, for practical research purposes, it is not useful to distinguish between an ontic and an epistemic component of internal validity. Yet, as will become clear in the subsequent discussion, this distinction can help clarify the relation between internal and external validity.

## **2.2 External validity**

Guala’s definition highlights the importance of the researcher’s knowledge of the causality of the relationship between two factors. In fact, his definition seems to suggest that internal validity is a two-part concept, with an epistemic and an ontic component. On this

reading, for an experimental result to be internally valid, it must fulfill an epistemic requirement, i.e. the experimenter must identify a certain factor X as a causal factor of Y, and an ontic requirement, i.e. this factor X must really be a cause of Y. The epistemic component of internal validity is the part that is accessible to the researcher. Thus, for practical research purposes, it is not useful to distinguish between an ontic and an epistemic component of internal validity. Yet, as will become clear in the subsequent discussion, this distinction can help clarify the relation between internal and external validity.

Broad external validity, on the other hand, is the transferability of a result to other contexts in general, not just to a particular “context of interest”, so-labelled by the researcher. In what follows, I will explain the two types of external validity in more detail and examine whether the epistemic/ontic distinction can be applied to them as well. This will be useful in the subsequent discussion of the relation between the two types of validity.

Narrow external validity is also sometimes referred to as “ecological validity” (see e.g. Roe and Just 2009). It refers simply to the applicability of an experimental result to another context that is of interest to the researcher. Narrow external validity can be split up into the same two components as internal validity. The ontic component concerns the question of whether the experimental result under consideration, e.g. a qualitative causal relationship, actually holds in the context of interest. The epistemic component,

in my interpretation, is then the researcher's degree of confidence that the experimental result holds in the context of interest, under the assumption that the effect the experiment identifies is, in fact, causal.

Broad external validity, on the other hand, is both a vaguer and more complex concept than narrow external validity. Broad external validity concerns the general transferability of an experimental result to any number of distinct settings. Specifically, if a result has a high degree of broad external validity, it is highly generalizable, meaning that it holds in many other contexts.<sup>1</sup> In physics, for example, some findings have high broad external validity because they are based on laws of nature, which are typically considered to apply over very broad spatio-temporal domains. In the social sciences, a high degree of broad external validity may often be more difficult to achieve.

---

<sup>1</sup> I use Matthewson and Weisberg's definition of "a-generality" as adapted to experiments, i.e. generality as a measure of how many [actually possible] targets the experimental result applies to. However, since I focus on narrow rather than broad external validity in what follows, I do not elaborate on this concept. For details, see Matthewson and Weisberg 2008.

### **3 The conflict**

#### **3.1 The prerequisite view and the tradeoff view**

Having introduced these crucial concepts, I will now sketch the two conflicting views of internal and external validity. According to Guala, “there exists a trade-off between the two dimensions of experimental validity. The stronger an experimental design is with respect to one validity issue, the weaker it is likely to be with respect to the other. ” (Guala 2005, 144). Even though Guala’s remarks seem somewhat tentative, the tradeoff between internal and external validity has come to be known as a “well-known methodological truism ” (Cartwright 2007, 220).

Interestingly, Guala also seems to hold the other common view of the relationship between external and internal validity, which is often seen as conflicting with the tradeoff view. He writes: “Problems of internal validity are usually chronologically and epistemically antecedent to problems of external validity: it does not make much sense to ask whether a result is valid outside the experimental circumstances unless we are confident that it does therein ” (Guala 2003, 1198). In other words, on this view, internal validity is (usually) understood to be a prerequisite for external validity. There is a clear tension between the tradeoff and prerequisite views. This is because if there is in fact a tradeoff between internal and external validity, while at the same time there can be no external validity

without internal validity, then both relationships cannot be true in all cases. It could be the case, for example, that an experimental result must be internally valid to some degree for external validity to be achievable. But can internal validity be fulfilled “to some degree” or is it a dichotomous property? In the ensuing discussion, I hope to shed some light on the relation between these two types of validity.

While the conflict between the prerequisite and tradeoff views has been documented in the literature (see e.g. Persson and Wallin 2015 and Jimenez-Buedo and Miller 2010), there have been few attempts to resolve it.

### **3.2 Dropping the tradeoff**

Jimenez-Buedo and Miller (from here on JM) make one such attempt, wherein they argue that the two views cannot be supported simultaneously and conclude that the tradeoff view should be abandoned. Their argument proceeds as follows:

The tradeoff view has two implications at least one of which must hold true if there actually is a tradeoff between external and internal validity. Specifically, if there is a tradeoff relationship, then it must be the case that either

- (1) a given experimental setting can be altered such that the results of the experiment become more internally valid at the



expense of external validity or

- (2) it can be altered such that the results gain external validity at the expense of internal validity.

JM then proceed to examine the logical implications of the prerequisite view: If internal validity is a prerequisite to external validity, then

- (3) it is impossible to change an experimental design to both increase external validity and decrease internal validity, and
- (4) an increase in internal validity should either increase external validity or leave it unchanged.

Both (3) and (4) are necessarily true if internal validity is indeed a prerequisite for external validity. However, clearly, it cannot be the case that an increase in internal validity entails both

(1) a decrease in external validity and (4) no decrease in external validity. And it cannot be the case that an increase in external validity entails both

(2) a decrease in internal validity and (3) an increase in internal validity.

However, for both the prerequisite and the tradeoff view to be true simultaneously, both (3) and (4) must be true and one of (1) and

(2) must be true. This means that holding both views to be true inevitably leads to a contradiction, either between (1) and (4) or between (2) and (3). JM thus conclude that the prerequisite view and the tradeoff view cannot be held simultaneously and that researchers might simply be mistaking the difficulty of making predictions about real-world phenomena based on experimental results for a tradeoff between internal and external validity.<sup>2</sup>

## 4 The resolution

### 4.1 Reinterpreting the tradeoff

Since the tradeoff view is nevertheless ubiquitous in the literature on internal and external validity, I propose a different approach to resolving the tension between the tradeoff and prerequisite views. The intuition leading to my approach stems from a question JM pose: “Is internal validity a quality that an experiment either has or not, or can experiments be more or less internally valid?” (p.7).

The tradeoff view implicitly presupposes that internal and external validity come in degrees. This is because even though there could be a tradeoff relation between the two if they took on binary values, this would mean that researchers would have to opt for one of the

---

<sup>2</sup> What JM describe seems to be a relationship of simple attenuation according to Matthewson and Weisberg’s taxonomy of tradeoffs. For details, see Matthewson and Weisberg 2008.

two, completely sacrificing the other, which does not seem to align with the kind of options practitioners face. In the prerequisite view, on the other hand, internal validity might well be a binary concept. If internal validity must be given in order for external validity to be possible, it may well be the case that an experiment either is or is not internally valid. Thus, the question of whether internal validity is scalar or dichotomous promises to illuminate the relation between internal and external validity.

To help resolve this tension, consider again Guala's definition of the two validity concepts and the distinction introduced above. Given the distinction between ontic and epistemic internal validity, it becomes clear that internal validity may in fact be part dichotomous and part scalar. This would mean that there is no genuine conflict between the tradeoff and prerequisite views. To see this, consider whether ontic and epistemic internal and narrow external validity are dichotomous or scalar properties. Ontic internal validity is a dichotomous property. It either is or isn't the case that there is a causal relationship between two variables. On the other hand, the fulfillment of the epistemic condition of internal validity is a matter of degree, specifically degrees of belief or credences. To make this claim more plausible, consider the definition of internal validity. The epistemic condition is fulfilled if "the experimenter attributes the production of an effect Y to a factor X." What does it mean for the experimenter to "attribute" an effect to a particular cause? To attribute the effect Y to a given cause X, the experimenter must be confident that X is the cause of Y. Confidence is a matter of degree.

*The relationship between internal and external validity*

The experimenter can be more or less confident in the validity of the causal inference, i.e. her degree of belief in the validity of the inference can be high or low. Thus, epistemic internal validity seems best understood as a scalar property.

Unlike in the case of internal validity, the ontic component of narrow external validity is not necessarily dichotomous. Rather, it could be the case that factor X is causally relevant for the production of Y both in the experimental context and in the context of interest but that its impact differs in magnitude in the two contexts. Ontic narrow external validity is thus a scalar property. Epistemic narrow external validity is also a scalar property, since it is measured in terms of degrees of confidence. Similarly, broad external validity is determined by scalar properties, such as to how many other contexts a result can be generalized and to what degree the result is similar to the result in the target setting of interest. Thus, broad external validity also seems to be a scalar property.

These answers to JM's question suggest that the tradeoff and prerequisite views might be reframed. If the tradeoff view implies that internal and external validity are measured in degrees, then perhaps this view refers only to the epistemic clause of the internal and narrow external validity definition. In other words, the researcher cannot design her experiment in a way that allows her to be highly confident about both the internal and the external validity of her results.

The prerequisite view, on the other hand, implies that internal va-

validity could be a dichotomous concept. This suggests that the prerequisite view might be compatible with the ontic components of internal and external validity. These components however, focus on a different set of issues. They ask whether there really is a causal relation between the variables under consideration and whether that same causal relation also holds in the context of interest.

This reinterpretation of the tradeoff can resolve the debate about the relationship between internal and external validity because it makes the tradeoff and prerequisite views compatible: It is entirely possible for the epistemic components of internal and narrow external validity to be in a tradeoff relation while the ontic components of internal and narrow external validity are in a prerequisite relation. On this interpretation, researchers face a tradeoff between epistemic access to the narrow external validity of a result and epistemic access to the internal validity of a result. This means that the more they change their design to gain confidence that the observed relation between the experimental variables is actually causal, the more difficult it becomes to extrapolate that result to another context under the (thus not easily verifiable) assumption that it holds in the experimental context.

At the same time, ontic internal validity must be fulfilled (whether the researcher knows it or not) for ontic external validity to be possible. In other words, the relationship between the variables in the experimental context must be causal for the result to be applicable to the context of interest.

To illustrate this reading of the tradeoff, I will apply it to a classic instance of the validity tradeoff, to wit, the tradeoff researchers face when choosing a research design.

#### **4.2 The epistemic tradeoff in “Tradeoffs between Experiments, Field Experiments, Natural Experiments and Field Data”**

In their 2009 paper “Internal and External Validity in Economics Research: Tradeoffs between Experiments, Field Experiments, Natural Experiments and Field Data, ” Brian Roe and David Just describe the tradeoff between different kinds of study designs. They differentiate between four types of studies and order them along a spectrum according to the “degree of verifiable exogenous variation within the economic context that produces the data ” (p. 4). According to them, observational studies have the lowest internal and highest external validity, followed by natural experiments, which have higher internal validity but somewhat lower external validity. Field experiments have still higher internal validity because they allow for manipulation by the researcher, which also further decreases external validity. Laboratory experiments have very high internal validity but are weak as far as external validity is concerned. There is thus a tradeoff between external and internal validity when choosing between these four types of designs.

The tradeoff they describe can be explained in terms of the tradeoff

between the epistemic components of internal and external validity.

In order to obtain a high degree of narrow external validity, a study must be set in a context that is as similar as possible (if not identical) to the context of interest. If it is set in the context of interest, narrow external validity is trivially fulfilled. This is the case in observational studies, where the researcher observes the very subjects she is interested in. On the other hand, internal validity is difficult to establish in observational studies. While the ontic component of internal validity may be fulfilled in such studies, i.e. the observed effect really is caused by the under investigation, the epistemic component cannot be fulfilled as it is difficult to establish causality in an observational study.

At the other end of Roe and Just's spectrum are laboratory experiments. There, the study context is necessarily different from the context of interest. Lab experiments in economics are often conducted to draw conclusions about real-world settings. The context of interest is part of the real world, whereas the study is set in a laboratory. That means narrow external validity is not automatically fulfilled but needs to be established. While it might be the case that the observed effect holds in the context of interest, it is difficult to know for sure that this is the case. In other words, the ontic component of narrow external validity may be fulfilled but the epistemic component is difficult to fulfill in a lab experiment. On the other hand, since the researcher has full control over the laboratory setting, internal validity in lab experiments is high. More precisely, the

researcher can manipulate the experimental setup to gain epistemic access to the internal validity of the result. The epistemic component of internal validity is thus easily fulfilled in lab experiments. Overall, while the internal validity of the results of laboratory experiments is highly epistemically accessible to the researcher, the narrow external validity of the results is not and must be argued for using additional information beyond the study's findings.

On Roe and Just's validity spectrum, field experiments and natural experiments are located between lab experiments and observational studies. Field experiments are closest to lab experiments as far as validity is concerned. They are fairly strong on internal validity. Because the researcher can manipulate the experimental setup to some degree, she gains epistemic access to the internal validity of the results. At the same time, this very manipulation also limits the researcher's epistemic access to the narrow external validity of the results. So while the epistemic component of narrow external validity can be fulfilled to some extent because the experiment is conducted within the context of interest, it cannot be completely fulfilled because the researcher manipulates that context to conduct the experiment, thereby making the experimental context potentially relevantly dissimilar from the (non-manipulated) context of interest.

Natural experiments are closer to observational studies as far as validity is concerned. This is because they are set within the context of interest and involve no manipulation of the experimental context



on the part of the researcher. This means that the researcher has high epistemic access to the narrow external validity of her results, i.e. she can be fairly confident that the observed effect holds in the context of interest. On the other hand, due to the lack of manipulability, the researcher also cannot be fully confident that the observed effect is really causal. The epistemic component of internal validity is thus not so easily fulfilled.

As the preceding description of the validity tradeoff between different types of studies suggests, the epistemic interpretation of the tradeoff between internal and external validity is applicable to this classic instance of the tradeoff.

## **5 Conclusion**

My aim was to clarify the relationship between the internal and external validity of experimental results. In doing so, I first examined the concepts of internal and external validity in detail. From this analysis I concluded that internal validity and narrow external validity might be two-part concepts, each comprising an epistemic and an ontic component. This distinction, proves helpful in illuminating the relationship between the two types of validity. It inspired my reinterpretation of the tradeoff and prerequisite views as referring to different sub-concepts of internal and external validity. On this reading the ontic component of internal validity is a prerequisite for the ontic component of external validity, while at the same time,

researchers face a tradeoff between the epistemic components of internal and external validity. What the researcher is trading off is not internal validity simpliciter, but rather the degree of confidence with which she can attribute the observed effect to the experimental treatment.

As I have pointed out, the distinction between the epistemic and ontic components of internal and external validity by itself is uninteresting to the researcher. What matters in practice is merely what knowledge the researcher can obtain regarding her results' internal and external validity, i.e. their respective epistemic components. This, however, has an interesting implication for practitioners wondering which aspect of the relationship between internal and external validity is relevant to experimental research. The prerequisite relationship between the ontic components of internal and external validity happens "behind the scenes "; it is irrelevant for the researcher because she has little control over whether or not a causal relation actually exists. When designing an experiment, the only thing she can control is her own epistemic access to the internal and external validity of the results, and any attempt to optimize her experimental setup to gain maximal insight into the causality of the observed effect and its transferability to her context of interest will be constrained by the tradeoff. On the other hand, the fact that a causal relationship must exist in the experimental context for the results to be transferable to the context of interest places no constraint on her experimental design, since she cannot control whether or not a causal relationship exists. The relevant constraint regarding

*Clarissa Busch*

the relationship between internal and external validity is therefore the validity tradeoff and not the prerequisite relation.

## References

- [1] Guala, F., 2005. External Validity. In: *The methodology of experimental economics*. Cambridge University Press, pp. 141-160.
- [2] Jimenez-Buedo, M. and Miller, L.M., 2010. Why a trade-off? The relationship between the external and internal validity of experiments. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, 25(3), pp.301-321.
- [3] Hogarth, R.M., 2005. The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, 12(2), pp.253-263.
- [4] Campbell, D.T., 1957. Factors relevant to the validity of experiments in social settings. *Psychological bulletin*, 54(4), p.297.
- [5] Guala, F., 2003. Experimental localism and external validity. *Philosophy of science*, 70(5), pp.1195-1205.
- [6] Persson, J. and Wallin, A., 2015. The (misconceived) distinction between internal and external validity. *Against boredom: 17 essays on ignorance, values, creativity, metaphysics, decision-making, truth, preference, art, processes, Ramsey, ethics, rationality, validity, human ills, science, and eternal life to Nils-Eric Sahlin on the occasion of his 60th birthday*, pp.187-195.
- [7] Cartwright, N., 2007. *Hunting causes and using them: approaches in philosophy and economics*. Cambridge University Press.

- [8] Matthewson, J. and Weisberg, M., 2008. The structure of tradeoffs in model building. *Synthese*, 170(1), pp.169-190.
- [9] Roe, B.E. and Just, D.R., 2009. Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5), pp.1266-1271.