Oxford Development Studies, 2014
Vol. 42, No. 4, 455–472, http://dx.doi.org/10.1080/13600818.2014.961414

Randomisation, Causality and the Role of Reasoned Intuition

KAUSHIK BASU

ABSTRACT *The method of randomisation has been a major driver in the recent rise to prominence of empirical development economics. It has helped uncover patterns and facts that had earlier escaped attention. But it has also given rise to debate and controversy. This paper evaluates the method of randomisation and concludes that while the method of randomisation is the gold standard for description, and does uncover what is here called "circumstantial causality", it is not able to demonstrate generalised causality. Nor does it, in itself, lead to policy conclusions, as is often claimed by its advocates. To get to policy conclusions requires combining the findings of randomised experiments with human intuition, which, being founded in evolution, has innate strengths. Moreover, even non-randomised empirical methods combined with reasoned intuition can help in crafting a development policy.*

JEL Classification: B41, O20, Z18

Preamble

This paper is based on the Hirschman lecture delivered in Mexico City in October 2013. I owe a special debt to Albert Hirschman who invited me to spend a year in Princeton, which opened up a lasting interest in political economy and related social sciences. Following Hirschman's lead, and stealing a favourite expression of his, I learned to trespass on neighbouring disciplines and have never given up on that. I am grateful to have been able to pay a tribute to Albert Hirschman in this way, and deliberately chose an area of wide methodological significance to economics, especially in the context of development policy-making. That is the method of randomised trials and its role in discovering causal links and designing policy. I chose this topic somewhat

Kaushik Basu, The World Bank, 1818 H Street, NW, Washington, DC 20433, USA; Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14853, USA. Email: kbasu@worldbank.org; kb40@cornell.edu

This article was originally delivered as the Albert Hirschman Lecture, at LACEA-LAMES, Mexico City, 31 October 2013.The method of randomisation has been a topic of interest to me for quite a while. During this long period and, more recently, in the course of preparing for the Albert Hirschman lecture and in writing it up as a paper, I have accumulated many debts and would like to thank, in particular, Abhijit Banerjee, Talia Bar, Chris Barrett, Alaka Basu, Karna Basu, Tito Cordella, Augusto de la Torre, Francisco Ferreira, Karla Hoff, Vamsee Krishna Kanchi, Daniel Lederman, Luis-Felipe Lopez-Calva, Mattias Lundberg, Nora Lustig, Celestin Monga, Dilip Mookherjee, Ted O'Donoghue, John Roemer, Neelam Sethi, Vito Tanzi and Merrell Tuck-Primdahl. Finally, I am grateful to two anonymous referees of the journal for their comments and criticisms.

self-consciously because there are things that are easier to do than to explain “how to do”. This paper is about the “how to do” of research.

1. Introduction

Human knowledge owes a lot both to the ability of human beings to think and to their inability to think deeply enough. This is not as paradoxical as it may sound. It implies people typically know less than they think they do. A careful examination of how we acquire knowledge, even of the most scientific nature, compels us to recognise that a lot of what we take to be knowledge is illusory. The present paper is focused on one of the most important and rapidly spreading scientific methods of knowledge acquisition in economics, that of randomised trials.

The method of randomisation has long been one of the instruments of investigation in epidemiology, and it is often treated as the gold standard for making causal inferences. Randomisation and probabilistic inference as a method for acquiring scientific knowledge goes back to the 19th century, at least. An interesting early example of probabilistic inference occurred in the study of telepathy and parapsychology. In an experiment in 1884, Charles Richet wanted to see whether the information acquired by one person drawing and looking at a playing card could be transmitted to others. His study showed that of the 2927 guesses about the card, 789 were correct whereas the expected number of correct guesses if they were completely random would have been 732. This allowed him to conclude that there is a small and error-ridden but, nevertheless, positive transmission of knowledge that occurs between people’s minds. These early parapsychology studies and debates drew in some prominent thinkers including philosophers and economists, such as Henry Sidgwick and Francis Ysidro Edgeworth. For a fascinating account of this early history of the subject, including some proper randomisation studies, see Hacking (1988).[1]

The arrival of randomisation as a systematic method of investigation in economics and, in particular, development economics, is, however, relatively new. But it has arrived here with such gusto that randomisation (and relatedly the use of instrumental variables) has become a major story of our time, giving rise to widespread use around the world by development economists and, in equal measure, debate and discussion about its value and validity.[2]

It will be argued here that the rise of randomisation and randomised control trials (RCTs) is of immense value to economics and to the crafting of economic policy. It gives us insights into history that we did not have earlier. At the same time, many of the claims widely made on behalf of RCTs are exaggerated and invalid. It will be shown that RCTs do not give proof of any universal causality. Provided that we treat as axiomatic that the world is causal, it enables us to establish what will here be called “circumstantial causality”. Some of the help that it provides, in indirect ways, in shaping policy is also provided by non-randomised methods and studies of pure correlations. What RCTs do well, and on this it is indeed the gold standard, is to *describe* large populations over multiple periods. Its true value lies in good description, and the worth of good description is not something to be dismissed (Sen, 1980). To go from this to policy requires the use of intuition. If we deny intuition any ground, RCTs are of no value. But the paper argues that the use of intuition is justified and, hence, RCTs do act as an “aid” to policy-making.

Why do we make the mistake of thinking that studies based on exogenous randomisations, RCTs and even the method of carefully selected instrumental variables give us insights into universal causality? The next three sections try to answer this question.

2. Randomisation and Causality

To understand why we often tend to think that RCTs demonstrate universal causal links, consider two celebrated studies that used randomisation and led researchers to some striking findings—Chattopadhyay & Duflo (2004) and Miguel & Kremer (2004). Chattopadhyay and Duflo raise a pertinent question: does having an elected woman leader to head a local government make a difference to the quality of life in the village or locality? There is, it can be shown, a positive correlation between having women leaders and the quality of provision of public goods, such as water. But, a priori, it would be reasonable to say that causality can run either way. On the one hand, it is possible that women have a greater sense of social responsibility and so once elected they make sure that villagers get good water. On the other hand, it is not difficult to think of the causality running the other way around. In a society such as India's, it is traditional for women to be in charge of getting water for the household. It is arguable that if fetching water takes up a lot of time in a village, women are unlikely to have the time to run for office and so would be unlikely to be the elected head of the local government, the *panchayat*.

It is therefore possible to argue, a priori, that the correlation between the election of a woman leader and the better provision of public goods, such as water, reflects causality running either way. Chattopadhyay & Duflo (2004) sorted this out through a deft use of a natural randomisation exercise.[3] The Indian government had decreed through a constitutional amendment that in one-third of the *panchayats* all over India, selected by lottery, the leader must be a woman. Utilising this random choice of women leaders, they showed that one of the a priori hypotheses, namely, the choice of elected women leaders was caused by the better provision of water, was false. From this, they concluded that it was the choice of a woman leader that caused the better provision of public goods and, in particular, water.

What their study did, virtually beyond question, was to rule out the hypothesis that causality ran from villages with superior provision of water to women being elected leaders. However, it never proved that causality ran in the reverse direction. That impression was created in our minds by the unwitting use of a set of prior possibilities and picking the residual after all other possibilities were ruled out by trials and experiments. To jump from this to the policy conclusion that we should elect a woman to the head the *panchayat* to improve village water supply would be incorrect.

Similarly, consider the seminal work of Miguel & Kremer (2004). In developing countries, there tends to be a correlation between the use of deworming medicine by school children and superior participation in schools by children. Why might this be so? Regular participation in school may teach the children the importance of good hygiene and, in particular, keeping oneself free of stomach worms and so they take deworming medicines. It could be that some households are smarter and smartness means being aware of the importance of deworming and going to school regularly. It could also be that deworming makes children fit and fitness enables them to attend school regularly. Miguel and Kremer used an excellently designed RCT. They administered deworming medicine to children in a number of randomly selected schools in Kenya[4] and found that participation rates of students improved. This clearly ruled out the first two causal hypotheses.[5]

It should be clear that in neither of the two studies just discussed is causality established. Based on these findings, we cannot assert that electing a woman leader in a Tamil Nadu village will improve water provision there and administering deworming medicine to all school children in a school in Kenya in 2014 will improve school attendance. All these two RCTs do—and RCTs are ideal for this—is to rule out certain possible causal links. They give the impression of having pinned down causality only because we carry certain prior assumptions about causal links and when all but one of our prior assumptions are ruled out, our mind jumps to the conclusion that causality has been established[6] and we feel that we are now ready to use these findings to design actual policy interventions.[7]

To be aware of our own minds' propensities, it is at times useful to take cognisance of the very different beliefs found in other cultures, which those cultures consider perfectly reasonable. The Gurung tribesmen of Nepal spend a lot of time climbing dangerously high trees to gather honey. When the French photographer, Eric Valli, who worked extensively among the Gurung, asked them whether they ever fell down from their high perches, they replied: "Yes, you fall when your life is over" (National Geographic, vol. 193, no. 6, June 1998: p. 92).

The above argument warns us that these "conclusions" about causal links and policy-making can be wrong. By the use of examples and a simple argument, I now show they are wrong.

Suppose researchers come to a town and do an RCT on the town population to check whether the injection of a green chemical improves memory and has adverse side effects. Suppose it is found that it has no side effects and improves memory greatly in 95% of cases. If the study is properly done and the random draw is truly random, it is likely to be treated as an important finding and will, in all likelihood, be published in a major scientific journal.

Now consider a particular woman called Eve who lives in this town and is keen to enhance her memory. Can she, on the basis of this scientific study, deduce that there is a probability of 0.95 that her memory will improve greatly if she takes this injection? The answer is no, because she is not a random draw of an individual from this town. All we do know from the law of large numbers is that for every randomly drawn person from this population the probability that the injection will enhance memory is 0.95. But this would not be true for a specially chosen person in the same way that this would not be true of someone chosen from another town or another time.

To see this more clearly, permit me to alter the scenario in a statistically neutral way. Suppose that what I called the town in the above example is actually the Garden of Eden, which is inhabited by snakes and other similar creatures, and Eve and Adam are the only human beings in this place. Suppose now the same experiment was carried out in the Garden of Eden. That is, randomisers came, drew a large random sample of creatures, and administered the green injection and got the same result as described above. It works in 95% of cases. Clearly, Eve will have little confidence, on the basis of this, to expect that this treatment will work on her. I am assuming that the random draw of creatures on which the injection was tested did not include Eve and Adam. Eve will in all likelihood flee from anyone trying to administer this injection to her because she would have plainly seen that what the RCT demonstrates is that it works in the case of snakes and other such creatures, and the fact that she is part of the population from which the random sample was drawn is in no way pertinent.

Indeed, and the importance of this will become evident later, suppose in a neighbouring garden, where all living creatures happen to be humans, there was a biased-sample (meaning non-random[8]) trial of this injection, and it was found that the injection does not enhance memory and, in fact, gives a throbbing headache in a large proportion of cases, it is likely that Eve would be tempted to go along with this biased-sample study done on another population rather than the RCT conducted on her own population in drawing conclusions about what the injection might do to her. There is as little hard reason for Eve to reach this conclusion as it would be for her to conclude that the RCT result in her own Garden of Eden would work on her. I am merely pointing to a propensity of the human mind whereby certain biased trials may appear more relevant to us than certain perfectly controlled ones.

We shall return to this later. For now what is important to note is that even if a result is proved by a proper randomised controlled trial conducted on a population to which you belong, you have no reason, on the basis purely of this study, to believe that its finding will apply to you.

The same reasoning carries over when one moves from a finding based on one population to another population. The fact that the choice of a woman leader of the village council in Bengal and Rajasthan led to the better provision of public goods does not mean that you have reason to expect this will be true in Bihar, Bahia or Bahrain, or, for that matter, in Bengal and Rajasthan tomorrow because tomorrow's population is a different one. At least, if we were as fastidious about deducing conclusions as the practitioners of RCTs want us to be, we could not draw those conclusions about other places and other times. That would be like drawing lessons from biased samples. This, in an interesting converse way, is related to an important problem noted by Deaton (2010). Even if we learn the facts from properly controlled randomised trials, if the intervention is then to be made not on a random draw from the population but on a specially selected group, we should be prepared for failure.[9]

3. A Limited Causality Claim

The claim that can be put forward, but only after some prior assumptions are made, is that randomised trials can establish certain "circumstantial" causalities. The prior assumption that we need is to accept the determinist claim that the universe proceeds by causality and so the future that lies ahead of us is as determined as our history. To state this more precisely, assume time progresses in pixels, instead of a continuum, so that for each point of time, there is a well-defined immediate previous point of time. Let X be the set of all possible states of the world at all possible points of time. Further, X is a Cartesian product of n sets, $X_1, X_2, \ldots, X_n$, where each of these sets consists of all possible states of some particular feature of the world. Thus, X_1 could consist of positions of the sun; X_2 could be states of the weather, etc. Hence, at any point of time, an $x \varepsilon X$ is a full description of the state of the world.[10] Of course, n will be a very large number.

In its minimal form, the determinist claim says that if at two points of time, t and k, the world is in state $x \varepsilon X$, and at time $t + 1$, it is in state $y \varepsilon X$, then at $k + 1$ it must also be in state y. In other words, being at x causes the world to be at y in the next instant.

This simple axiom is untestable but has many implications. It means, for instance, that if at two consecutive points of time the world is in the same state, then from then onwards it will forever remain in that state. It follows that, since the world recently has been changing, it must have always been changing. This in itself does not mean that the world never comes back to the same state. Put simply, it is not that history never repeats itself but either history never repeats itself or it repeats itself endlessly. The determinist axiom also has large implications for our attitude to crime, responsibility and punishment, and much has been written by philosophers on this. But that is not my concern here.

To see what RCTs show, let me define the Cartesian product of $X_2, X_3, \ldots, X_n$ by Z. What RCTs show is that there exists some $z \, \varepsilon \, Z$, such that if we have the world in state $(x, z) \, \varepsilon \, X$ instead of $(y, z) \, \varepsilon \, X$, the world in the next period will be in state $a \, \varepsilon \, X$ instead of state $b \, \varepsilon \, X$. This is like saying, other things being the same (that is, z), if you vaccinate people, in the next period, there will be no influenza. But if you do not vaccinate them, there will be influenza. If we accept the determinist axiom, as many do,[11] then this demonstration means that whenever we switch from (y, z) to (x, z), the world will switch in the next period from b to a. It is the "whenever" that makes this a causal claim. This is what I am referring to as "circumstantial causality". Given a certain set of circumstances, changing y to x has a predictable consequence.

The discovery of circumstantial causal connections, as has happened with the rise of RCT studies, is valuable and, at the same time, of limited consequence, more so than the proponents believe. On the one hand, RCTs have given us numerous valuable descriptions of what happened in the past and numerous instances of causes in the past (provided of course that one is willing to accept the determinist axiom). On the other, what they show is very limited. This is because when they show that it was the switch from y to x that caused the switch from b to a, what they are saying is that this was true under certain historical conditions (z), but they cannot tell you what those historical conditions are. RCT discoveries never graduate from something "was a cause" of something else to something "is a cause". RCTs give us no insight into universal causality because they cannot tell us what it was that was being held constant (z in the above example), when we switched some intervention b to a. For Bengal, in a certain period, electing a woman leader of the local government caused water provisioning to be better. This is no guide to the future because we do not fully know what Bengal in a certain period is like. Henceforth, a reference to causality without a qualifying epithet should be taken to be a reference to universal causality because for policy purposes, that is what is of essence.

One clarification is called for here. When we hold the z constant and change y to x, in the above example, why do we say that what is being held constant is not known and that we simply know that *something* is being held constant? In the Chattopadhyay & Duflo (2004) example, for instance, the result is true for Bengal of a certain period. Why is this not a useful description? The reason is that this is a "dated description"— one in which the date or period of the experiment is a part of the description of what is being held constant. Because a date never recurs, we can never use this description in the future. So we will never be able to say that the same conditions now hold and so the same result is to be expected (recall the determinist axiom). What is needed, at a minimum, is a description of what is being held constant when a particular experiment is being run without reference to a date. RCTs do not give us this and this is what makes the findings of RCTs non-portable.

Cartwright (2010) is right in terms of her reservations about RCTs. However, contrary to Cartwright's claims, I am arguing that RCTs do provide insights into circumstantial causality and are the gold standard for *describing* large populations over time. At the same time, however, they fail to demonstrate any form of universal causality. They show us that by the use of the law of large numbers, we can describe the average characteristics of a large population and changes over time, by appropriately studying a small sample drawn from the population. RCTs do this extremely well, though even here one should add the reminder that average characteristics are not the only pertinent features of populations.

In other words, what I have tried to demonstrate is a critique similar to that of Cartwright (2010, 2011) but with a significant difference. On description and on historical causal connections (without any statement of what it is that needs to be held constant for the causal link to work), RCTs are quite exemplary; but on universal causality, which is the main claim of the practitioners of the method of randomisation and which is the crux of using research to formulate policy, RCTs get no marks. On universal causal connections that tell us how to design policy interventions, the method of randomised controlled trials is silent.[12]

4. Learning from Success *and* Failure

The advantage of randomised experiments in describing populations creates an illusion of knowledge, which is not always easy to unravel. This happens because of the propensity of scientific journals to value so-called causal findings and not to value findings where no (so-called) causality is found. In brief, it is arguable that we know less than we think we do.

To see this, suppose—as is indeed the case in reality—that thousands of researchers in thousands of places are conducting experiments to reveal some causal link. Let us in particular suppose that there are numerous researchers in numerous villages carrying out randomised experiments to see whether M causes P. Words being more transparent than symbols, let us assume they want to see whether medicine (M) improves the school participation (P) of school-going children. In each village, 10 randomly selected children are administered M and the school participation rates of those children and also children who were not given M are monitored. Suppose children without M go to school half the time and are out of school the other half. The question is: is there a systematic difference of behaviour among children given M?

I shall now deliberately construct an underlying model whereby there will be no causal link between M and P. Suppose Nature does the following. For each child, whether or not the child has had M, Nature tosses a coin. If it comes out tails the child does not go to school and if it comes out heads, the child goes to school regularly.

Consider a village and an RCT researcher in the village. What is the probability, p, that she will find that all 10 children given M will go to school regularly? The answer is clearly

$$p = \left(\frac{1}{2}\right)^{10}$$

because we have to get heads for each of the 10 tosses for the 10 children.

Now consider n researchers in n villages. What is the probability that in none of these villages will a researcher find that all the 10 children given M go to school regularly? Clearly, the answer is $(1-p)^n$.

Hence, if $\varphi(n)$ is used to denote the probability that among the n villages where the experiment is done, there is at least one village where all 10 tosses come out heads, we have:

$$\varphi(n) = 1 - (1 - p)^n.$$

Check now that if $n = 1$, that is, there is only one village where this experiment is done, the probability that all 10 children administered M will participate in school regularly is $\varphi(1) = 0.001$. In other words, the likelihood is negligible.

It is easy to check the following are true:

$$\varphi(100) = 0.0931,$$

$$\varphi(1000) = 0.6236,$$

$$\varphi(10\,000) = 0.9999.$$

Therein lies the catch. If the experiment is done in 100 villages, the probability that there exists at least one village in which all tosses result in heads is still very small, less than 0.1. But if there are 1000 experimenters in 1000 villages doing this, the probability that there will exist one village where it will be found that all 10 children administered M will participate regularly in school is 0.6236. That is, it is more likely that such a village will exist than not. If the experiment is done in 10 000 villages, the probability of there being one village where M always leads to P is a virtual certainty (0.9999).

This is, of course, a specific example. But that this problem will invariably arise follows from the fact that

$$\lim_{n\to\infty} \varphi(n) = 1 - (1 - p)^n = 1.$$

Given that those who find such a compelling link between M and P will be able to publish their paper and others will not, we will get the impression that a true causal link has been found, though in this case (since we know the underlying process) we know that that is not the case. With 10 000 experiments, it is close to certainty that someone will find a firm link between M and P. Hence, the finding of such a link shows nothing but the laws of probability being intact. Yet, thanks to the propensity of journals to publish the presence rather than the absence of "causal" links, we get an illusion of knowledge and discovery where there are none.

One practical implication of this observation is that it spells out the urgent need for a *Journal of Failed Experiments*, or at least a publicly available depository of such experiments so that we know how many people found a relationship between M and P and how many did not.[13] This will not solve all problems, because arguably, in each of the villages where researchers were looking for the relationship between M and P, it is possible that M and P stood for different things from the things others were looking for in other villages. If each underlying process entails the toss of a coin as in the above description, then we will not have evidence on the same test in many places, but different tests in different places. Nevertheless, it is important to keep track of failed experiments, in the sense of experiments that did not reveal any link between two variables. Such a journal, there can be little doubt, will have a sobering effect on economics, making evident where the presence of a result is likely to be a pure statistical artefact.

5. Knowledge, Evolution and the Induction Principle

Despite the scepticism of the previous sections, the fact remains that human beings know a lot. We "know" when we release a ball in mid-air, it will move downwards; we "know" what time the sun will rise tomorrow; we "know" that a virus is the source of the common cold. Of course, all this could be the luck of the draw, implying that all this supposed knowledge worked thus far, but will not tomorrow. This cannot be ruled out formally, but intuition suggests there is more to these bits of knowledge than pure luck. At any rate, that is the presumption under which I will work. Human beings know a lot about nature or, more minimally, our minds are somehow synchronised with nature. We tend to observe patterns in nature, where indeed there are patterns. Most of what our mind knows it does simply by picking up knowledge along the way without pausing to think whether the knowledge has scientific basis and whether it will stand the test of randomisation.

A child soon learns that a person rolling over with laughter is happy, the woman weeping quietly in a corner is sad, the man with ruffled hair running towards another man with an open knife is likely to do harm. If we acted as custodians of knowledge by randomisation and stopped the child at each stage and asked the child to reject these pieces of knowledge until they had verified them by conducting RCTs, that is, by making sure that randomly selected persons rolling over with laughter were happier than others, and so on, then it is likely that the child's cognitive abilities would be deeply damaged.

If we truly paused to think, we would recognise that the proportion of knowledge that we absorb informally is substantially greater than what we know based on scientific enquiries. To dismiss the former out of hand would greatly deplete our knowledge.[14]

As the example of the Garden of Eden, above, shows, in acquiring knowledge there is scope and indeed need for a catholicity of methods because our intuition can help fill the gaps. A recent study (Paul & Dredze, 2011) shows that there is an amazingly strong correlation between the incidence of flu in the USA and the extent of chatter related to the flu on twitter. Seeing this, most of us will tend to assume that it is not the chatter that causes the flu. This would seem to me to be the right conclusion, and there would be no reason to insist that this be established by conducting proper randomised trials.[15]

But how do children pick up knowledge from such ascientific processes? While there is no clear answer, one incomplete answer relates to evolution. Only those human minds that are in reasonable synchrony with nature and can glean knowledge from experience have survival value. Hence, over long stretches of time, minds with this ability have been selected, and we, at the start of the 21st century, are lucky to be endowed with such ability.[16]

What will be argued here is that we have no choice but to use this faculty of ours to go from experience—both statistically sound experience based on RCTs and experience based on biased samples—to the crafting of policy. There is actually no other option. There is a lot of discussion in the literature on beginning with experiments and then going on to check "external validity". But to imagine that there is a scientific way to achieve external validity is, for the most part, a delusion. As we saw in the previous sections, RCTs do not in themselves tell us anything about the traits of populations in other places and at other times. Hence, no matter how large the population from which we draw our random samples is, because it is impossible to draw samples from tomorrow's population and all policies we craft today are for use tomorrow, there is no "scientific" way to go from RCTs to policy. En route from evidence and experience to policy, we have to rely on intuition, common sense and judgement.[17] It is evidence coupled with intuition and judgement that gives us knowledge. To deny any role to intuition is to fall into total nihilism.[18]

Some critics have pointed out that the method of experiments and random trials deflects us from the important aim of explaining "why" certain regularities in nature occur instead of merely finding such regularities. This is an important observation made by very prominent commentators in this debate (see, for instance, Heckman & Smith, 1995; Deaton, 2010) and answering the "why" is indeed an important ingredient of human understanding. However, I would argue that as a critique of RCTs, this is invalid (see also McKenzie, 2012). To see this, let me point out, as will become evident later, that some of our so-called understanding is delusional. It is based on the mind observing regularities and, after some time, treating these regularities as facts of nature. Then, when we see something new happen and say that we have understood it, it is because we have managed to fit this new observation into the partly solved jigsaw puzzle where the other parts are provided by our previous observation of regularity and our propensity to treat those as immutable parts of nature. To the extent that RCTs feed our minds with more and more observed and statistically verified regularities, they actually help with our understanding of why certain things happen, instead of deflecting us from such an understanding. So this arrow, despite having been shot by such shrewd observers, does not slay the claims made on behalf of the method of randomisation. But there are other arrows.

The first ingredient required to go from past regularities to expectations about the future is the "induction principle". Broadly, this says that if we see something that has happened repeatedly in the past, such as the sun having risen every day, we have reason to expect this will happen in the future. We have no escape from using the induction principle in acquiring (falsifiable) knowledge.[19] That qualifier, "falsifiable", is necessary because there are certain kinds of deductive knowledge that do not require any evidence or data. This is true, for instance, of the Pythagoras theorem on right-angled triangles, or Arrow's Impossibility theorem on voting. The knowledge of these results does not require facts and so is independent of the principle. But, in this paper, I am keeping such non-falsifiable knowledge aside and focusing attention exclusively on knowledge which is, in principle, falsifiable.

The problem with the induction principle, so essential for this kind of knowledge, is that it is inherently fuzzy, and it is possible for intelligent people not to accept it. This stems from the fact that the validity of the induction principle depends on our priors, and priors, by their very nature, cannot be proved.[20]

Let us see what kinds of priors get us to the induction principle. Suppose you have the prior that Nature sits with an urn containing a very large number of balls which could be

(1) all white, (2) all black and (3) some white and some black. You believe that just before dawn Nature picks a ball. If it is white, the sun is made to rise and if it is black, the sun is stopped from rising. Now consider your first dawn. If the sun rises, you know (2) cannot be valid. Then as the sun continues to rise every morning, you gradually begin to rule out (3) and base your expectations on (1). The more ample the evidence is, the stronger your belief becomes that the sun will rise tomorrow as it has done in the past.[21]

But consider a person with different priors for instance, someone who believes either (2) or (3), above, is true. Now when the sun rises the first day, (2) is ruled out. Then as the sun rises each morning, her prediction with every passing day will go the other way around. She will become more and more firm in the expectation that the sun will not rise tomorrow.[22] She will reject the induction principle and there is no way to persuade her that she is wrong. It is my prior versus hers.[23]

The induction principle cannot of its own reach definite conclusions. As Manski (2013, pp. 30 and 31) observes, "The logic of inference does not enable any conclusions about future or hypothetical situations to be drawn based on observed tendencies per se. Assumptions are essential". This inherent ambiguity of the induction principle and the need for tempering it with assumptions or human judgement opens us to the risk of "anything goes". Human beings have often been subject to blind faith, strange delusions of knowledge and superstitions, and large mistakes in decisions have been made and continue to be made because of this. Hence, it needs to be emphasised that the method of using the best available evidence and intuition to derive policy has to guard against this risk. Just as we have intuition, we also have reason. While, as I have tried to show, there is no getting away from the use of intuition, we need to harness reason to vet our intuition. What we have to rely on is best described as reasoned intuition.

6. Reasoned Intuition

Making good policy entails the use of the best available evidence. But to do this right we need reason and that is often a stumbling block. Moreover, the fact that we should use the best available evidence does not imply that policies must invariably be based on evidence. To see this, consider the common criticism whenever someone proposes a new policy. Assume that a person recommends a new policy intervention, called X, without providing any evidence on whether or not X works. X could be a new way to make conditional cash transfers to the poor. It is common in such a situation to find critics who will oppose this policy X on the ground that there is no evidence as to whether or not X works. What these critics do not realise is that such a criticism is self-contradictory and thus invalid.

To understand this, let "not doing X" be called "doing Y". Clearly, if there is no evidence as to whether or not X works, there is no evidence as to whether or not Y works. If it is wrong to implement X if there is no evidence as to whether or not X works, it is wrong to implement Y if there is no evidence as to whether or not Y works. But because Y is the negation of X, it is meaningless to say that we should not implement X or Y. This contradiction establishes that the initial criticism must be invalid.

This is simply a cautionary tale about the ubiquity of unreason and the errors into which it leads us. Of course, the best available evidence must be used, and we must learn from past patterns and regularities about what to expect in the future. What is often not recognised is that when people try to follow the widely used dictum, "policies should be evidence based”, the hardest part of the rule is contained in the word “based”. This is the reason why we so often hear glib references to some evidence and then a hand-waving switch to one’s favoured policy, ignoring the fact that the policy should be *based* on the best available evidence.

But how do we make this transition from evidence to policy? As I argued earlier, our intuition has something innately valuable in it. This may well be an outcome of evolution. If human intuition were completely out of line with nature and read patterns wrongly, presumably it would have been weeded out by the process of natural selection. However, as the above example shows, our pure intuition and gut feeling may not be quite flawless. They need to be held under the scanner of reason before we use them to translate experience and evidence into rules of behaviour and policy. I use the expression “reasoned intuition” to refer to intuition after it has been subjected to thought and scrutiny by the other faculty that we all possess, namely, reason. It is not possible to give “reasoned intuition” a hard definition. Its importance is founded on the realisation that our intuitions are not flawless. Major blunders have been committed in the past by humankind on the basis of intuition and a refusal to subject intuition to scrutiny.

Fortunately, we also possess the ability to reason. What I am arguing is that by bringing these two faculties—intuition and reason—together, we can greatly improve our ability to make policies. Is my intuition consistent with the evidence we have and the other things I already know? Is it consistent with the other intuitive beliefs I hold? Reasoned intuition is intuition that has been subjected to this kind of interrogation.

Indeed, a large part of the present paper tried to demonstrate the value of the use of reasoned intuition. Empirical studies, no matter how carefully done, cannot in themselves lead us to particular policies. Going from evidence, data and statistics to policy will invariably entail a leap of imagination. This is facilitated by our intuition, honed over millennia by the processes of evolution. But we must stand ready to subject that intuition to reason. This cannot be stated as a hard rule; but to ignore it on that ground would be a mistake.

I can illustrate some kinds of common mistakes that our intuition often leads us into. The remedy is not to discard the use of intuition but to pause and check that it stands the test of reason. One empirical finding which has been at the base of a lot of development policy discussion is the role of growth in mitigating poverty. Dollar *et al.* (2013) have recently shown, using a remarkably comprehensive data-set spanning multiple countries and several decades, that the bulk of poverty reduction that the world has seen in the past was due to overall GDP growth. Intuition then prompts us to propose that we should therefore rely on growth rather than specially designed market interventions to battle poverty. For the untutored, the growth versus poverty debate is an emotive one and this observation quickly gathers support among those who have a predilection to leave it all to the market. Without going into this larger debate here (see Basu, 2013, for such a discussion), let me simply point to an obvious mistake that leads us to draw this policy conclusion from this robust empirical finding. The fact that 75% of the drop in poverty in the past occurred because of growth in no way shows that growth is more effective in eradicating poverty than, say, a new conditional cash transfer programme. The mistake in making that deduction would be the same as someone studying infections in the 1930s asserting that we should rely on non-penicillin medicines because past data shows that 99% of all cures were because of non-penicillin drugs. This ignores the fact that penicillin was discovered in 1928 and so the lack of evidence of the success of penicillin is not a sign of penicillin not working but of penicillin not existing.

In other words, if we want to use the induction principle to decide whether policy X is likely to achieve a certain result R, it is not enough to show that R was achieved in the past because of Y, and from this conclude X will not work. It is important to check whether X was tried in the past and then make the deduction. In case X never achieved R because X was never tried, we do not have evidence on the efficacy of X one way or the other. To assume otherwise is to make the same mistake as the one discussed at the start of this section, whereby the lack of evidence either in favour of or against a new policy is taken to be a verdict against the new policy. As was pointed out earlier, the induction principle does not have a precise definition; but, upon cogitation, we can nevertheless conclude that supposedly inductive conclusions are wrong. It is this which makes the inductive principle useful.

The mistake about growth and poverty just discussed is the same one that we so often find being made in arguments over job creation. People often assert that we have to rely on the private sector for job creation because past data show that 80% of jobs were created by the private sector. If this logic were correct, we would also have to accept the logic of an economist in Soviet Russia in 1980 arguing that we have to rely on the state for job creation because past data show that 90% of all jobs were created by the state.

These are all examples of using our intuition wrongly to go from induction to certain policy prescriptions. Because the induction principle is not one that can be formally defined, it is not always possible to define mistakes deriving from the principle formally. Yet, if we pause to reason carefully or listen to someone else provide the reasoning, we would, on our own, back off from many a wrong policy. This is what reasoned intuition is all about.

What is the role of theory in all this? It is common to pay lip service to this and assert that we need both theory and data to make policy prescriptions. There are many instances where we do not need any theory. Spotting regularities in data, coupled with reasoned intuition, can take us to useful policy prescriptions. There can be no need for theory. Likewise, there are areas where theory, coupled with what we intuitively know, can take us to useful policy decisions.

The real role of theory is to make consistency checks on our intuitive beliefs[24] and, based on our intuitive beliefs, to take steps towards a deeper understanding using pure deduction. Can all the beliefs we hold be valid together? Theory is a method for checking this. It also helps us take deductive steps in conjunction with what we know, whether through formal empirical studies or simply through the informal acquisition of knowledge. The role of theory and empirics can be illuminated through an analogy with Sudoku. Suppose we have a slightly different Sudoku game in which the numbers that are typically entered in the squares to start with are not shown explicitly but have to be found out using clues (look under a pillow for the number that goes into the left-hand square). The search for these pre-existing but hidden numbers is akin to empirical investigation. However, once we have found out all the numbers specified in advance, the rest of the Sudoku proceeds by pure reasoning. All the other numbers have to be deduced rather than found. This is akin to what theory does. It enables us to move on using the given facts and by using pure deduction.

7. Concluding Remarks

The rise of the method of randomisation has been a major stimulus to empirical development economics and, as such, it is a welcome advance. Akin to what this method did earlier for epidemiology, it has in its short life in economics brought to light several new facts that were earlier languishing unseen. For this, it deserves credit and justifies further use. However, the claims made by its proponents concerning its ability to reveal causal links that lead directly to policy conclusions are wrong. That is what this paper has tried to argue. RCTs have no advantage on these scores. What RCTs do usefully and well is to describe. They give us ways to describe static features and also temporal features (M happened in period 1 for some members of the population and P happened in period 2 for 90% of those to whom M happened and to 50% of those to whom M did not happen) with no clues to universal causality. They can reveal circumstantial causality, namely, that, in certain historical circumstances, doing x instead of y, led to a instead of b, in the next period, but because we do not have a full description of the historical circumstances, there is, strictly, no lesson in this which is portable through time.

Universal causality is a mental construct for observers, something that helps them apply intuition and judgement, in conjunction with what they know or think they know, to decide how to behave and which policies to choose. What human beings know comes from many sources, and to deem only one method valid and all others invalid is to slow the process of knowledge acquisition. The catholicity of methods currently used—from anthropological notes, analysis of large data-sets, everyday experience and randomised trials—all have a role to play in this enterprise. When it comes to seeking good descriptions of past facts, RCTs constitute the best method to aspire to. When it comes to the search for universal causality and best policy, there is a role for multiple sources, including reasoned intuition, honed over millennia of human evolution. All this should be tempered with a shot of scepticism—an awareness that for all our best efforts, we may be wrong.

Notes

[1] As must be obvious from this paragraph, what constitutes a proper randomisation experiment is not beyond dispute. Even in modern development economics many variants, around the broad idea of randomisation, have been tried and critically evaluated (see Bruhn & McKenzie, 2009). Further, there are some engaging historical studies, for instance, by Wantchekon *et al.* (2013), of the long-run impact of setting up regional schools, where the randomisation is, unavoidably, less than perfect but nevertheless an attempt is made to capture this broad idea and draw insights.

[2] See, for instance, Banerjee (2005), Banerjee & Duflo (2009, 2012), Bardhan (2013), Barrett & Carter (2010), Basu (2005, 2011), Cartwright (2010), Cartwright & Hardie (2012), Deaton (2009, 2010), Duflo *et al.* (2008), Elbers & Gunning (2013), Heckman & Smith (1995), Heckman & Urzua (2010), Imbens (2010), Mookherjee (2005), Ravallion (2009) and Rodrik (2008).

[3] This rightly celebrated paper has spawned other papers shedding interesting light on the status of women in some parts of India (for instance, Beamen *et al.*, 2009; Iyer *et al.*, 2012).

[4] In this brief statement I am, of course, leaving out important details of the experiment.

[5] These are just two papers I am using for my argument. But RCTs have given us a rich haul of findings on past regularities in a range of domains in development economics: see, for instance, Ashraf *et al.* (2006), Beasley & Huillery (2013), Benhassine *et al.* (2013), Bernard *et al.* (2013), Habyarimana & Jack (2012), Jensen (2012), Hoff & Pandey (2006) and Ifcher & Zarghamee (2011).

[6] The reason magic surprises us has roots in a similar phenomenon of the human mind. The skilful magician instils in us certain prior beliefs about how a rabbit may have got into a hat. Then he or she demonstrates that none of these are possible and when the rabbit is pulled out of the hat, we are left wondering if this was a miracle.

[7] The journey from (allegedly) finding causes to designing policy is fraught with dangers and is the subject matter of considerable writing (see, for instance, Basu, 2000; Cartwright, 2007; Das *et al.*, 2011). I shall return to this later.

[8] For instance, the trial was done early in the morning and so, inadvertently, only early risers were chosen for the test.

[9] A stark example of this is provided by Worrall (2007). The drug benoxaprofen, for arthritis and muscular skeletal pain, passed the test of randomised trial very well. But when the drug was tried on actual patients it was found to cause a disproportionate number of deaths. What later became clear is that this is a drug that is not used on a random sample of people, but pre-dominantly on older persons and they are the ones most likely to get an adverse reaction. This powerful argument is also discussed and elaborated upon by Deaton (2010).

[10] It should be clarified that the time when a state exists is not a feature of the state and so there cannot be a reference to the temporal position of the state.

[11] I personally do believe in the determinist axiom, while remaining prepared to be corrected (Basu, 2000).

[12] Causality in science and economics has long been the source of philosophical debate and skirmish: see, for instance, Sen (1959), Hicks (1979), Hacking (1975), Skyrms (1980) and Hoover (2001). In its purest form, the existence of causality simply means that if there were two worlds which were identical just before time *t*, they cannot be different at time *t*. If one accepts this as an axiom, as many prominent philosophers through the ages have done, one adopts a "deterministic" view of life. Among economists, Amartya Sen took a similar line in one of his earliest writings— Sen (1959). I should hasten to add that though he has not subsequently written on this directly, there is some smoking-gun evidence from his other writings that his views on this have changed.

It may appear at first sight that this pristine form of causality does not have any implication for what we do and how we behave and so would be dismissed at least by logical positivists as a debate of no consequence. This is however not right because whether or not one is a determinist has important implications concerning the moral responsibility of individuals for their actions, with implications for systems of justice and the punishment that is meted out to those whose actions are found unacceptable to society.

[13] To give due credit, there are some existing initiatives, such as the Campbell Collaboration (http://www.campbellcollaboration.org/), which have tried to create depositories and compile systematic reviews of experiments done, whether or not they yielded publishable results, as long as they conformed to some scientific standards.

[14] This also implies that when designing an RCT, ignoring the informal knowledge of the local people is not as costless as might appear at first sight (Barrett & Carter, 2010).

[15] One can go further and argue that even without correlation, the pure act of measurement facilitates understanding. Gates (2013) gives numerous examples, though without going into why measurement works. It is arguable that measurement makes it easy for our minds to process information and develop an intuitive and heuristic understanding, the role of which must not be minimised (Gilovich *et al.*, 2002; Gigerenzer & Gaissmaier, 2011).

[16] I tried to develop this idea in Basu (2000), utilising the prior work of Lorenz (1977).

[17] As I venture into this controversial territory, it is useful to marshal some support from Leonard Savage: "I believe that here, as elsewhere, catastrophe is avoided, primarily because in practical situations commonsense generally saves all but the most pedantic of us from flagrant error" (Savage, 1954, p. 1).

[18] Total nihilism regarding knowledge is not a line to be dismissed out of hand and several prominent philosophers have taken such a line and there are probably even more philosophers who have taken such a line than we know, because it is reasonable to expect that many of them would not write anything and maybe not even speak about it.

[19] Philosophers have long battled with analysing and trying to understand induction (for instance, Mill, 1843; Bunge, 1959; Skyrms, 1980). For an historical account, see Hacking (1975).

[20] Moreover, as John Stuart Mill had noted in the mid-19th century (1843), the induction principle, unlike deduction, does not lead us to apodictic certainty.

[21] Russell (1912), in an essay on induction, has made this very clear with the example of the chicken trained in the art of induction who expects to have its neck patted every day only to have it wrung one fine morning. This risk will always be there but one can amass a greater and greater amount of data from the past to provide strength to the inductive process. Thus, we may have begun by arguing that the

sun will rise tomorrow because it has done so in the past. But if we could base this on the larger body of evidence that we now have about spinning bodies continuing to spin, we will have an even longer database on which to base our induction. In some sense, this is the heart of the scientific enterprise.

[22] If her prior is that the urn contains a large and finite number of balls, she will not be able to compute the probability that the sun will not rise tomorrow but will be sure with each passing day that the probability of the sun not rising will become larger. If the urn has z balls and w of those are white, then after the sun has risen for y days, her estimated probability of the sun not rising the next day will be $\eta \equiv (z-w)/(z-y)$. Clearly, as y increases, so does η.

[23] There are many examples where a blind adherence to the induction principle does not work. Folklore attributes the example that follows to Galileo. Modernising the example a little, suppose a man, rooted to the induction principle and impatient about wasting time understanding calendars, asks his neighbour each morning, "Is today 1st November 2013?" and gets the answer "no". This continues for days, months and years and he finally predicts: "No morning will be 1st November 2013". But of course he will, as I pointed out when delivering this paper as the Albert Hirschman lecture on 31 October 2013, "get a big surprise tomorrow".

This is a light-hearted argument but cannot be dismissed out of hand. This argument alerts us to the fact that science, which is so rooted in induction, should be taken with a dose of scepticism. This is my one argument with Dawkins (2006). The scepticism he expresses about religion and faith is right; but he falls into some of the same traps he warns us against when it comes to science. The human proclivity for superstition does not end with religion.

[24] As Rubinstein (2012, p. 129) observes, "My view of game theory is consistent with my approach to economic models in general .... Game theory does not try to describe reality or be normative. Game theory investigates the logic of strategic thinking".

References

Ashraf, N., Karlan, D. & Yin, W. (2006) Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines, *Quarterly Journal of Economics*, 121(2), pp. 635–672.

Banerjee, A. (2005) 'New development economics' and the challenge to theory, *Economic and Political Weekly*, 40, pp. 4340–4344.

Banerjee, A. & Duflo, E. (2009) The experimental approach to development economics, *Annual Review of Economics*, 1, pp. 157–178.

Banerjee, A. & Duflo, E. (2012) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (New York: Public Affairs).

Bardhan, P. (2013) Two ideas about fighting global poverty, *Boston Review*, available at: http://www.bostonreview.net/world-books-ideas/pranab-bardhan-little-big

Barrett, C. & Carter, M. (2010) The power and pitfalls of experiments in development economics: some non-random reflections, *Applied Economic Perspectives and Policy*, 32(4), pp. 515–548.

Basu, K. (2000) *Prelude to Political Economy: A Study of the Social and Political Foundations of Economics* (Oxford: Oxford University Press).

Basu, K. (2005) New empirical development economics: remarks on its philosophical foundations, *Economic and Political Weekly*, 40, pp. 4336–4339.

Basu, K. (2011) *Beyond the Invisible Hand: Groundwork for a New Economics* (Princeton, NJ: Princeton University Press and New Delhi: Penguin).

Basu, K. (2013) *Shared Prosperity and the Mitigation of Poverty: In Practice and in Precept.* World Bank Working Paper, available at: http://www.documents.worldbank.org/curated/en/2013/11/18506691/shared-prosperity-mitigation-poverty-practice-precept

Beamen, L., Chattopadhyay, R., Duflo, E., Pande, R. & Topalova, P. (2009) Powerful women: does exposure reduce bias? *Quarterly Journal of Economics*, 124(4), pp. 1497–1540.

Beasley, E. & Huillery, E. (2013) *School Resources, Behavioral Responses and School Quality: Short-Term Experimental Evidence from Niger* (Science Po, Paris: Mimeo).

Benhassine, N., Devoto, F., Duflo, E., Dupas, P. & Victor, P. (2013) *Turning a Shove into a Nudge? A "Labelled Cash Transfer" for Education*. NBER Working Paper, No. 19227.

Bernard, T., Dercon, S., Orkin, K. & Taffesse, A. (2013) *Learning with Others: A Field Experiment on the Formations of Aspirations in Ethiopia* (Mimeo, Wolfson College, Oxford University).

Bruhn, M. & McKenzie, D. (2009) In pursuit of balance: randomization in practice in development field experiments, *American Economic Journal: Applied Economics*, 1(4), pp. 200–232.

Bunge, M. (1959) *Causality and Modern Science*, 4th edn, 2009 (New Brunswick: Transaction Publishers).

Cartwright, N. (2007) *Hunting Causes and Using Them: Approaches to Philosophy and Economics* (Cambridge: Cambridge University Press).

Cartwright, N. (2010) What are randomized trials good for? *Philosophical Studies*, 147(1), pp. 59–70.

Cartwright, N. (2011) The art of medicine: a philosopher's view of the long road from RCTs to effectiveness, *The Lancet*, 377, pp. 1400–1401.

Cartwright, N. & Hardie, J. (2012) *Evidence Based Policy: A Practical Guide to Doing It Better* (Oxford: Oxford University Press).

Chattopadhyay, R. & Duflo, E. (2004) Women as policy makers: evidence from a randomized policy experiment in India, *Econometrica*, 72(5), pp. 1409–1443.

Das, J., Devarajan, S. & Hammer, J. (2011) *The Welfare Economics of Behavioral Economics and Randomized Trials*. World Bank Blog, available at: http://www.blogs.worldbank.org/files/africacan/DDH_The%20Welfare%20Economics%20of%20Behavioral%20Economics%20and%20Randomized%20Control%20Trials(1).pdf

Dawkins, R. (2006) *The God Delusion* (London: Bantam Books).

Deaton, A. (2009) Instruments of development: randomization in the tropics and the search for the elusive keys to development, Keynes Lecture, British Academy, in: *Proceedings of the British Academy 2008 Lectures*, Vol. 162 (Oxford: Oxford University Press), pp. 123–160.

Deaton, A. (2010) Instruments, randomization, and learning about development, *Journal of Economic Literature*, 48(2), pp. 424–455.

Dollar, D., Kleineberg, T. & Aart, K. (2013) *Growth Still is Good for the Poor*. World Bank Policy Research Working Paper No. 6568 (Washington, DC: World Bank).

Duflo, E., Glennerster, R. & Kremer, M. (2008) Using randomization in development economics research: a toolkit, in: T. Paul Schultz & J. Strauss (Eds) *Handbook of Development Economics*, Vol. 4. (Amsterdam: Elsevier), pp. 3895–3962.

Elbers, C. & Gunning, J. (2013) Evaluation of development programs: randomized controlled trials or regressions? *World Bank Economic Review*, 27, pp. 1–14.

Gates, B. (2013) Measuring progress, in: *Annual Letter* (Gates Foundation), available at: http://www.annualletter.gatesfoundation.org/#nav=section1&slide=0

Gigerenzer, G. & Gaissmaier, W. (2011) Heuristic decision making, *Annual Review of Psychology*, 62, pp. 451–482.

Gilovich, T., Griffin, D. & Kahneman, D. (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press).

Habyarimana, J. & Jack, W. (2012) *State Vs. Consumer Regulation: An Evaluation of Two Road Safety Interventions in Kenya*. NBER Working Paper No. 18378.

Hacking, I. (1975) *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference* (Cambridge: Cambridge University Press).

Hacking, I. (1988) Telepathy: origins of randomization in experimental design, *Isis*, 79(3), pp. 427–451.

Heckman, J. & Smith, J. (1995) Assessing the case for social experiments, *Journal of Economic Perspectives*, 9(2), pp. 85–110.

Heckman, J. & Urzua, S. (2010) Comparing IV with structural models: what simply IV can and cannot identify, *Journal of Econometrics*, 156, pp. 27–37.

Hicks, J. (1979) *Causality in Economics* (Oxford: Clarendon Press).

Hoff, K. & Pandey, P. (2006) Discrimination, social identity, and durable inequalities, *American Economic Review*, 96(2), pp. 206–211.

Hoover, K. (2001) *Causality in Macroeconomics* (Cambridge: Cambridge University Press).

Ifcher, J. & Zarghamee, H. (2011) Happiness and time preference: the effect of positive affect in a random assignment experiment, *American Economic Review*, 101(7), pp. 3109–3129.

Imbens, G. (2010) Better late than nothing: some comments on Deaton and Heckman and Urzua, *Journal of Economic Literature*, 47, pp. 399–423.

Iyer, L., Mani, A., Mishra, P. & Topalova, P. (2012) The power of political voice: political representation and crimes against women in India, *American Economic Journal: Applied Economics*, 4(4), pp. 165–193.

Jensen, R. (2012) Do labor market opportunities affect young women's work and family decisions? *Quarterly Journal of Economics*, 127(2), pp. 753–792.

Lorenz, K. (1977) *Behind the Mirror: A Search for the Natural History of Human Knowledge*, (translated from German by R. Taylor) (New York: Harcourt Brace and Jovanovich).

Manski, C. (2013) *Public Policy in an Uncertain World* (Cambridge, MA: Harvard University Press).

McKenzie, D. (2012) *Does Angus Deaton Worry Too Much About Wolves Eating His t-Values?* World Bank Blog, available at: http://www.blogs.worldbank.org/impactevaluations/does-angus-deaton-worry-too-much-about-wolves-eating-his-t-values

Miguel, E. & Kremer, M. (2004) Worms: identifying impacts on education and health in the presence of treatment externalities, *Econometrica*, 72(1), pp. 159–217.

Mill, J. S. (1843) *A System of Logic, Ratiocinative and Inductive*. Cambridge Books Online Edition 2011.

Mookherjee, D. (2005) Is there too little theory in development economics today? *Economic and Political Weekly*, 40, pp. 4328–4333.

Paul, M. & Dredze, M. (2011) You are what you tweet: analyzing twitter for public health, *Association for the Advancement of Artificial Intelligence*. http://www.cs.jhu.edu/~mdredze/publications/twitter_health_icwsm_11.pdf.

Ravallion, M. (2009) Should the randomistas rule? *Economists' Voice*, 6(2), ISSN (online) 1553-3832, available at: http://www.siteresources.worldbank.org/INTPOVRES/Resources/477227-1142020443961/2311843-1229023430572/Should_the_randomistas_rule.pdf

Richet, C. (1884) La suggestion mentale et le calcul des probabilities, *Revue philosophique de la France et de l'etranger*, 18, pp. 609–674.

Rodrik, D. (2008) The new development economics: we shall experiment but how shall we learn? in: J. Cohen & W. Easterly (Eds) *What Works in Development? Thinking Big and Thinking Small* (Washington, DC: Brookings Institution Press), pp. 24–47.

Rubinstein, A. (2012) *Economic Fables* (Cambridge: Open Book Publishers).

Russell, B. (1912) *The Problems of Philosophy*, Re-published by Project Gutenberg, Salt lake City, UT, 2009.

Savage, L. (1954) *The Foundations of Statistics* (Page references are to the revised Dover publication, 1972) (Hoboken, NJ: John Wiley & Sons).

Sen, A. (1959) Determinism and historical predictions, *Enquiry*, 2, pp. 99–115.

Sen, A. (1980) Description as choice, *Oxford Economic Papers*, 32(3), pp. 353–369.

Skyrms, B. (1980) *Causal Necessity* (New Haven, CT: Yale University Press).

Wantchekon, L., Klasnja, M. & Novta, N. (2013) *Education and Human Capital Externalities: Evidence from Colonial Benin*. BREAD Working Paper.

Worrall, J. (2007) Evidence in medicine and evidence-based medicine, *Philosophy Compass*, 2/6, pp. 981–1022.