

1 Theory + Experiment Do Not a Science Make

Preface

Science has undoubtedly produced remarkable achievements from deep theories to technological devices to new ways to measure things. These achievements, I claim, are secured by a dense interwoven net of scientific constructions that constrain and support each other – the concurrent, mutually feeding back-and-forth development of ideas, concepts, theories, experiments, measures, middle-level principles, models, methods of inference, research traditions, data and narratives that make up a scientific endeavour, with rich interconnections with other bodies of work on very different topics that also constrain and support it.

I think it's significant that the images of scientists that Tabi and I found on Google for our drawings are of Einstein. Of course, if you are drawing a scientist you need to make the person recognisable as a scientist, and that's hard when it's theory you want to point to. Einstein has become the canonical emblem for a theoretical scientist. But that's not all there is to it. Einstein represents a certain take on theory – created by a genius, original, explaining nature's deepest secrets. I Googled 'science theorists' and the very first entry was 'revolutionary theories' – grand sweeping theories with broad explanatory stretch, like relativity, quantum theory, evolution, plate tectonics and game theory lately

so popular in the economic and social sciences.¹ Not the tens of thousands of more local, domain-specific or lower-level theories across the sciences that make up the bulk of scientific theorising, like the theory of high-temperature superconductivity, the theory of democratic peace, the theory of thermal neutron scattering, ecological niche theory, theories about institutional racism, protein folding, the natural rate of unemployment, cognitive dissonance and so on and so on.

Nor is this focus on theory and experiment and concomitantly the explanations that theory provides – and on getting those right – peculiar to philosophy and popular imagination. The US National Academy of Sciences' 2008 report, *Science, Evolution, and Creationism*, builds this right into their 'Definition of Science': 'The use of evidence to construct testable explanations and predictions of natural phenomena, as well as the knowledge generated through this process.' They go on to describe 'how science works':

Scientific knowledge and understanding accumulate from the interplay of observation and explanation [which sounds to me very much like 'theory and experiment']. Scientists gather information by observing the natural world and conducting experiments. They then propose how the systems being studied behave in general, basing their explanations on the data provided through their experiments and other observations. They test their explanations by conducting additional observations and experiments under different conditions. Other scientists confirm the observations independently and carry out additional studies that may lead to more sophisticated explanations and predictions about future observations

and experiments. In these ways, scientists continually arrive at more accurate and more comprehensive explanations of particular aspects of nature.²

Similarly, the UK Science Council claims: ‘Science is the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence.’³ Again, this sure sounds like ‘theory and experiment’. So – it seems – science is all about theory, knowledge and explanation and the experiments and other observations that confirm these.

Sometimes it seems that even experiments do not get much of a look in. Consider this lament by a particle physicist turned historian and philosopher of science, Allan Franklin:

One of the great anticlimaxes in all of literature occurs at the end of Shakespeare’s *Hamlet*. On a stage strewn with noble and heroic corpses – Hamlet, Laertes, Claudius, and Gertrude – the ambassadors from England arrive and announce that ‘Rosencrantz and Guildenstern are dead’. No one cares. A similar reaction might be produced among a group of physicists, or even among historians and philosophers of science, were someone to announce that ‘Lummer and Pringsheim are dead’. And yet they performed some of the most important experiments in the history of modern physics. It was their work on the spectrum of black-body radiation, along with that of Rubens and Kurlbaum, that showed deviations from Wien’s Law and formed an important part of the background to Planck’s introduction of quantization.

This is symptomatic of the general neglect of experiment and the dominance of theory in the literature on the history and philosophy of science. In Thomas Kuhn's history of quantization, *Black-Body Theory and the Quantum Discontinuity, 1894–1912*, Lummer, Pringsheim, Rubens, and Kurlbaum are, at best, peripheral characters. The title indicates what Kuhn thinks is important. We never see what the experimental results were or find a discussion of how they were obtained.

But, it might be said, that it is only an isolated case. Surely everyone is aware of the famous experiments of Galileo and the Leaning Tower of Pisa, of Thomas Young's double-slit interference experiment, and of the Michelson-Morley experiment. What seems to be generally known, particularly by scientists, about these experiments shows the mythic treatment of experiment. Real experiments and their roles are not often dealt with.⁴

I too, like the Academy of Sciences and the Science Council, see much knowledge and explanation when I look at science, and I agree that much of what is mentioned by them can be conducive to achieving these. But I lament the narrowness of focus, and my lament extends far beyond Franklin's. When I look at science I see far more than theory and experiment, knowledge and explanation. As I've already indicated, I see a hotch-potch: models, concepts, validation procedures, measures, classification schemes, statistical techniques, study designs, data collection, curation and coding methods, mathematics high and low, methods of inference, narratives and more. These are resources ready to hand, for science to use to build successful technological devices, make precise

predictions and create compelling accounts of the world around us. Experiment and theory definitely must be there. But finding them – especially those ‘high’ revolutionary theories – among all the parts that play an essential role is a bit like playing Where’s Wally.

For science-related institutions to endorse the common image that science = theory + experiment reflects a lack of intellectual humility in those institutions about the importance of theory and experiment to the conduct of science and to what the sciences deliver. But I do not urge attention to these just because every labourer is worthy of their hire but also because these different products of science are mutually supporting. Each successful endeavour in science depends on these being up to the job they are needed for if the endeavour is to be successful. That’s why intellectual humility about theory and experiment is important in the institutions in and around science. The different kinds of enterprise in science are intricately interwoven and interdependent. We need due attention, policing and support for all of them if any are to perform as best they can. And failures to attend to them can have seriously harmful consequences.

I see science and how it operates as like a giant Meccano set, with scientists akin to a vast network of hard-working, practised Meccano builders labouring together in different teams on different creations. The more usual image pictures science as uncovering nature’s deepest secrets, highlighting big breakthroughs, grand theories and brilliant experiments, done by men of genius, insight and finesse. As historian Jaume Navarro notes, ‘[t]he folk history of science stresses crucial experiments, moments of revelation,

great achievements by geniuses'.⁵ I see science operating differently. I see a science that sails, with effort and dedication, between the Charybdis of a hubris that assumes our scientific successes are due to heroically wresting nature's secrets from her and the Scylla of diffidence that urges that we don't *really* know anything and should proceed only with extreme caution.

Science produces remarkable and reliable outputs. But not primarily by ingenious experiments and brilliant theory. Rather by learning, painstakingly, on each occasion, how to discover or create and then deploy together a panoply of different kinds of highly specific scientific products to get the job done. Every product of science – whether a piece of technology, a theory in physics, a model of the economy or a method for field research – depends on huge networks of other products to make sense of it and support it. Each takes imagination, finesse and attention to detail, and each must be done with care, to the very highest scientific standards, so that it can do the immediate job we expect of it and because so much else in science depends on it. There is no hierarchy of significance here. All of these matter; each labour is indeed equally worthy of its hire.

I begin by looking at theory and experiment and all they bring with them.

The Centrality of Theory and Experiment, Knowledge and Observation

I begin by considering a simple principle in physics that we probably all know: electrons are negatively charged. This is a

central part of physics theory. Later I will talk a bit about biology, chemistry, neuroscience and the social sciences, but I want to start with physics because it is the usual paradigm. I want to show how misleading it is to highlight theory and experiment even here in this most likely of places.

Here is the principle that tells us just how big the negative charge on electrons is:

EC principle: The charge of the electron is -1.602×10^{-19} Coulombs.

Already this one simple principle raises further questions. What's an electron? What's charge? What are Coulombs?

Let's suppose for the moment that we know something about what charge is and aren't too fussed about the units – Coulombs. Probably you know that an electron is the smallest unit of negative charge. More questions. This supposes that negative charge comes in units. It's not continuous, like the electromagnetic field or as we normally envisage water in a beaker. How do we know that it comes in discrete units?

All that, and much more, is answered for us in the theory of the electron. I do not propose to say anything much about what this theory says nor how it answers those questions; just a short paragraph to locate it, after which I make some general observations about it that tend to be true of theory generally, whether in physics, biology or economics.

The discovery of the electron is usually attributed to J. J. Thompson in 1897 in his cathode ray experiments. But theory about the electron both predates and postdates these

experiments. As to predating, from the introduction to an impressive collection on the history of the electron:

The electromagnetic effects produced by moving charges had been explored by Thomson and by Oliver Heaviside in the 1880s and had been developed into a fully fledged 'electron theory' of matter by George Fitzgerald and, especially, Joseph Larmor in the 1890s.⁶

Plus ever so much more you can read about in the collection.

As to postdating, the theory got a dramatic rewrite with the advent of quantum mechanics and the theory of relativity, at first by Louis de Broglie and Irwin Schrödinger on the quantum side, which left us the image of the electron as sometimes a small compact particle and sometimes a smeared-out wave, and by P. A. M. Dirac, who produced a relativistic wave equation for it. And even the latest developments in quantum electrodynamics and quantum field theory leave many questions open. So, it is a live theory that is still expanding and developing.

But it is not just developing in conversation with observation and experiment as suggested in the National Academy of Science's discussion of how science works. There's far more going on.

The Melange of Theory Ingredients

You might think of theory as just a set of claims – laws or principles or equations. But there is far more to it than that. In particular, in this section I outline the role of a number of other ingredients that constitute theory, making it

meaningful and useful. In particular, I look at concepts; models and narratives; and diagrams, illustrations and graphs. I am going to look at concepts in considerable detail because they are after all the meat and potatoes of theory, then turn more briefly to these other ingredients of theory.

Concepts

You can't have theory without concepts. It is also widely accepted that you shouldn't admit concepts into science that do not have exact and unambiguous meanings or that lack clear indications of ways of connecting them with the world. It is important for science to ensure that its concepts have genuine empirical content – that they have a grip in the empirical world – even if what they get a grip on is not much as we conceive it to be. There are many other bodies of thought, like various complex theologies, that resemble the sciences in being highly articulated and regulated, with a network of accepted claims and methods that tightly constrain what further moves can be made. What separates the concepts of science from these is the reassurance demanded in science that its concepts get a grip on the empirical world.

We typically do this through measurement and experiment. When we do so the measurements and experiments can play a dual role. We frequently use experiment to test whether a theory is true or not and simultaneously to help secure empirical content for concepts in it. In the final section, 'You Can't Build an Experiment without a Gigantic Meccano Set', I describe in some detail Robert Millikan's famous oil drop experiment, which served both to measure

the charge on the electron, thereby providing empirical content to the concept of ‘the electron’, and to test the theory that negative charge comes in discrete units. The reason I bring this up here is to point out that measurement and experiment are not just necessary to test theory but without them we do not have empirical concepts at all, and without empirical concepts we do not have empirical theory. So, just as we need concepts to have theory, so too we need experiments or other methods of measurement to have empirical theory – and *empirical* theory is what we want in science!

Defining Concepts

The concepts we find in our theoretical sciences do not of course spring into life fully grown. They have a prehistory: an often long period during which they are developed, contested and eventually, as the field of science studies puts it, ‘stabilised’. I’ll talk about this process of stabilising concepts later. For the moment let’s just think about mature concepts in what we take to be reasonably well-established theories. How are we to understand these theoretical concepts, how do we give meaning to them? I’m going to spend some time discussing this both because these are interesting questions in their own right and also because the long road we travel in thinking about these issues leads us to the conclusion I want to underline in this chapter – how important all the other pieces beyond theory and experiment are in the Meccano set of science, so much so that the very notion of theory does not make sense without them.

Some theoretical concepts have esoteric names and are unfamiliar to those who are not specialists in the field. In particle physics we learn about *leptons*, in low-temperature superconductivity about *Cooper pairs*. So, what is a lepton, what is a Cooper pair? Cell biologists study the *cellular cytoskeleton*, including *actin filaments*, *microtubules* and *intermediate filaments*. They examine the *endoplasmic reticulum* and the *Golgi apparatus*. Cancer biologists study *proto-oncogenes* and *tumour suppressor genes*. What are actin, microtubules and intermediate filaments? What are the endoplasmic reticulum and the Golgi apparatus? What are proto-oncogenes and tumour suppressor genes? Social psychologists study *positioning theory*. What is a position? Other theoretical concepts have familiar labels, especially in the social sciences, like *democracy*, *socio-economic status* and *aggressive behaviour*. But as these come – our everyday versions of them – these concepts are too vague and open-ended to play a role in proper science theory. The claims of science are meant to be clear and unambiguous. So, whether the concepts they employ have familiar or unfamiliar monikers, the concepts themselves need to be made clear and unambiguous.

The easy way to give meaning to a concept is simply to define it. For instance, ‘leptons are considered to be fundamental particles. They have a spin $1/2$ and do not partake in strong interactions [interacting only via electromagnetic and weak forces]. As fundamental particles, some leptons are negatively charged.’²⁷ That’s fine – sort of. The problem is that this definition is clear only if what it is to be a ‘fundamental particle’ is already clearly defined, along with ‘spin $1/2$ ’, ‘strong interaction’, ‘weak force’, ‘electromagnetic force’ and ‘negatively

charged'. Or consider 'proto-oncogenes'. These are defined by the US National Cancer Institute as: 'Gene[s] involved in normal cell growth. Mutations (changes) in a proto-oncogene may cause it to become an oncogene, which can cause the growth of cancer cells.'⁸ The clarity of this definition also depends on the clarity of other terms such as 'normal cell growth', 'mutations', 'oncogenes' and 'cancer cells'.

What we see here is typical: one theoretical concept defined in terms of other theoretical concepts that are themselves in need of clear and precise definitions. When it comes to social science, this problem is often compounded by the fact that there may be multiple different definitions on offer, definitions that classify items in the world differently. Consider 'democracy'. Definitions range from:

[G]overnment with the consent of the governed. This formula is indeterminate with respect to institutional forms, or the procedures by which consent is to be expressed – questions on which consent theorists have historically differed⁹

to

[A] competitive political system in which competing leaders and organizations define the alternatives of public policy in such a way that the public can participate in the decision-making process¹⁰

to definitions that provide more formal criteria. For instance, for a state to count as a democracy, its citizens must have:

1. Effective participation
2. Voting equality at the decisive stage

3. Enlightened understanding
4. Control of the agenda
5. Inclusiveness.¹¹

Note that, just like the examples from the natural sciences, these too end up defining one theoretical concept in terms of other theoretical concepts. What, after all, is a ‘competitive political system’, or ‘control of the agenda’, or ‘public participation in political decisions’ – what even is a ‘political decision’?

Can we ever break out of this theory circle, to define theoretical concepts in less esoteric terms? That was the hope in the heyday of operationalism, which was championed by the American physicist P. W. Bridgman. Bridgman’s method for breaking out of the theoretical circle was to define each theoretical concept by an operation by which it is measured: ‘We mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations.’¹² Readers may be familiar with this kind of doctrine from behaviourism – sometimes called ‘rat psychology’ – which was dominant in psychology from the 1920s through the 1950s. B. F. Skinner is its most well-known advocate. Skinner maintained that psychology is the study of behaviour, not of some invisible ‘inner mind’, and he urged that all psychological concepts, like anger, revulsion or reflection, must be defined in behavioural terms. Note *defined*. Behaviour is not a *reflection* of psychological states or a *clue* to what someone else is experiencing. Behaving in particular ways is what it is to be angry, revulsed or reflective.

Behaviourism was widely abandoned in psychology because it just didn't seem to work. We couldn't articulate distinct kinds of behaviour that fitted with each different nuanced psychological concept that we regularly employ and would surely not want to abandon on the grounds that it must be a chimera since we can't find a behavioural definition of it. Behaviourism also completely misses out on the 'what-it-feels-like' to be in these different states. So, gradually the claim that all psychological phenomena comprise behaviour lost dominance to the more everyday view that they are causally related – though, as I discuss in Chapter 2, in the section 'Physicalism and Materialism', objections to the genuine reality of the inner mind and its states remain alive in the programmes to reduce mental states to states of the brain.

Operationalism more widely suffers from two general problems. First, every different measurement procedure introduces a new concept even if, on the more usual way of thinking about it, these procedures measure the same concept but in different ways. Then it seems nature must be littered with huge numbers of principles to bind together the different concepts, all of which take the same value in the same circumstances and that we normally think of as all the same concept, just measured differently; principles like this: 'Concept 1 and Concept 2 and Concept 3 and . . . Concept N always have the same value.' Second, we generally make big efforts to *defend* the idea that our procedures are good for measuring what they are supposed to. As I noted, when I turn to discussing experiments I use the famous early twentieth-century oil drop experiments for which Robert

Andrews Millikan won the Nobel Prize for measuring the charge of the electron. As you will see in the final section in this chapter, ‘You Can’t Build an Experiment without a Gigantic Meccano Set’, Millikan did not just assert that what he calculated from the reading on the voltmeter in his experiment was the charge of the electrons in his experiment. He argued for this claim, both with a theoretical model – you’ll see the core of this in Figure 1.8 – and also in a detailed description of the actual materials which his experiment employed to show that they could play the part required of them in securing an accurate and precise measure of the charge.

Consider another example, this time from genetics. In his 1929 article on Heredity in the *Encyclopaedia Britannica*, British biologist J. B. S. Haldane provided what can be viewed as an operational definition of *phenotype* and *genotype*. He wrote that ‘[a] class of organisms whose members cannot be distinguished from one another by observation is called a phenotype; a class which can be distinguished from another by breeding tests is called a genotype’.¹³ Such definitions were given because genes or the genetic material had not yet been discovered. At the time, genes were accepted as some unobservable ‘units of heredity’ without any direct experimental handle on them. Hence, a quasi-operational definition was beneficial. Yet, today, with the advent of molecular genetics, which characterises genotype as the ‘variant forms of a gene that are carried by an organism’ and phenotype as the ‘observable physical properties of an organism’ – both of which accept a molecular characterisation of the genetic material in the

form of deoxyribonucleic acid (DNA) and the theoretical models used to explain gene expression – we would consider such operational ‘definitions’ superficial, serving at best as indicators of genotype and phenotype but falling short of proper definition. This shows that establishing the proper definition of genotype and phenotype required a lot of material experimental work within some broad theoretical framework that operational definitions fail to appreciate.

Operationalism makes nonsense of these elaborate efforts to defend that our measurement procedures are up to the job of measuring the concept they are supposed to, since after all operationalism holds that the concept just *is* what is measured by those procedures.

Just looking to the procedures by which a concept is measured may, though, be too narrow a focus in attempting to break out of the spiral of definition of one theoretical concept by another by another by another. In the 1930s, 1940s and 1950s, philosophers – especially logical empiricists – tried valiantly to define theoretical concepts using purely *observational* terms. They were called *logical empiricists* because they argued that scientific claims should be made entirely explicit, for instance theories should be formulated formally as systems of axioms from which further claims could be deduced as theorems. The *empiricist* label was because they wanted science to be able to confirm each claim by empirical observation. What exactly is meant by observation was up for grabs – did it for instance mean observable with the naked eye or was it to allow the use of sophisticated instruments? Whichever way that is decided,

what matters for breaking out of the circle of theoretical definition is that for each claim to be confirmable entirely by observation, each theoretical concept in that claim needs to have an observational correlate – some observational states that obtain if and only if the theoretical concept obtains. So, hurray, these observational correlates can serve as definitions for these concepts and the circle is broken for each concept.

The difficulty is that this programme failed miserably. There just seems to be no way to carry it off. As Carl Hempel, who is widely acknowledged as the leading philosopher of science of the time, concluded in a famous paper written originally in 1958: '[I]t is clear that theoretical formulations cannot be replaced by expressions in terms of observables only.'¹⁴ This continues to be true even if we become for more liberal, not looking just for correlates among features that can be observed but allowing that theoretical concepts be defined by any terms that we already grasp the meaning of, as Hempel suggested: '[W]e might qualify a theoretical expression as intelligible or significant if it has been adequately explained in terms which we consider as antecedently understood.'¹⁵ The result of these repeated failures is that philosophers settled for *implicit* definitions of theoretical concepts rather than explicit ones that defined them using concepts outside the theoretical circle. All that means is that we give in and take the meaning of a theoretical concept to be given by the axioms of the theory. As the *Stanford Encyclopedia of Philosophy* notes: "This idea has become almost constitutive of the very notion of a theoretical term in the philosophy of science."¹⁶

The problem with this is that implicit definition leaves it open what our theories are talking about. That's because no matter how detailed a theory is – how many axioms we add to its formalisation – there can never be enough detail to pick out uniquely what the theory is about. There will always be unintended or 'non-standard' interpretations for it. This follows from theorems in model theory in logic, but it is easy to get the gist of why by looking at a caricature example.

Consider the simple theory that has one axiom, an axiom familiar from school physics: $F = ma$. We think of this as telling us that the force on a material object is equal to its mass times its acceleration. But all we really know from the formula is that there are three quantities and the first is equal to the product of the other two. That is just as true of the area of a rectangle with respect to the length of its two sides. So F could mean area of a rectangle, m the length of the rectangle and a its width. Now let us add more detail. In a world where only gravitation acts we can add the law of gravity to our axiom set: $F_G = GMm/r^2$. This tells us what force a system with mass m will experience in the presence of another of mass M a distance r away, where G is the constant of gravity. From this we infer that in this gravity-only world, $ma = F = GMm/r^2$. But this works just as well for rectangles if we suppose that the rectangles come overlaid with kites, as in Figure 1.1.

Just suppose $G = 2$, $M =$ the area of the kite and $r = \sqrt{d}$.

We can go on thickening the theory of course. As we do so, more and more alternative interpretations will get

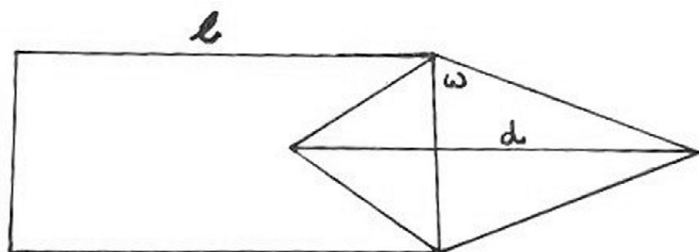


Figure 1.1 Kites and rectangles instead of forces, masses and accelerations

Drawn by Lucy Charlton especially for this book. Thanks Lucy!

ruled out. Maybe it will no longer be possible to see the equations of classical mechanics as all about rectangles. It naturally helps in narrowing down the interpretation if some of the concepts that figure in the theory are among that nice collection of concepts that are antecedently understood so that what they refer to is nailed down. For instance, we may suppose that a in the formula $F = ma$ refers to *acceleration*, which is a concept in our ordinary language vocabulary and we are clear just what it means. (But beware of even that. We take acceleration in the physics formula to be the rate of change of velocity with time (dv/dt) but in some medieval physics the acceleration of a falling body was understood as the rate of change of velocity with distance traversed (dv/dx .) Given this assumption, it gets harder – though not impossible – to interpret F as the area of a rectangle. Until each theoretical term is identified with something we antecedently understand, we cannot be sure of a unique interpretation. Without this, no set of axioms no matter how long can guarantee a unique interpretation. As the

philosopher Hilary Putnam argued, you can't fix what you are talking about just by talking more and more.

So, implicit definition of theoretical concepts by laying out the theoretical principles that are supposed to be true of the concepts doesn't work to fix what those concepts refer to. Nor can we define them explicitly in terms of other concepts we already understand outside the confines of the theory.

Philosophers worried about this problem a lot. But I don't know of any scientists who do. I think there's a good reason for this. Theory isn't just a set of claims that stands there to inform us of what the world is like – it's not there just to describe the world. Theories are tools that we use to do things in the world. We use theories to build models and we use the models to make predictions about what will happen in the world and to design experiments and technologies and policies and measurement procedures that we then implement, jostling the world, picking up bits and changing them around, to learn more about the world and to try to make it more to our liking. These are what break us out of the spiral of theory defining theory defining theory. When we position our specially designed radar sensors at the ends of the court at Wimbledon, train these on the ball that Serena Williams serves and read out 122 miles per hour, we know we are dealing with *velocity* and if we calculate its time rate of change (so $a = dv/dt$) and associate *that* with the formula $F = ma$, then we can be sure that $F = ma$ is not about the areas, lengths and widths of rectangles.

We must be careful, though, about how much all these successful interjections of theory into the world can

buy for us. They serve to rule out unintended interpretations of our theoretical concepts, but they do not ensure that these theoretical concepts refer to things that are really there in the world as we conceive them in our theories. Just think about all those theories that we used in the past to make successful predictions about what will happen in the world and to change things, like phlogiston theory that was used successfully to produce breathable air, inflammable air, shiny metal and ‘calx’ (metallic ash).¹⁷

Stabilising Concepts

Scientific concepts do not, of course, appear full blown and fully formed out of nowhere. Rather, they are constructed within and by science and its surrounds, contested and over time stabilised. There are various ideas of how this happens. To the early twentieth-century bacteriologist and philosopher Ludwick Fleck, concepts were developed within a community of experts and spread outward to the general public. At the same time, the public reinforces the thinking of experts. As Fleck urged, scientists are ‘more or less dependent, whether consciously or subconsciously, upon “public opinion”’.¹⁸

By contrast, in the late 1970s the philosopher Bruno Latour and the sociologist Steve Woolgar argued that a fact is produced within the scientific community through persuasion – scientists must constantly convince one another that certain statements using those concepts should be treated as facts. The point at which a concept becomes stable, Latour and Woolgar write, is when a statement using the concept ‘rids itself of all determinants of place and time

and of all reference to its producers and the production process'.¹⁹ The meaning of a concept becomes fixed when it does not seem to have been constructed in the first place. And of course, as time goes on and new things are learned and new ideas and influences arise, concepts can also destabilise and even eventually disappear.

Aside from the obvious social processes it takes to get a concept entrenched, the stabilisation of any one concept in science depends on many products of science, including not just other concepts. Concepts become stabilised in part because of how they are gradually interrelated with other pieces of knowledge and practice, especially those from fields of research other than the one that first developed the concept.

Take, for example, the concept of a 'neuron type'. According to neuroscientists, there are not just different types of cells in the nervous system but different types of *nerve cells* in terms of their shape, electrical properties, gene expression and, perhaps most importantly, function. Although the concept of a neuron type is not new, since the early 2000s neuroscientists have developed materials and methods that support and constrain what the concept of a neuron type can amount to. They have used genetically modified animals and/or genetically modified viruses to express light-sensitive proteins in specific neuron types within the brains of mice. Upon delivering flashes of light, they could manipulate the activity of these neuron types, such as those that produce a particular neurotransmitter. Consequently, they could also explore which behaviours these types of neurons control.²⁰ In this case, products that

were not concepts or theories, and specifically coming from genetics and virology, helped to stabilise the current concept of a neuron type. It is now well accepted that the genetic variability of neurons is important in understanding the structure and function of the brain.

Misusing Concepts

Since at least the Scientific Revolution on, it has been commonly – though definitely not universally – supposed that proper scientific theory is required to lay out exact relationships between cleanly delineated features of the world, as in physics and economics. When I come to talking about powers, tools and laws in Chapter 3, in the section ‘Where Is Physics Successful?’, you will see that this is not something I entirely subscribe to. But where it is required I have wanted to underline how important it is to characterise the concepts that theory employs to refer to those features carefully and unambiguously and to provide ways to measure them precisely and accurately.

These demands that concepts be well specified and cleanly measurable creates problems however. As one of the founding members of the Vienna Circle, Otto Neurath, noted, many concepts that we find useful and that science helps us learn more about are loose congestions of different ideas and criteria with vague boundaries, like socio-economic status, learning readiness or implicit bias. Much of the world that science studies does not lend itself to description by precisely, rigidly delimited theoretical concepts.

That means that theoretical concepts that are tightly constrained by a web of mathematical laws and by highly

precise criteria for application may not be universally applicable but at best constrained to pockets of reality – and that attempting to apply them more widely can often lead us into trouble. To underline that this is no idle philosophical worry, I will illustrate it in detail with a well-known concept at the heart of social science – probability. The lesson to be drawn from the case of probability about concepts and what I shall term ‘small worlds’ hold widely across both the social and the natural sciences.

You will already be familiar with the fact that broad swathes of social science research are given over to establishing, analysing, generalising, theorising about and using statistical associations that are manipulated with the assumptions of probability theory.

This makes sense if probabilities can be attached to broad swathes of the phenomena that social science is meant to deal with. But can they? Here we face the same issue that you will meet when I discuss the assumption of universal determinism: is the social world really that orderly? Perhaps it is my failure to see the forest for the trees, but when I look at various studies across the social sciences, from psychology, sociology and political science to economics and public health, I often cannot see grounds for this assumption, I sometimes see good evidence against it and I also see places where it seems to be leading us astray, with respect both to the accumulation and the use of knowledge.

The popular view that the reliability of real science requires precision is reflected in the account I have been giving of the importance of clear concepts and measurement in producing scientific theory. If the social sciences are to aspire

to the same reliability, surely they too must aspire to precision! For example, cost–benefit analysis purports to combine inputs and outputs by quantification which allows the computation of an unambiguously best outcome: the optimum. This appears to require very precise knowledge of causes and effects, for example that if you increase demand by this (precise) amount prices will go up by this (precise) amount. But everybody knows that you don't know *exactly* what will happen. You are uncertain. Maybe 2 per cent, maybe 3.5 per cent. Not precise. Precision is to be re-established by the notion of the probability distribution. If you assign probabilities, as you do with horse-racing odds, to various outcomes, then you can maximise by calculating the expected outcome of this or that input, and you are back with what is called mathematical precision. This is more than assessing carefully in qualitative terms which outcome is more likely than that. It requires quantification to enable commensurability and allow calculation of a precise optimum.

This account says that you assign probabilities. For that to be legitimate, there have to be probabilities to assign. And the paradigms of tossing dice and drawing cards do a lot of work to reassure us that this is a fair assumption. Of course, in real life, where the social sciences operate, it may be hard to find probabilities. And you may not be very sure that you have got them right. But surely it is reasonable to ask of any input, 'what are the odds that it will produce this result?' and thereby produce usable probability distributions. And you should use them, it would be irrational not to use them, not just because they have the properties you need to carry out the manipulations you want but because, even if

shaky, they are better than nothing, they are better than the alternatives. So this familiar trick is used in the social sciences to tidy up the messiness of uncertainty and turn it into a well-behaved set of numbers.

But what if there are some problems which do not allow this? What if there are some uncertainties which are not like 'I am uncertain whether I will throw a 6', uncertainties which are fundamental? This is a source of troubles for the social sciences that I have long worried about. But if this is an obstacle to successful social science and successful social decision-making, as I think it is, it is an obstacle we bring on ourselves. For it involves the misuse of the otherwise valuable tool, reasoning with probabilities. Probabilities are a genuine aspect of the social world – under some circumstances. But problems arise when we assume there are probabilities where there are none and then base our theories and our expectations on them.

Where there are none? Really? Isn't there always some probability or other for an event? I think not. And that's part of what makes trouble for social science. Probability is a superb tool, like a thirty-piece set of first-class chromium screwdrivers. But it may be that not all of our problems are loose screws.

What do social scientists say about these worries about the assumption that there are probabilities here, there and everywhere? One person who has defended them to me is the Oxford econometrician David Hendry. Hendry assumes that econometrics would ideally like to discover 'the complete structure of the process that generates economic data and the values of all its parameters'.²¹ He seems

to presuppose that this process is always appropriately modelled with a probability measure. In conversation he has supported this assumption with the remark that the sampling procedure can ensure that a probability measure is appropriate. I can make sense of this from the point of view of the statistician looking at samples from a particular population in order to infer characteristics of the population as a whole. But even in this case it makes a lot of difference how one thinks the sample has been drawn.

Consider a simple example. We have a pack of ten playing cards: four hearts, two diamonds, four spades. You draw one card at random. What is the conditional probability that the card drawn is a diamond, given it is red? Here there is an answer: $1/3$. Next consider a more complicated procedure. First, we do two totally separate and independent flips of a fair coin. If the results are HH, you draw one card at random from the 'pointed' cards in my population (i.e. diamonds and spades), otherwise from the non-pointed cards. Again, it is entirely correct to assign a probability measure to the outcomes, and under that measure the conditional probability that a drawn card is a diamond given it is red is $1/4$. But this depends on more than the sampling procedure – drawing one card at random. It depends on what is 'sampled'.

To see how it does so, imagine I give the cards to the person sitting closest to me to order on their aesthetic preferences – put the card they find most attractive first, then the second most attractive next and so on; then I pick the top card. Now we have to argue the case, and I do not know any good reasons to suggest that this situation should

be represented by a probability measure. What argument is there to show that there'll always be some probability or other for the outcomes from a data-generating method like this, characterised by features that we have no idea how to assign a probability to – like the various possible preferences of the person who happens to be sitting next to me? To insist that every data-generating process, regardless of its features, is correctly described by some probability measure or other is just to assert the point at issue, not to defend it.

Not all economists take the view that Hendry seems to have been defending. The economists John Kay and Mervyn King express the same worries I have in their recent book, *Radical Uncertainty: Decision-Making for an Unknowable Future*, which takes issue with the attempts to use what they call 'small-world' methods for making decisions in real-life large-world settings. Kay and King never explain what they mean by this but they make clear that a small world contrasts with much of 'the world as it really is',²² full of uncertainty, ambiguity and vagueness. I have more to say about small worlds in Chapter 3, in the section 'Where Is Physics Successful?', but I think that you'll be able to see the point here by thinking about their illustration, which involves a version of a television gameshow:

You are offered the choice of two envelopes and are told that one contains twice as much money as the other. You make your choice, open envelope one, and find that it contains \$100. The referee asks if you would prefer envelope two. Since one envelope contains twice as much money as the other, but you do not know whether you have chosen the larger or the smaller, you know that the

second envelope contains either \$200 or \$50, so you stand to gain \$100 or lose \$50 by switching from envelope one to envelope two. If you apply the Indifference Principle, and judge each of these outcomes equally likely, this sounds a good deal – expected value of \$25 – so you switch.

But suppose you had initially chosen envelope two, which contains either \$50 or \$200. If it had contained \$50, you would have stood to gain \$50 or lose \$25 by switching. If \$200, you would either gain \$200 or lose \$100. In both cases, the possible gain is twice the possible loss. So, if you had chosen envelope two, you would now want to switch to envelope one. Yet this conclusion cannot be right. Your initial choice is random, and it cannot be the case that if you selected envelope one you will always want to switch to envelope two – while if you chose envelope two you will always do better to switch to envelope one. But no one has ever come up with a clear and simple explanation of why the recommendation to switch is wrong. The hidden assumption is that there is a 50–50 chance of gaining or losing by switching irrespective of the amount in the envelope. But is that right? Who is putting the money in the envelope, and what are their financial resources? There appears no coherent way to identify the possible states of the world that characterise the problem, and hence no sensible basis for assigning probabilities. And this is true even though the rules of the puzzle appear to have been fully described.²³

Philosophers have been worrying about this for a good while. The philosophical basis for it was nicely laid out by the philosopher Ian Hacking in his 1965 book, *The Logic of*

Statistical Inference. Hacking taught that probabilities are characterised relative to what he called ‘chance set-ups’ – like drawing balls randomly from an urn with six white balls and four black balls or picking cards randomly from a fair deck – and they do not make sense otherwise.

Not every situation in the world can be characterised as a chance set-up. In fact, I don’t suppose that many situations that come naturally can be. Looking at my account and that of Kay and King, we can extract two key necessary features a situation should have in order to be classed as a chance set-up. One is that the outcomes and the process are fully specified. The other is that there are enough probabilities built in at the start to ensure that the probabilities you aim to calculate fall out logically. The slogan for this last is: no probabilities in, no probabilities out.

We can easily see how the two requirements are satisfied in the two cases I noted as exemplary chance set-ups. Consider the first case:

1. The outcomes are: {the one card drawn is a heart, the one card drawn is a diamond, the one card drawn card is a spade}, where it is clearly meant to be supposed that hearts and diamonds are red and spades are black.
2. The probability ‘in’ comes from the description that one card exactly is drawn ‘at random’ from a pack of ten cards with the frequencies of each suit as specified. ‘Random’ is a technical term. It does not mean just ‘drawn without thought’ or ‘arbitrarily’, It means: each card has an equal probability of being drawn. The fact that all hearts and diamonds are red and all spades are black guarantees

certain conditional probabilities about the colour of a card given its suit: Prob (a card is red/the card is a diamond) = Prob (a card is red/the card is a heart) = Prob (a card is black/the card is a spade) = 1. These together are enough to fix the conditional probability we want: Prob (the card is a diamond/the card is red).

We can do the same with my second example.

Why is the game described by Kay and King not a chance set-up? Well, start with the very first step. You are invited to choose envelope one or choose envelope two. If you do either you are guaranteed some money – or you are so long as it is part of the defining characteristics of the game that what you are told is true. Let us suppose that is certain. Then it is a no-brainer to join the game. Well, yes – unless it is run by some dreadful organisation you don't want to be associated with, you haven't time, there'll be publicity you don't want, etc. The possibilities are endless and they fall in no orderly scheme. Perhaps we can use the ordinary English idiom of likely/unlikely for many, but there is no way to put a proper probability measure on them.

Suppose you join the game. Now you have to choose envelopes. How do you decide? You might decide to choose the envelope with the number of the age of your youngest child or the odd number because you associate 'odd' with 'eccentric' and you like eccentrics, etc. Still no genuine probabilities yet.

Now suppose you have indeed chosen an envelope – envelope one – and it contains \$100. How do you decide to keep it versus switching? Knowing a little about rational choice theory, you might think in terms of expected gain and

loss and try to maximise the ‘expected value’. And you might be happy to take the face values of the dollar amounts as appropriate measures of how much value those amounts provide for you, which you might not always do. (Suppose for instance you drew envelope one and the \$100 seems a godsend – it will just allow you to pay your mortgage but half that definitely will not.) Let’s suppose that the game show claim that the contents of the second envelope are either half or twice that of the first can be trusted. In that case the possible outcomes of switching and of staying put are fixed too. But to calculate expected gains and losses you need probabilities.

So where do the probabilities come in? Kay and King say, ‘[t]he hidden assumption is that there is a 50–50 chance of gaining or losing irrespective of the amount in the envelope’. That’s not given in the information defining the situation. That comes about from the assumption that ‘you apply the Indifference Principle, and judge each of these outcomes *equally likely*’.²⁴ The Cambridge Press collection, *The Theory of Decision under Uncertainty*, tells us:

The principle of indifference, also known as the principle of insufficient reason, is attributed to Jacob Bernoulli, and sometimes to Laplace. Simply stated, it suggests that if there are n possible outcomes and there is no reason to view one as more likely than another, then each should be assigned a probability of $1/n$. Quite appropriate for games of chance, in which dice are rolled or cards shuffled, the principle has also been referred to as the ‘classical’ approach to probability assignments.

However, this principle has to be used with great care.²⁵

One conventional reason for this is that you are told, ‘when you don’t know anything that favours one outcome over another, it is rational to treat each possible outcome as equally likely’. But you are not told how to divide up the outcomes. Real-life problems seldom come with the outcomes pre-set. *The Theory of Decision under Uncertainty* gives this example:

[A]ssume that I ask you what is the probability that it will rain tomorrow. If you think of the two outcomes, ‘rain’ and ‘no rain,’ you come up with the probability of $1/2$. But if the possibilities are ‘rain,’ ‘snow,’ and ‘no precipitation,’ the probability drops to $1/3$. Typically, one can partition the state space in a multitude of ways, resulting in the principle of indifference assigning practically any desired probability to the event in question.²⁶

This is not the problem in the envelope-switching game though. There the outcomes are clear. Keep the envelope you drew and take home the amount of money found in it. That we have supposed to be certain. If you switch envelopes, we may suppose you have not been lied to and you take home either twice or half that amount. The problem is with the principle of indifference. This instructs you to treat outcomes as equally probable when ‘there is no reason to view one as more likely than another’.

What King and Kay point out is that given the information supplied there is plenty of reason to think they are not equally probable. For instance, as they suggest, you may well have views about ‘[w]ho is putting the money in

the envelope' and 'what . . . their financial resources [are]'. You may also have views about whether the game show likes to produce big winners or not, or when across their series of shows they like to do this and how far along in that process you are, or whether their audience share has dropped recently and they want a big win to entice back viewers or much else relevant.

Of course, you may not know which of these pieces of information you have are really reliable and you may not be sure which option they favour, if either, or what it amounts to all put together. I suppose in this case you might throw up your hands and say, 'oh well, might as well flip a coin'. But we should hardly count this as a way to proceed that rationality dictates. We would certainly be shocked to learn that a jury did this in deciding on a verdict of guilty or not.

This, though, is all about invoking the principle of indifference and assigning a 50–50 probability to the two outcomes that you'd get from switching. Behind Kay and King's worries, and mine, is a more serious problem: why think there are any probabilities applicable in this case? If there aren't any, then surely it cannot be rational just to act as if there are, then assign them in some way or another and act accordingly.

Consider where these probabilities could come from. How for both envelopes would we get 50–50 chance of there being half or double that amount in the other? The most obvious starting idea I suppose is that the game show organiser chooses some amount of money they are prepared to start with – say \$100 – and flips a coin to settle which

envelope to put it in. Imagine that's envelope one. Then they again flip a fair coin to decide whether to put \$200 or \$50 in envelope two. In this chance set-up, $\text{Prob}(\$50 \text{ in envelope } 2) = \text{Prob}(\$200 \text{ in envelope } 2)$. But the reverse is not true. Whether there is \$50 or \$200 in envelope two, $\text{Prob}(\$100 \text{ in envelope } 1) = 1$. So, the odds of doubling versus halving are not equal for both envelopes, as was proposed for calculating expected values.

'Ah', one might object, 'still we don't know that's the chance set-up we are in'. That's part of the point of calling this a 'decision under uncertainty'. Here we are not just uncertain what is in the other envelope, but we haven't a clue how that was decided, nor how it was decided to put \$100 in envelope one to begin with. Maybe they decided to put \$50 in envelope two and then flipped a coin for envelope one's contents, or maybe they were short of money so only had \$25 to put there. In these circumstances, whether we act as if we don't know anything else and apply the principle of indifference or whether we look at all our information and come up with some different numbers, these just represent our *subjective* probability. And here I mean purely subjective, not anything based in facts we know about the world. It is just an arbitrary number we decide to use so we can bring the apparatus of expected loss and gain to bear to guide our choice.

Now, there's a very great deal written back and forth about this: what is the rational way to make decisions when we know very little? There's a whole field devoted to it, echoed in the title of Kay and King's book: decision-making under uncertainty. But that's not the issue I am discussing. Rather, here I am exploring reasons why the social sciences

may not be so good at prediction as we might hope, nor at establishing detailed theories on which there is widespread agreement, nor at devising stable explanations of social phenomena – ones that stick around. So here the issue is whether part of the reason for this could be that social science predictions, explanations and theories are often based on models that presuppose that there are probabilities where in fact there are none.

I'll illustrate this with a case I've worked on recently – the idea of an 'effect size' of a treatment or a social intervention. Effect size is big business in social science right now, as witnessed in these remarks:

- Nearly every discussion about educational improvement today refers to 'effect sizes'. Education organizations compare effect sizes in planning professional learning programs. District and school leaders consider effect sizes when selecting the strategies to include in school improvement initiatives. Even classroom teachers evaluate effect sizes in deciding what practices will be most effective in helping their students learn.²⁷
- Effect sizes are the currency of psychological research.²⁸
- Increasing emphasis has been placed on the use of effect size reporting in the analysis of social science data.²⁹

The effect size of a cause on a given outcome is an average in the numerator plus a denominator that is supposed to make these averages comparable across different interventions, like smaller class sizes versus more teaching assistants. The denominator introduces very serious problems of its own, but here I want to focus on the fact that the effect size

numerator involves an average. Thus enters probability. It is an average across the individual effects of the cause across a given population, so each of these individual effect sizes is supposed to occur with some probability in that population in order that the average makes sense.

Here I should pause for a moment's clarification. Take any given population, like the students in my UCSD 2018 Philosophy of Social Science class, and any given quantity that we can assign a number for to each individual, like their numerical marks in the first exam. I can always calculate an average of those in the usual way – multiply each grade outcome by the number of students who got that grade, add those numbers together and divide by the total number of students who received grades. Consider the grade 66. We can call the relative frequency of students receiving a 66 the 'probability' of getting a 66 in the population of students who received a mark. And it is true that relative frequencies are probabilities in the sense that they satisfy the fundamental axioms of probability. But to then use that probability to calculate an effect size for that kind of exam or my kind of teaching that educators should sit up and take notice of requires much more than the fact that it is a true relative frequency in my class. It must have some broader, more stable significance. There are many ways of thinking of probability that do not involve frequencies at all – probabilities are disposition-like objective chances or, as 'Bayesians' do, degrees of belief. But those who do think in terms of frequencies are not content with just finite frequencies in a single case, like the twenty-two students in my class. Rather they define

probabilities as *limiting relative frequencies in an infinite run of repeat trials* on similar populations. The infinite repetition is supposed to supply the generality or stability required for attaching the label ‘probability’.

It is just this kind of stability that I worry about with respect to these many effect sizes calculated and promulgated in social science nowadays. To explain why, I have to say a few things about the kinds of causes or interventions whose effect sizes are under consideration. The cause or intervention under discussion is very seldom if ever enough on its own to produce a contribution to the named effect. It needs supporting or helping factors to be enabled to do so. The same is generally true of causal claims we make in both science and daily life. The stock philosophers’ example is that striking a match causes it to light. Well yes, supposing there is oxygen in the room, there are no heat-sensitive sprinklers that will douse the match as soon as the match head begins to heat up, there’s no eager firefighter ready to blow on it at the merest sign of a flame, etc. etc. etc.

In statistics these supporting factors are called ‘interactive’ or ‘moderator’ variables because they interact with the cause to moderate how much, if anything, it will contribute to the effect. They are often graphed in causal pies or cakes, like those pictured in Figures 1.2, 1.3 and 1.4. In formal treatments, as in econometric models or proofs about what can be accomplished with randomised controlled trials, the relation between causes and effects are represented in equations like this

$$Y_i = b_i X_i + W_i$$

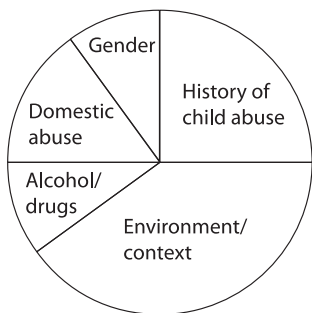


Figure 1.2 A cake diagram for Person A (abused as a child)
From Munro et al. 2016.

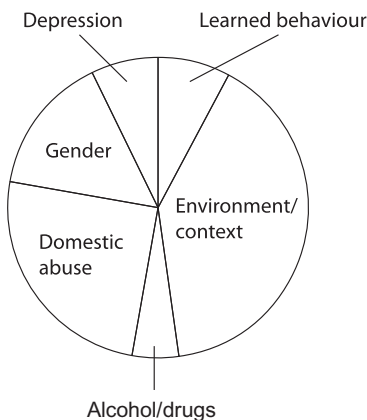


Figure 1.3 A cake diagram for Person B (not abused as a child)
From Munro et al. 2016.

where X_i is the cause in question of the outcome Y_i for individual i , b_i is the interactive variable representing the net effect of all X 's supporting factors that obtain for i and W_i is the net contribution of all the factors that contribute to Y for i independently of X . Note that the value of the

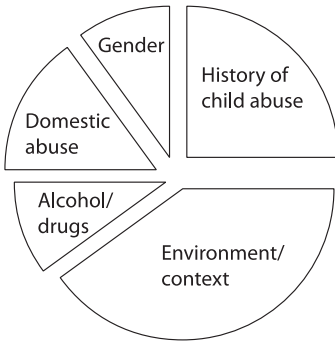


Figure 1.4 A cake diagram for Person C (abused as a child: conditions not combined)

From Munro et al. 2016.

supporting factors b moderates how much a given value of X contributes to Y as it should.

To see my worries, consider a simple two-valued case: either the cause is present for i (e.g. student i is in the small class) – $X_i = 1$ or it is not (that student is in the usual large class) – $X_i = 0$. The difference in Y 's value between the two, holding fixed all the independently operating causes for i (represented in W) is just what we mean by the *individual effect size of X on Y for i* . By inspection you can see that the average of the individual effect sizes across all the different individuals in a population is just the expectation of b in that population.

But b represents the supporting factors. And there's the rub. Why think that the causes and interventions for which we are purporting to calculate effect sizes are operating in small worlds where these probabilities exist rather than in the larger worlds of the kind that Kay and King worry about?

Indeed, before even we get to probabilities over a set of supporting factors, we should worry about supposing that there are these fixed sets of supporting facts in the first place to assign probabilities to – namely, that my first requirement that outcomes be fixed is satisfied. Why suppose that the ones necessary for one individual are the same as for others?

Consider, for example, the following discussion from Eileen Munro, who authored the UK's 2011 *Munro Review of Child Protection*:

Consider Person A, who was abused as a child. In this *particular* context, for this *particular* person (Person A), all the factors depicted in Figure [1.2] are necessary to bring about the outcome of becoming an adult perpetrator of abuse. A history of child abuse is by itself insufficient in Person A to cause the effect. It requires all the other factors to be present at the same time in order to lead Person A to perpetrate abuse on a child.

This may explain why some people go through periods of abusing children in their care and not abusing, because at certain times, some support factors will be missing or present. However, we are not proposing that the factors that are present for Person A are applicable to all. The conditions vary between individuals, as demonstrated in Person B (Figure [1.3]). With Person B a different set of *insufficient but necessary* factors combine to lead to adult perpetration. For Person B, a completely different set of factors is associated with being a perpetrator of abuse. Person B was not abused as a child, but a number of factors combine to create an environment for perpetration of abuse to occur in this particular person's life.

As depicted in Figure [1.4], Person C, like Person A, has undergone a history of child abuse, but Person C differs from Person A in that other factors never combined with this history. They are separated out in the cake diagram. For this particular person, adulthood does not include perpetration of abuse upon children. Although some of the same necessary conditions are present, they are insufficient unless combined with others. A whole set of supporting factors need to come together to make the outcome likely.

However, all of the above are *unnecessary* in the sense that another cluster of factors may be responsible for abuse in another person. So the six factors associated with Person A, for instance, are only *necessary* for that particular person and only necessary for that person when all combined together.

It's because of these kinds of problems with probabilities that I worry that talk of effect sizes is very often misplaced. The concept of probability on which it depends is too often stretched beyond where it will reach. We end up calculating something that has no real significance in the large worlds that effect sizes are supposed to help us navigate in. As a result, we adopt policies that fail and neglect others that could succeed in the setting at hand.

I turn now from concepts to look briefly at some other ingredients that are also necessary to constitute meaningful, useful theory.

Models and Narratives

Theory is pointless if it cannot connect with the world. And securing measurement techniques for its concepts is not

enough to do this. We need to know where and when and in what guise its laws and principles are meant to obtain. We do that with models and narratives. These play a major role in understanding the import of theories and in knitting together their claims and practices. Despite that, both were long neglected by twentieth-century philosophy of science, which under the influence of logical empiricism was keen on theory as well-formulated axiomatic schemes, as I noted in the section ‘The Melange of Theory Ingredients’. The Cambridge philosopher Mary Hesse is normally credited with forcing philosophical attention on models in her 1963 book, *Models and Analogies in Science*. This was a major influence on work developed by historian and philosopher of economics Mary Morgan, philosopher of physics Margaret Morrison and me, which thrust models into the centre stage in philosophy of science decades later, at the turn of the twentieth century, culminating in an influential collection by Morgan and Morrison, *Models as Mediators*. Now at last history and philosophy of science have turned to exploring and explaining the importance of narratives, again spearheaded by Mary Morgan.

Models as Mediators has a slightly misleading title. It could suggest that models are important because of their role in mediating between theory and the world. But one of the collection’s major themes is that, borrowing a phrase you will see more of when I turn to experiment, models ‘have a life of their own’, independent of the service they provide to theory. For instance, we probe models to see what they yield, often to suggest new results not envisaged beforehand or to illustrate what is possible not what is actual or to study what happens

when – contrary to what is true in any actual setting – some specific factor (like Coulomb attraction between charged particles in physics, or in economics the isolated effect of skill loss during unemployment on perpetuating unemployment) works all on its own. In this section though, my topic is theory. My claim is that without models, theory is just what is often meant by the term in common parlance – ‘just theory’, that is, with no foothold in reality.

Think about the model of the planetary system as a number of smaller point masses circulating a much larger point mass, affected only by the pull of gravity. This model simultaneously gives life to two abstract Newtonian principles that I discussed in the section ‘The Melange of Theory Ingredients’, $F = ma$ and $F_G = GMm/r^2$. Even supposing we have given real content to each of these symbols in these equations, so that F means force, m mass, a acceleration, r a distance of separation and G the constant of gravity, then these equations just tell us about relations between abstract quantities. They don’t tell us about anything that happens. The model tells us about things that can actually happen and which happen to things that, albeit idealised, are much like things in the real concrete empirical world that they are meant to represent – the little masses that represent the planets circulate the big one that represents the sun in elliptical orbits.

Or consider the very basic starting demand and supply equations in economics: $Q_d = a - bP$ and $Q_s = -c + dP$. Again, these are just relations among abstract quantities, even once we know that Q_d is quantity demanded, Q_s is quantity supplied, P is price and b and d

are demand and supply elasticities that are deemed constant, with a and c also constants. Now the model. The model has consumers and producers in it. The producers produce a single good that consumers want but that is not a necessity of life. In this small world there are no substitutes for the good, consumers expect the price of the good to stay fixed and there is no product differentiation – all instances of the good are the same and they are each sold at the same price. Consumers get more utility from buying the good at a lower price and producers from selling at a higher price. The constants are all set by circumstances outside the model. In this small world the good sells at a price where the supply and demand intersect. The equations are exemplified in something that actually happens in that world.

You see a further example in Figure 1.8, which is an idealised model for what happens in Millikan's oil drop experiment, which I discuss in the section 'You Can't Build an Experiment without a Gigantic Meccano Set'. There I point out the need for a model like this to make sense of what is going on in the experiment. Here I use it as another example of an idealised model that gives life to equations in classical mechanical and classical electromagnetic theory. It not only exemplifies each equation by showing something that happens in accord with it but it also shows one way in which these equations from different theories can be knitted together to produce an outcome jointly.

Narratives do a similar job. Here is how UCLA historian of science Norton Wise describes their importance, using the idea Mary Morgan introduced of 'colligation':

One quite general insight involves the coherence-making power of narratives, their capacity to fit together in a coherent pattern a variety of elements that otherwise would seem disparate. In interesting ways, this power of narratives for colligation recovers one of the central features of [Thomas] Kuhn's paradigms, their holistic character as pictures or patterns. That is the same power that he and others have ascribed to historical narratives.

My own efforts focused initially on how computer simulations generate narratives that provide understanding of otherwise inscrutable processes, such as the formation of snowflakes, and how simulated movies of unobservably fast chemical reactions provide visual narratives that unveil highly contingent processes. Most recently I have been looking at earlier historical cases, such as the narrativizing role of Carnot diagrams in understanding the Second Law of Thermodynamics and Maxwell's use of 'physical analogies' as fictional narratives to make lines of force in electromagnetic fields comprehensible. All of these model-based narratives do their work by making the processes they investigate seem familiar in the everyday world of concrete and sensible things.³⁰

And: 'When coupled with models, the narratives tell us how the models work and how they relate to the world.'³¹

The narrative that accompanies the simple supply/demand model I described serves as a familiar example. Here is a typical version of it:

Consumers typically look for the lowest cost, while producers are encouraged to increase outputs only at higher costs. Naturally, the ideal price a consumer would pay for a good would be 'zero dollars'. However, such a

phenomenon is unfeasible as producers would not be able to stay in business. Producers, logically, seek to sell their products for as much as possible. However, when prices become unreasonable, consumers will change their preferences and move away from the product. A proper balance must be achieved whereby both parties are able to engage in ongoing business transactions to the benefit of consumers and producers. (Theoretically, the optimal price that results in producers and consumers achieving the maximum level of combined utility occurs at the price where the supply and demand lines intersect. . . .)³²

You will also see an extended narrative associated with the model for the Millikan experiment when I turn to that in the section ‘You Can’t Build an Experiment without a Gigantic Meccano Set’.

Diagrams, Illustrations and Graphs

Wise mentions the ‘narrativizing role’ of diagrams. But diagrams have other roles to play as well in constituting theory and delineating what it says.

The historian of science David Kaiser has shown in his book *Drawing Theories Apart* how Feynman diagrams (named after the physicist Richard Feynman, who shared the Nobel Prize in physics in 1965 for contributions to the development of quantum electrodynamics) became a ‘calculational *tool*’, or a way for theoretical physics to be practised, in the postwar years. As tools, they were adapted to many uses: ‘where [Freeman] Dyson had *derived* Feynman diagrams’ form and use from quantum field theory, [Geoffrey]

Chew refashioned the diagrams as a tool with which to *eliminate* the quantum field theory altogether'.³³ Diagrams both followed from and preceded theory. We see this in biology as well, at least with respect to hypotheses. My colleague at UCSD, philosopher William Bechtel, and his colleagues have suggested that, as 'representational tools', diagrams do not just visualise how a biological phenomenon might be produced and what is still unknown about its mechanism, but they also delimit hypotheses about that mechanism. That is, a diagram 'provides constraints and affords possibilities for inference that influence hypothesizing about and investigating further elements of the proposed mechanism'.³⁴

Or, for a view from the socio-economic sciences, consider what Nobel Prize-winning economist Angus Deaton says about how he arrives at theoretical insights, like Anne Case's and his recent conceptualisation and account of the dramatic rise in the USA of 'deaths of despair' – deaths due to suicide, drug and alcohol poisoning and alcoholic liver disease:

We all have our preferred methods that we think are underused. My own personal favorites are cross-tabulations and graphs that stay close to the data; the hard work lies in deciding what to put into them and how to process the data to learn something that we did not know before, or that changes minds. An appropriately constructed picture or cross-tabulation can undermine the credibility of a widely believed causal story, or enhance the credibility of a new one; such evidence is more informative about causes than a paper with the word

‘causal’ in its title. The art is in knowing what to show. But I don’t insist that others should work this way too.

In sum, theory is nothing by itself, it can exist only in virtue of a great variety of other parts of the vast Meccano set of science. We can’t have theory without concepts, and it takes a melange of other activities from science to stabilise those concepts, to zero in on what bits of the world they pick out and to measure them. Diagrams often play an essential role in fleshing out what theory is claiming and narratives and models in showing what sense they make together. None of this is scaffolding that can be kicked away leaving theory free standing. This is the very stuff that constitutes theory. It would be like taking away the bricks and mortar or the steel girders and still expect to see a building standing there.

Experiments and the Testing of Theory

Theory is just half of the right-hand side of the equation ‘science = theory + experiment’. What about the other half: experiment?

The philosopher Karl Popper famously taught that it’s not real science if it’s not falsifiable. That is, if a principle or a theory is to count as genuine science it must make clear claims about the world that we can check to see if they are true. Should these empirical claims turn out false, the theory then is false, by simple deductive logic: if theory *T* implies observation *O* and we find that *O* is not the case, then *T* cannot be the case.

Popper was motivated by his concerns about the theories of Freud and Marx. These, he claimed, never imply anything specific; they can explain anything that happens, no matter which way things go. Popper insists that proper science should make claims that are so determinate that they could be shown to be false. No faffing about, what follows is *exactly this*. This is echoed in more recent philosophical work by Christopher Hitchcock and Elliott Sober in their worries about theories that merely accommodate (i.e. are consistent with) but do not really imply the data: if a theory is 'sufficiently plastic that it can accommodate any data that may come along, it is in no position to make predictions about what data will come along'.³⁵

When it comes to falsifiability, experiments – the other part of our equation – play a central role. Theories predict outputs but only from specific kinds of inputs. You can calculate the distance that a heavy body falls from a resting place in a given time t by the formula $s = 1/2gt$, where g is the acceleration due to gravity. You can do this *supposing* that there is no wind resisting its fall, no magnet pulling it back up, no golf balls striking it and so forth – that is, supposing gravity is all that affects its motion. This formula doesn't predict what happens if other factors than the pull of gravity intrude. So for a fair test of the formula, you need to look at how far a body falls in the very special circumstances – the special small world – where only gravity is at work. Sometimes situations like that happen naturally. In those cases, we merely have to observe what happens. But usually you have to manufacture them: you have to construct a *controlled experiment*. A controlled experiment to

test a principle is a very special environment, which is so regimented and shielded that nothing can affect the outcome except the inputs that the principle represents – nothing affects the body's fall than the pull of gravity. We then look to see if those inputs yield the outputs predicted by the theory. The controlled experiment gives us a fair test of the theory.

Experiments: A Life of Their Own

That gives us both components in the equation, theory and experiment. But it gives us a very narrow picture of experiment. Experiment is meant to test theory. This was the dominant view in philosophy for a long time – from what is sometimes called the ‘theory-centric’ view of science. On this view, experiments are not there to create new ideas but merely to make you feel comfortable with the ideas you already have. But experiment has much more to do than that. As philosopher Ian Hacking notes, ‘[e]xperimental work has a life of its own’.³⁶ Historian Ted Porter says the same: ‘[Theory testing] is often taken as the decisive role of experimental quantification in the practice of science. It is not. Researchers on topics that lack mathematical theory are often equally assiduous in reporting methods as well as results in quantitative form, and filtering out findings that cannot be so expressed.’³⁷ So, if they are not just suggesting and testing theory, just what do experiments do in this life of their own? Quite a number of things.

Exploring

What philosophers call ‘exploratory experimentation’ is conducted without any regard to theory; it is not intended to test theory. Nor is it intended to develop theory or estimate values of theoretical parameters (like the gravitational constant) or to fill in missing details. For the most part, it does not take direction from theory at all, even low-level theory. It is there to probe, to explore, to discover. It is blithely living a rich life of its own, far from the beady eye of theory.

Philosopher and historian of science Richard Vagnino provides us with a nice example, illustrating the care and detail that go into this kind of exploratory experimentation:

Italian physicist Luigi Galvani’s work on *animal electricity* began in the early 1770s, culminating in the publication of *De viribus electricitatis in motu musculari, Commentarius* in 1792. While discussions of *De viribus* tend to focus on two significant and highly influential experiments, the text itself recounts, often in exhaustive detail, a litany of experimental manipulations which took place over the course of the preceding decade. Both experiments describe the production of muscular contractions in a frog ‘prepared in the usual manner’, which involved isolating the lower half of the animal by cutting just below the upper limbs and connecting the legs to the spinal cord by way to the crural nerves. Galvani’s ‘first’ experiment concerned the production of contractions at a distance when a scalpel was placed in contact with the exposed nerve at the same time as a spark was produced by a nearby electrostatic generator.

The event is, on Galvani's retelling, a fortuitous one, owing something to chance. Luck, however, figures little in what follows.

Beginning in 1781, Galvani undertook a series of experiments, each marked by a minor variation to some feature of the experimental setup that initially produced the phenomenon: the scalpel was grasped at various points along its surface; it was replaced by glass and then iron rods; the distance between the frog and the generator was varied, as were the materials affixed to its exposed spine and nerves. A similar battery of manipulations was employed using muscle instead of nerve. Again, Galvani made use of a wide variety of arrangements with the aim of establishing exactly which conditions allowed for or inhibited the production of the phenomenon. His observations collected over the course of this 'long series of experiments' led him to 'ascribe the phenomenon of such contractions to electricity' and perhaps more importantly 'to note the conditions and as it were certain laws by which it was governed'.³⁸

Galvani's 'second experiment' involved the addition of a metallic hook to the prepared animal's spinal cord which was then placed in contact with a silver plate. The frog was grasped by one leg while the second was allowed to come into contact with the same metallic surface, producing rhythmic contractions as the leg rose and fell with the completion of the circuit. As with the first, the second experiment is better understood as the product of a large number of closely related but discrete experiments. Galvani altered both the metals in the hook and the plate, beginning with a bronze hook and iron grating. He likewise investigated the effects of 'gum,

resin, stone, wood' and materials known to 'transmit little or no electricity'.³⁹ Further iterations controlled for the effects of 'atmospheric electricity', the composition of the circuit, and the effects of submerging the preparation in different conducting and non-conducting fluids.⁴⁰ This second experiment ultimately led Galvani towards the conclusion that 'double and opposite electricity' was to be found 'in the prepared animal itself'.⁴¹

The character of Galvani's experimental practice during this period is striking not only because of the volume and diversity of experiments executed, but also in that it resists straightforward analysis under the familiar banner of experimentation as hypothesis testing. This is not to say that his investigations lacked direction, or that the countless tweaks and subtle alterations to the same basic experimental setup were done at random. Rather, the experiments described in *De viribus* are a good example of *exploratory experimentation*, which differs from *theory-driven* research in that it does not set out to confirm or falsify a well-defined, predetermined hypothesis. Exploratory experiments are often carried out when the phenomenon of interest lacks an established theoretical framework, or such a framework is in flux or under development, and often involve a large volume of open-ended investigations, often with what Friedrich Steinle describes as the 'desire to obtain empirical regularities and to find out concepts and classifications by means of which those regularities formulated'.⁴² Even in cases where a single, definitive experiment was reported – as was, for instance, the case with Galvani's 'first' and 'second' experiments – those instrumental setups, and the specific outcomes they were designed to detect, were often the product of a long series

of open-ended experiments designed to systematically test the influence of various parameters on the phenomenon of interest.⁴³

Creating Phenomena

Hacking tells us that experiments create new phenomena that never before existed, like the Zeeman and Stark effects, which are the shifting and splitting of spectral lines in light emitted by atoms and molecules, both of which gave rise to Nobel Prizes. The second, which was important in the early development of quantum mechanics, is splitting due to an electric field. The first is splitting due to a magnetic field. It was first produced in Leiden in 1896 by the Dutch physicist Pieter Zeeman. It plays a significant role in our knowledge of the electron. As historian Theodore Arabatzis reports, the Zeeman effect ‘not only provided evidence for the existence of the electron but also led to a specification of two of its properties, its charge to mass ratio and the sign of its charge’.⁴⁴

Hacking explains that we speak of physicists *discovering* these effects, but, he claims, this is misleading. It downplays the battery of experimental practices, including the construction of ingenious apparatuses that allow for their stabilisation and reliable reproduction. Only with this network of background activities and products of science do such effects turn into actual phenomena: noteworthy events that occur regularly in very specific – small-world – circumstances. Hacking’s point is that we have been so caught up with the use of experiment to test theory and with the

theoretician's attempts to explain phenomena that we have ignored how experiments often create the very phenomena to be explained.

Another of Hacking's examples is the Hall effect produced in 1879 by the American physicist Edwin Hall as part of the research for his PhD at Johns Hopkins University. The Hall effect occurs when a current passes through a conductor in a magnetic field. The magnetic field exerts a force on the moving charges perpendicular to their line of motion. This pushes them to one side of the conductor. The charge build-up on the conductor creates a voltage difference at right angles to the field and to the current.

Did Edwin Hall's experiments really create the Hall effect? Can it not take place in nature without the need for experiment? Hacking's answer is 'yes and no'. He explains:

If anywhere in nature there is such an arrangement, with no intervening causes, then the Hall effect occurs. But nowhere outside the laboratory is there such a pure arrangement [...] I suggest, Hall's effect did not exist until, with great ingenuity, he had discovered how to isolate, purify it, create it in the laboratory.⁴⁵

Reconstituting Phenomena

Besides testing theory and creating phenomena, there's lots else we do with experiments. We use them to help develop new concepts, to fill in missing gaps in narratives and to understand better how various measurement

techniques work. They not only produce new phenomena never before seen but we use them to find out more about phenomena that we are already familiar with. Often this involves getting a better measure of some of properties of the phenomena: like using a superior radar gun to find that a car's top speed is 220.335 kpm rather than 220 kpm. Sometimes, though, in the effort to find out more about an already familiar phenomenon experiments tell us that we weren't that familiar with it in the first place. These kinds of cases call for a more drastic, qualitative redescription of the phenomenon. When this happens, philosophers say that the phenomenon gets 'reconstituted'. Let's look at this in more detail as another illustration of what more we can do with experiments than just test theory. I use an interesting case in molecular biology, as described by philosopher Andrew Bollhagen:

Kinesin, the molecule in Figure 1.5, binds to cargo in a cell and 'walks' them down microtubule 'trails'. A tiny walking molecule is a fascinating phenomenon. Naturally, scientists are interested to describe it. This is no easy task. You can't just *watch* them walk. Kinesin is too small.

So, scientists used what they *could* see in video microscopes – relatively large beads being pulled along immobilized microtubules by single kinesin molecules or, alternatively, microtubules being pulled around by single immobilized kinesin molecules – to describe what they couldn't, the stepping of kinesin molecules. Their experimental design resembles the set-up in Figure 1.6 except that the vesicle in the figure is, in the experiment, a bead big enough to be microscopically observed as kinesin carries it along a fixed microtubule. Other versions of the

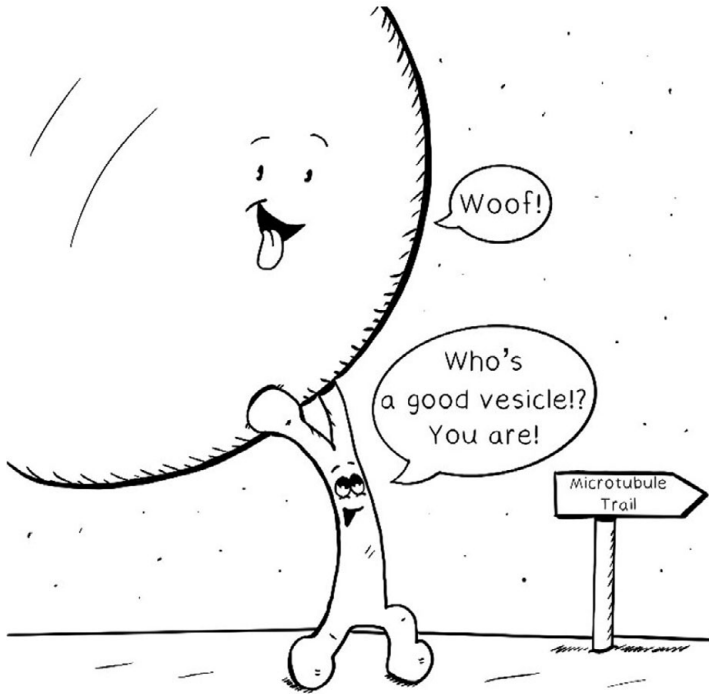


Figure 1.5 Kinesin walking cargo down a microtubule trail
Drawn by Nicki Shaw/@nickisdoodles, especially produced in black and white for this volume. Both Bollhagen and I are grateful for Shaw's permission to use it. Thanks Nicki!

experiment 'inverted' this design, like in Figure 1.6, where kinesin is fixed, and it pushes the microtubule along.

For over a decade, such studies led scientists to think that kinesin walked kind of like humans – one foot at a time, with each foot stepping forward and passing the other. Why did they think this? One reason is that they observed single kinesin-driven beads moving quite a

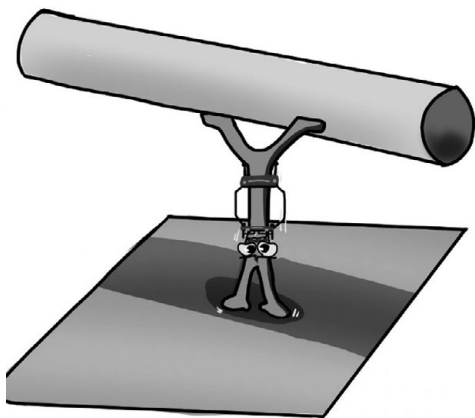


Figure 1.6 Kinesin pushing microtubule along
 Drawn by Faith Bollhagen especially for this book. Thanks to Faith from both Andrew and me!

distance without drifting away from the microtubule. Thus, at least one kinesin 'foot' must remain bound to the microtubule as it steps – the molecule doesn't 'jump' its way down the trail. This further implies that kinesin must *coordinate* its stepping so that each foot takes a step only after the other is firmly planted.

Setting out to study this 'hand-over-hand' walk in more detail, Hua and colleagues (2002) varied the basic design seen in Figure 1.5 using a modified kinesin with a 'stiff' neck, ensuring that whatever torque was generated as kinesin walked would be communicated directly to the microtubule which would observably rotate. These rotations could be used as a measure of torque. However, they didn't observe the expected rotations. This wasn't because there was something wrong with their design. Surprisingly, these researchers found, the molecule simply doesn't produce torque when it walks! This led

them to re-evaluate the familiar description of the molecule's step. Rather than walking 'hand-over-hand' – which would generate torque – the molecule walks like an 'inch-worm', they suggested, with one head always in the lead and the other stepping up from behind.⁴⁶

As the story illustrates, experimental efforts to further describe a phenomenon can lead to drastic revisions of 'familiar' descriptions. As philosophers like Bollhagen⁴⁷ put it, experimentation can lead researchers to 'reconstitute phenomena'.

You Can't Build an Experiment without a Gigantic Meccano Set

So experiment, like theory, has a life of its own. It is no mere handmaiden to theory. But what about all these other scientific endeavours I defend? What's so good about them? Here's *one* good reason we need to take them seriously, to ensure they are done properly and that they are up to the jobs we set them. Maybe you love theory and experiment – they are what concerns you about science. Still, there's no way your concerns can stop there. Without a good many of the other pieces in the Meccano set of science, *there simply isn't any real theory, nor experiment*. These other endeavours are part of the very stuff that makes them up. I've already discussed this in the case of theory. Let's look at a real case to see the kinds of pieces from the Meccano set that might be used to constitute an experiment.

The easy place to start is with concepts. As I discussed earlier, the principles of a theory are couched

in terms of concepts. The charge of the electron is -1.602×10^{-19} Coulombs. This principle doesn't say anything unless the concepts 'electron', 'charge' and 'Coulomb' are well delineated. How does that happen? It took a lot of complicated experimental work to stabilise the concept of the electron – to settle on a fixed core of characteristics that pick out what an electron is supposed to be. Each of these experiments only makes sense – they only are what we think they are – supposing a vast array of other concepts are in place.

And not just *concepts*. An experiment always needs a *model*. You see examples in Figures 1.7 and 1.8 for the

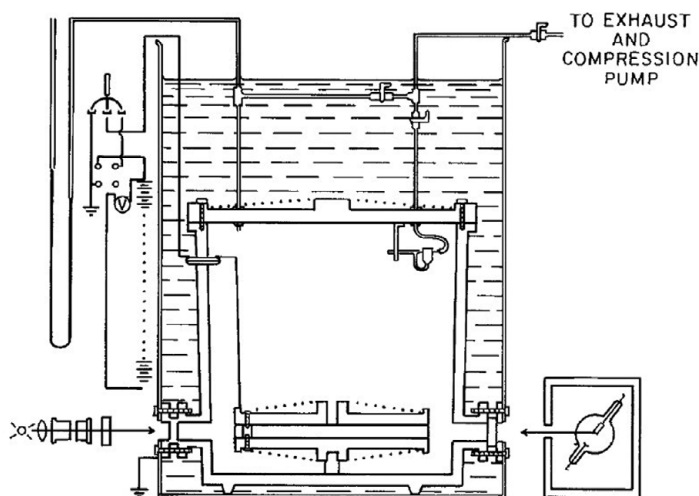


Figure 1.7 Millikan's own model of his measurement device
Reprinted figure with permission from R. A. Millikan, 'On the Elementary Electrical Charge and the Avogadro Constant', *Physical Review*, 2, 109 (1913), by the American Physical Society

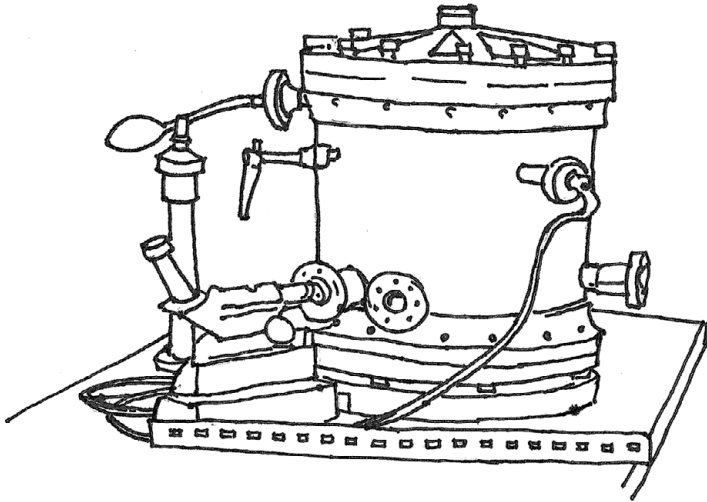


Figure 1.8 Millikan's actual apparatus

Drawn by Adrian Harris especially for this book. Thanks Adrian!

Millikan experiment to measure the charge of the electron, which I describe later in this section. The model shows just what the experiment is. It is a blueprint that the actual physical devices and procedures must live up to if the experiment is to do what it is supposed to. Then we need *specifications* – just how will that model be instantiated? The Millikan model asks for a very lightweight sphere to carry the electrons. What should be used for this? Millikan's original attempts were with water drops. Harvey Fletcher, Millikan's later associate on the experiment, suggested oil drops. These were much better because they don't evaporate readily. But of course, what counts for 'better' depends on what the model requires. The model for the Stanford Gravity

Probe B that I describe in Chapter 2, in the section ‘Even Physics Isn’t All Physics’, calls for gyroscopes. It took years of patient hunting and trials, plus lots of ingenuity and patient study, to decide that these would be fused quartz spheres.

Both to experiment and to ensure your concepts have a grip on the world, you need ways to *measure*. As I urged earlier, these again require a great deal of work in the background and of different kinds. To measure you need not just techniques, devices, technology and all that’s involved in designing and implementing them. You also need to know *why* these can measure the features they are supposed to.

For illustration, let’s look in some detail at Millikan’s experiment to measure the charge of the electron and thus establish our principle *EC*. In the experiments developed and conducted by Millikan and Fletcher over a handful of years around 1909, oil drops were injected into an air-filled container where they picked up charge from the ionised air inside. Recall my claim that without a model, there’s no experiment. Millikan’s own model for this one, from a 1913 paper,⁴⁸ is pictured in Figure 1.7. It is the model to which he had to build the real device, pictured in Figure 1.8.

Here’s what Fletcher and Millikan did in the experiment. They released tiny oil droplets one at a time into a chamber between two parallel capacitor plates, sprayed the droplets with electricity, then watched them drop across about a centimetre of fall while adjusting the charge on the capacitor to control the size of the electric field acting on the falling droplet. Eventually the negatively charged oil droplet

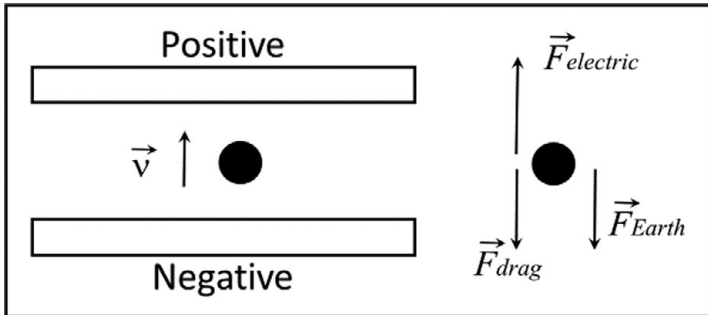


Figure 1.9 Why Millikan's and Fletcher's results constitute a measurement of the charge on a droplet
 Drawn by Adrian Harris especially for this book. Thanks Adrian!

comes to a stop hovering between the two charged plates, simultaneously pulled down by gravity and up by electric attraction. Millikan and Fletcher then recorded the exact size of the electric field at which the droplet came to a standstill. From this they calculated the charge on the droplets.

Now we come to the defence that these elaborate procedures measure what they are supposed to. I have just described what Millikan and Fletcher did in their experiment. Why does that count as providing a measurement of the charge on the droplets? The answer is represented in a different kind of model of the experiment, the one pictured in Figure 1.9.

The droplet is pulled down by the force of gravity and up by electric attraction. Due to air resistance, it also feels an upward drag force proportional to its velocity. The droplet comes to rest when the combined upward forces are just equal to the downward force.

Millikan measured the charge q on the droplet by adjusting the electric force ($F_{electric}$) till the droplet was at rest in the face of the force of gravity (F_{earth}) plus the upward drag force (F_{drag}). Then he could calculate $F_{electric} = qE = F_{earth} \oplus F_{drag}$. The two terms on the right, as well as the size of the electric field, E , are determined by a combination of theory and measurement. The charge q is supposed to be due to free electrons on the oil drop. It turns out that (as expected) this was always in multiples of the same number q_e , which Millikan calculated to be $4.774 (\pm 0.005) \times 10^{-10}$ electrostatic units, which is very near the value in our principle *EC* from the earlier section ‘The Centrality of Theory and Experiment, Knowledge and Observation’.

This shows that charge is discrete and tells us what the minimum is – this is the charge assigned to a single electron. Though the drops differ in charge, for each drop, $q = nq_e$. So the charge q_e of a single electron can be estimated by measuring q for a number of drops.

Millikan’s earlier experiments with Louis Begeman used water drops. They gave imprecise results because the water drops evaporated too quickly. Apparently mercury and a few other substances were considered as substitutes. In the end Millikan and Fletcher used oil drops, as Fletcher suggested – clock oil. You can see here how just this one tiny aspect of the Millikan–Fletcher experiment relies on a vast network of previous activities having been done well. As Millikan explains, ‘mankind has spent the last three hundred years in improving clock oils for the very purpose of obtaining a lubricant that will scarcely evaporate at all’.⁴⁹

Note what else Millikan says about the choice of droplets in his Nobel lecture:

[T]o take on the smallest possible charge [the charge carrier used in the experiment] had of course to be the smallest spherical body which could be found and yet which would remain of constant mass; . . . A non-homogeneous or non-spherical body also could not be tolerated; for the force acting on the [charge carrier] had to be measured by the speed of motion imparted to it by the field, and this force could not be computed from the speed unless the shape was spherical and the density absolutely constant. This is why the body chosen . . . was an individual oil droplet about a thousandth of a millimetre in diameter blown out of an ordinary atomizer and kept in an atmosphere from which convection currents had been completely removed by suitable thermostatic arrangements.⁵⁰

Just think what background work over decades had gone into devising thermostatic methods to keep out convection currents – and indeed into developing the very concepts of ‘convection current’ and ‘thermostatic’.

As I mentioned, as a result of his experiment, Millikan’s reported that the charge of the electron is $4.774 (\pm 0.005) \times 10^{-10}$ electrostatic units. Why should we take his experiment as confirming this claim? To do so, we clearly must trust that the equations of classical physics he employs are accurate enough in this setting for the job. Perhaps you think this is trivial: of course we can trust in that. I agree that probably we can. But I do not think that this is trivial. When you take Millikan’s experiment to

confirm that the charge of the electron is very near 4.774×10^{-10} electrostatic units you are trusting in science. And it is important to realise that, because not everything that we trust in and have trusted in with good reason is correct and not all of it can do the jobs we might expect of it. There is undoubtedly a vast amount of evidence in support of those equations. But every experiment that we think supports those equations (or their sufficient accuracy for Millikan's purposes) requires the same kind of mix of theoretical and concrete knowledge we see in Millikan and Fletcher's – like the characteristics of clock oil or how to seal the container to prevent air drafts. So, to back up that what Millikan did can count as an experiment to test principle *EC* we need an ever-expanding set of Meccano pieces.

Then think about the actual application of those equations. The equations are couched in highly abstract terms: the electric force, the force of gravity, the drag force. Now, Millikan can't just say, 'let there be an electric force' and then there it is. He has to do something – something concrete – to create that force. He has to figure out what it takes to physically instantiate this abstract theoretical concept in his setting. This is always the case. To do an experiment that relies on an abstract equation, it is necessary to create the features of the concepts in that equation in the real world. That is not easy. Consider $F_{electric}$ in Millikan's own words:

The potential difference is not reliably given by the battery voltage. It is measured in 6 parts with a device accurate to 1 part in 2000. This device in turn is calibrated by a second, whose accuracy is both certified

and independently measured by yet a third device. And 5,000 readings calibrated in two different ways were shown to be consistent. The electric fields were produced by a 5,300-volt storage battery, the P.D. [potential difference] of which dropped on an average 5 or 10 volts during an observation of an hour's duration. The potential readings were taken, just before and just after a set of observations on a given drop, by dividing the bank into 6 parts and reading the P.D. of each part with a 900-volt Kelvin and White electrostatic voltmeter which showed remarkable constancy and could be read easily, in this part of the scale, with an accuracy of about 1 part in 2,000. This instrument was calibrated by comparison with a 750-volt Weston Laboratory Standard Voltmeter certified correct to 1/10 per cent, and actually found to have this accuracy by comparison with an instrument standardized at the Bureau of Standards in Washington. The readings of P.D. should therefore in no case contain an error of more than 1 part in 1,000. As a matter of fact, 5,000 volt readings made with the aid of two different calibration curves of the K. 8c W instrument made two years apart never differed by more than 1 or 2 parts in 5,000.⁵¹

There was a great deal more theoretical activity as well that played a part in the Meccano set of science that supports the claim that Millikan and Fletcher's experiments provide good evidence that the charge of the electron is near 4.774×10^{-10} electrostatic units. But I think you have seen enough already to get a real sense that even when it comes just to experiments themselves:

It is not all theory and experiment after all.