Acting to know: a virtue of experimentation

DRAFT, though pretty final. to appear in *Synthese*

Adam Morton

adam.morton@ubc.ca

We often act in order to know. One celebrated instance of this is scientific experimentation, but as epistemic acts experiments in science have a lot in common with a variety of everyday activities, such as asking for the time or wiping your glasses. The important feature is that the act succeeds only if knowledge results. (The intention is usually directed to getting at the truth on some topic, and if the intention is satisfied *because of* the action, then the result is knowledge. What if a true belief results, but in an unintended accidental way? That's complicated – see Morton 2012a – but it is not the topic here.) Capacities of doing this well are thus both epistemic and practical virtues. In this paper I explore one central virtue of experimentation, which I eventually name the virtue of experiment-shopping. It is the virtue of knowing if an experiment is worth performing, and although some obvious examples of it are found in scientific practice, I believe it is important throughout our intellectual lives. To call this capacity a virtue is to link it to a particular kind of success, that of coming to know if an experiment is the one to carry out. You can also say that if the virtue is used then the determination whether to carry out the experiment is arrived at well, as long as you don't build into this any ideas of its following any particular rational method. All I mean is that it is sensitive to the factors that make experiments achieve their ends or fail. In the last section of the paper I connect the virtue to intellectual virtues in general. But to begin I discuss the ubiquity of experiment

1. **experiments everywhere**

Simple everyday experiments involve no special equipment, and ideas about experimental design are rarely consulted. Yet they fit the fundamental pattern that in order to learn something one does something, making information emerge which would not have otherwise. For example, you are on a committee interviewing candidates for a job which involves dealing with a range of people on a range of topics. A letter for one candidate says that he does not suffer fools gladly, that he is inclined to be brusque and visibly impatient with people who he takes to be confused or wasting his time. The letter may be exaggerating or malicious, and you would like some better evidence. So you ask a stupid question. You put a lot of thought into your stupidity, and at the interview you make an elaborate suggestion about his area of expertise that rests on a conflation of two similar-sounding words. The outcome is unpredictable. It may be that he seethes with contempt, that he patiently and tactfully unravels the confusion, that he deflects the question, or something in-between. Some of these outcomes will tell you more than others.

That experiments are causal interactions to epistemic ends was noted by Ian Hacking some time ago. See chapters 2 and 9 of Hacking 1983. The theme has been ignored in a lot of more recent work but see Radder 1996, and Woodward

2003. I do not find in any of this otherwise admirable work recognition of the continuity between the scientific and the everyday, of the kind that the interview example illustrates.

Several basic points are illustrated by informal experiments such as the interview case. Most basic of all, the experiment is an act – the realisation of an intention by causing some change in the world - which can be well-thought out or not, and can be successful or not. It is an act whose purpose is epistemic, but the thinking behind it does not fall into a traditional category of belief-directed reasoning. One reason for this is that what belief it is that results depends on something unpredicted that happens outside the person's cognition. The opposite is also common, where you form a belief in order to achieve a practical aim, as when you look at the weather forecast in order to choose the best day for the picnic, but we are now concerned with walking to the hill where you can see the clouds in the west. Often, of course, we perform an action in order to gain knowledge in order to be able to do something: walking to the hill in order to predict the weather in order to time the picnic. In the interview example you do the experiment to learn if the candidate is tactful in order to appoint the best person. Epistemic and practical are usually entwined. Experimentation overlaps with thinking when a person wonders what she thinks on a topic ("would we be happier if nothing was secret", "are there really fundamental rights") by posing various hard questions to herself and seeing how she reacts. This is a kind of self-experiment similar to those one performs conversationally with other people. With oneself or with others, it produces information, material for thinking about, which one could not have got just by thinking or passively perceiving.

The interview experiment is also typical in that it has a cost. In asking the stupid question you make the candidate think less of you, and this may have repercussions. You use up time in the interview that could be used on other topics. You affect the atmosphere later in the interview. If you are thinking whether and how to perform the experiment you have to formulate these costs and risks, which have to be considered together with the benefits of the information you might gain.

Thirdly, this experiment like many others has an unpredicted outcome. The unpredictedness is hard to state carefully. It is reminiscent of epistemic paradoxes such as Kripke's observation that when one has good evidence for a belief one also has good evidence that evidence against it is likely to be misleading, and therefore to be ignored. (The idea comes from Saul Kripke, but its first appearance in print was p 148 of Harman 1973.) In the interview case you can expect several possible general types of response from the candidate. You may well consider some of these more likely than others. But you don't take yourself to know what the outcome of the experiment will be. Often one is surprised when an experiment turns out as it does, but in planning the experiment one does not take it for granted that it will not turn out this way. One does, though, make more elaborate contingency plans for following up the more expected outcomes than the less expected ones. In planning the interview you may think that it is pretty unlikely that the candidate will simply ignore the mistake in the question, but you still prepare a follow-up question to highlight it in case he does. You think it pretty likely that he will use some abusive language to you, and so on the one hand you prepare a pretence of injured pride in order to test his reaction to information that he is causing distress, and on the other hand you think how to get across to him later that no harm was done (except to his job prospects.) The situation also resembles the strategic interactions studied in game theory. There although one player may have expectations about what another is more or less likely to do, a prediction of the other's actions cannot be separated from a decision of what to do oneself (since the other is basing their action in part on a prediction of what the first player will do.) An experiment is a game against (or with) nature in this respect: your moves depend not on what you expect the other to do but on what values the possible outcomes have for you. (Considering experiments as games against nature opens up formal ideas, due to Abraham Wald. See Gigerenzer and others 1990. The similarity of experiment to strategic interaction described here is more basic than, and independent of, these ideas.)

These features are found in formal scientific experiments, too, and in innumerable everyday information-eliciting procedures. One finds out if someone is awake by whispering a message; one finds out if there is water in the well by dropping a pebble into it; one finds out if the enemy is still out there by sticking one's head above the parapet. It is important in distinguishing these from non-experimental inquiry to emphasise that the procedure has a causal effect that allows the information that would not otherwise be available to be produced. The nearest that simple perception comes to this is in some uses of one's tactile sense, as when one feels how many coins are in one's pocket by actively moving them around. (Fingers are special in that they both move and feel, in ways that are often inseparable.) Just opening one's eyes is an action, and can be intended to produce a situation in which information is available, as is flipping a light switch, but these should be seen as at most limiting cases of experimentation. I shall take it that in even very informal experimentation one performs an action, the action produces a situation that would not otherwise have existed, and the existence or features of this situation provide data one wants in order to form opinions. When a person opens her eyes she is producing a situation - her eyes being open and light striking her retinas - in which information is available, but it is information about the scene rather than about the state of her eyes or the effects of opening them. When the bandages are removed from someone recovering from an eye operation and she first opens her eyes, that *is* a real experiment.

2. **success**

An experiment has gone well when the intended situation has been produced and it provides information that is relevant to the question at issue. (In science, experiments are usually directed at fairly specific questions; less so in everyday life, though in both there is a virtue of asking and probing at the right level of generality.) Then it has been successful, in more than the minimal sense of producing knowledge. (It is a frustrating success when you get the information you wanted, but in terms of the questions that matter to you, you are none the wiser.) Experiments often succeed inadvertently: a situation is produced which is not among those anticipated and one is not prepared for the information that results. When you ask your stupid question a feature of the wording may produce a reply that reveals a completely different flaw than the one you were probing for. The experiment has then failed in that you did not get an answer to a particular question, and has succeeded in that you did get an answer to a more general question that also interests you, such as "is he qualified?" Suppose that events take a completely unexpected turn and all that you learn is that the candidate has halitosis. Then, I would say, the experiment has failed *as an experiment* though it has provided information which other intellectual virtues can use. (As when you turn the switch on the accelerator and blow every circuit in Geneva, thus failing to learn anything about the Higgs but a lot about the Swiss power grid.) If an experiment has gone well then it is an accomplishment and results in knowledge. If it is conducted well then it exhibits virtues, of planning and anticipation and use of resources. Of course a well conducted experiment will often not go well. (Thinking of experiments as primarily sources of information is of course very common in the philosophy of science. For a discussion of evidence that makes a place for the results of experiment see Achinstein 2001.)

What is the relation between knowledge, accomplishment, and the virtues of experimentation? I am interested in intellectual virtues that are epistemic in the sense that they concern the conduct of inquiry but also practical in that they aim at an accomplishment, the production of specific information. One very basic such virtue is the capacity to devise the situation that will produce the information.

Virtues are double-edged. In the external direction they are directed at the information-giving situations and the production of opinions from them. And in the internal direction they are directed at making use of the information, and thus at the cognitive economy of the agent, the use of her pattern of beliefs and desires and the shape of her reasoning. It is be a bad experiment if it produces loads of information which cannot be made sense of. And it is unreasonable to undertake an experiment if there is good reason to expect that instead of having the desired effect it will frustrate the experimenter's deeper aims. It is unreasonable for that agent at that moment even if it turns out to result in the perfect informative clue to the question.

Since experimentation is aimed at a result it is subject to a basic constraint of practical reason, the need to find a means to an end that accommodates other competing ends. You don't drive a Mercedes because although you would then be safe and elegant you would be hungry and indebted and harming the environment. Since experiments have costs, the experiment has to be designed so that it provides information without disrupting other projects. These other projects can themselves be epistemic, but there is not a lot to be gained by distinguishing between competition from epistemic and other aims, since there is no end of things it would be good to understand, most of which would gain from non-trivial experiments, and a very finite time for any single person to devote to them. Doing all but the simplest experiments means renouncing others. And accomplishing all but the simplest experimental or practical aim means renouncing others, of both kinds. So we might as well throw all a person's aims together into the same pan, all to be balanced against all.

One of the features of an experiment that is crucial to this balancing is the amount of light it might shed on questions the person has reason to be interested in. (Curiosity is a good reason, often.) The issue is impossibly complicated. An experiment - even a pretty trivial one as in the interview example - has many possible outcomes, and the facts any of these reveal can be inputs to many different lines of thought. There are several kinds of linked imponderables. What will the physical outcome of the experiment be? How much relevant information will it provide? How successful will one be in exploiting the information, to refute a conjecture, formulate a new one, or adjudicate between existing hypotheses? Against these imponderables there is one manageable fact, the likely cost of the experiment. (It's certainly not a given, since the consequences of the experiment-as-act ramify into the future, but it is usually more nearly something one can get a comparative grasp on than the other questions.) The ability to handle situations of this shape, with these uncertainties deriving from these projects, is my main interest in this paper.

One important kind of experiment is a continuation or repeat of an experiment that has already been performed. After the candidate responds with only mild irritation to your question you ask him an even stupider one, in order to find his explosion point. In science it is important that experiments can be replicated, and in everyday life we sometimes fail to repeat them in part because of the familiar fact that we underestimate the importance of a person's situation on her actions. (We think that having behaved one way a few times tells us that this is her constant mode of operation.) A repeated or continued experiment produces more evidence to add to the evidence we already have, so in deciding whether to do it we have to decide whether the cost is justified given that we could instead do other experiments or throw a party to celebrate the results we already have. One particularly important case arises when the initial data in an experiment suggests that the experiment itself is doing harm. This can happen when a drug being tested is worsening the condition of subjects. Then the experiment is giving information about its own cost, and this is relevant to the question of continuing it. A pre-scientific analog is sticking your head above the parapet. If you immediately attract enemy fire you are reluctant to repeat the experiment in order to get more information about the number of enemy shooters.

So questions of cost are ubiquitous in experimentation. Not all experimenters have to face the most intractable forms of them. The budget for many scientific experiments is set in advance, or at any rate severely limited, by allocations in a department budget, a research grant, or other similar factors. So taking cost in this very narrow sense, there is often an upper bound to how much a proposed experiment can cost. Still, within a fixed budget, variant experiments are possible, and the experimenter has to decide which ones to run. That means comparing different possible experiments, and to do this one has to face the unpredictability of their results and the problems of anticipating what one will be able to make of these results.

3. **intractability**

I am now in a position to describe the virtue that is the target of this paper. It is the capacity to evaluate possible experiments, in order to decide whether to do them. I do not mean simply the capacity to plan an experiment sensibly, maximizing the chances of getting desired results. I mean the externalist capacity actually to proceed when the objective situation will result in both the need for knowledge and the need for solvency (etc) being satisfied. There is obviously no such infallible capacity, and there are obviously many component skills of sensitivity to the environment and to one's own proclivities, different ones being relevant to different situations for different people. But this is a large part of what makes it a virtue and not a simple skill: its essence consists in getting a certain kind of result in a certain kind of situation.

One central consideration in the choice between experiments is that different experiments give different amounts of evidence. This can be a result of such familiar factors as sample sizes and the effort made to randomise within blocks. Generally speaking, the experiments that give more evidence cost more. Experiments that promise more significant evidence also tend to be more expensive. In one kind of experiment more varied samples are required, the randomization is more thorough, or the block structure allows protocols that might eliminate more alternative hypotheses. In another kind, more sensitive equipment is used, or it is applied to a richer variety of cases. The consequence is that one often has to decide how much and what quality evidence to try for. As a result, we do not, nor should we, always go for the most and the best. So how are these decisions made?

I do not think they can be made on the basis of a simple cost-benefit comparison. These are feasible - the project of making them makes sense - when there is a manageable variety of comparable values of specific outcomes and an intelligible probability distribution over them for every action under consideration. Under values of outcomes I am including gains of understanding and expenses of performance, and the actions in questions are ways of carrying out the experiment. In the simple ideal case there would be a series of ways of carrying out the experiment, graded in order of expense - you pay $1k and you get the basic experiment, you pay $2k and you get a more careful one, you pay $100k and you get a super one with many control groups and loads of randomization - and the likelihood of getting a given amount of information for a given expense could be assessed. But nothing like this is almost ever the case. There are problems of comparability and problems of prediction.

The major problems of comparability are between the costs of experimentation and the information gained. Suppose for the sake of argument that the costs can all be expressed in terms of money (though in many cases this does not seem plausible.) The benefit of an experiment is the light it throws on some uncertain question. The outcomes cannot normally be expressed in terms of units of information, as if outcome gives twice as much information as another. (Remember that in order to compare expected values we need cardinal comparisons of the values of outcomes, and not just an ordering of them.) Issues about comparability and the problem they make for cost-benefit or (equivalently) expected utility thinking are discussed in Morton 1990 and the essays in Chang 1996. Issues of incomparability have gone quiet lately, but they beg to be connected with questions about the value of knowledge raised in Kvanvig 2003.

Instead, the manageable way to think of the outcomes is as settling very simple questions, causing one to know their answers. Did the applicant lose his temper; did the subjects respond more quickly to the items they had been primed for? Then the benefits in question are the information these answers give to the questions of primary interest. Is the applicant likely to be a difficult colleague; is there an unconscious representation of some category of information, playing some given functional role? If we could measure the degrees of support that these possible simple answers give to the primary questions then we would have something to match against money. But the issue is notoriously hard and even with formalised simple hypotheses there is no consensus how to do it. The existing formal accounts of comparative strength of evidence will not apply, for example, when the hypotheses contain higher-order terms such as "there is some unknown factor which correlates phenomenon A with phenomenon B". And as noted above ordinal comparisons will not do: we would need numerical measures of evidential strength. All this is before we even try to introduce the different interests of the different hypotheses that might get the different degrees of support. Or factors other than support, such as understanding why a hypothesis might be true or how a causal mechanism might operate.

Some problems of comparability are mollified by the fact that an experiment often has a budget, with an upper limit, and we are often reluctant to leave any of it unspent. (We don't like returning any of the research grant, and we are not allowed to donate it to famine relief.) So some experiments are ruled out and a central question is simply "how can we best spend $N?" Even then, less expensive ways of carrying out the experiment proper will have other benefits, some of them epistemic. We could do our consumer choice experiment with a large group of subjects, with payoffs in real money so that their motives are realistic, or we can save money by having a smaller sample and paying them with tokens for a lottery, and spend the rest of the grant on database software which will allow us to categorize the results of this and other studies. Comparability then re-enters the picture.

The problems of prediction are if anything greater than the problems of comparability. As noted above, the outcome of each proposed experimental avenue is open as a matter of principle: if we had much confidence how it would turn out we would have less reason to do the experiment. So it is hard to have more than the roughest assignment of probabilities of what I called the simple answers just above conditional on variant experimental procedures. And given these simple answers there is the problem of predicting the support they will give to ideas about the questions of interest. No doubt a competent experimentalist will have thought out the consequences of various anticipated outcomes, so that she can say that *if* one of them occurs then evidence of a given force for or against a given hypothesis will have been gained. But she will know that if one does occur then in thinking out its consequences, for example in preparing her results for publication, she will see more alternative possibilities more complications. There is a kind of circular trap here: the more time she spends working out the likelihood that a hypothesis will have been confirmed to a given degree the less time she will have to do the same for other simple outcomes and other hypotheses, and the more indefinite her expectation of getting any particular degree of support for any hypothesis will be.

Consider a simple prediction-testing experimental situation. We know that general relativity predicts that the paths of particles will follow geodesics shaped by the presence of mass, and gives predictions about the exact paths involved. We are lucky enough to have a neutrino-measuring instrument on the moon and can measure the influence of the presence of the sun on neutrinos from a neutrino star. (This is evidently a science fictional experiment, so objections of unfeasibility or physical implausibility are to be put aside.) We can be pretty confident in advance that if the deviation of the paths of the neutrinos, for example in producing a double image, is exactly what general relativity predicts then we will have added confirmation for it, though it may be hard to assess how much. And we can be somewhat confident that if the deviation is extremely different then we will have significant disconfirmation for general relativity. Of course we would be surprised by either of these. The most likely outcome is something near to the prediction of accepted theory, with the difference ascribable to experimental error. But what will we conclude if the observed result is between these extremes? We will have to consider the possibility that we are wrong about the mass and shape of the sun, or the speed and mass of the kinds of neutrino, or the physics behind the neutrino detector. We may have made some relevant simplification in modeling the interaction of enormous and tiny objects. What will we say if two thirds of the particles are within the expected range but one third of them are weirdly deviant? Will that lead us, and others, to suspect that the theory is correct and some unknown factor is causing a random deviation, or that something physically mysterious is going on? It will obviously take a while for the physics community to digest such a result and predicting their verdict is not something you want to charge experimentalists with. (The capacities required to handle such situations are related to those discussed in Fairweather 2012.) The situation the experimentalist would prefer to be in is to be given a theory and a consequence of it that will appear in a novel situation, and a budget. Then the experimentalist doesn't question the budget but tries to produce the novel situation within its limits.

4 **experiment-shopping**

So, whether planning job interviews or testing relativity, we do not decide which experiments to run by doing a cost-benefit analysis. How we do it?

We do it by being good experiment-planners, knowing which and how much data we want to collect. There is an intellectual virtue here, a mixed epistemic-practical virtue. (For epistemic virtues see Zagzebski 1996 and Sosa 2001. My own approach is different, as suggested below.) It mixes the epistemic and the practical in that one's aims affect *how much* one knows, rather than the possibility debated in the 'pragmatic encroachment' literature (Fantl and McGrath 2010) of whether one's aims affect *whether* one knows. It is distinct from the experimentalist's virtue of ingenuity: being able to devise the setups that will force nature into the situations where unexpected things may happen. I have nothing to say about the psychology of the virtue in question, except that one place it is found is in the largely middle-aged experiment-managers who advise the ingenious ones on what they might try, approve and administer research grants, and generally shoulder the burden of deciding whether a data-producing project is worth the trouble. In saying that it is a virtue and that there are places to look for it I do not mean to claim that it is usually exhibited ideally, or even well.

The right way to approach intellectual virtues, I believe and have argued elsewhere (chapter two of Morton 2012b), is in terms of their conditions for success. What situations are they applied to, and what outcomes do they aim at? The virtue we are discussing applies when there are several actions one can perform whose main benefit will be to provide evidence relevant to some questions of interest, and which have different costs one would like to minimize. The outcome it aims at has two sides, knowing the answer to the question and being satisfied with having paid what one did for it. Finally we know it well enough to name it: call this the virtue of experiment-shopping. It is a skill of buying a good enough experiment at a low enough price, of actually accomplishing these, not just worthily striving towards them or blindly fumbling in their direction. I have argued that we do not exercise this virtue by calculating and comparing costs and benefits. In fact we do not evaluate the desirability of outcomes and the likely results of courses of action independently at all. We consider whole situations, in which we or others face uncertainty about what to do in order to uncover uncertain information, and we assimilate new situations to them. At least that if what we do if the capacity here is a typical intellectual virtue of a bounded agent, and if I am right about how such virtues operate.

I have gestured at such an analysis in Morton (2004) and develop it at length in Morton (2012b). The essential elements are a database of past situations with the satisfactoriness of their solutions, and a similarity measure that can relate novel situations to stored ones. These will vary from one agent to another depending on their experience and how well they have assimilated it. Then given a new situation - a question needing information, a range of actions that might prompt it, background information - an agent can find solutions that are in a very general way like ones that have worked in the past - pushing out the boat for a grand and risky exploration, or a careful and tentative probe that might reveal whether the topic is fertile or recalcitrant. This may involve ingenuity and creative thinking, to see surprising similarities, or it may rely on rote learning of experimental paradigms in one's area of science. In either case it is likely to be very subject-specific: someone who makes the right probes when interviewing candidates may be disastrous in allocating money for DNA sequencing equipment.

Virtues understood in this way will be in a general way externalist, in that a capacity that is a virtue in one situation may not be a virtue in another, and the agent may not be able to tell one from the other. They could also be called reliabilist virtues, in that they are parts of reliable ways of getting true belief, and in fact reliable in ways that lead to knowledge, and more generally to accomplishment. For the purposes of this paper, these taxonomic issues are not important. What is important is the existence of the profitable species of thinking I have been describing, the necessity of using it throughout our activities, and the facts that it can be carried out more or less well.

The similarity of this virtue to others and my praise of my general analysis do not clinch the case. But look at the features of experimental choice that we can explain in this way. They can be gathered under three heads

- We make reasonable choices in situations whose complexity and incomparability prevent our thinking them out from first principles.
- We can train one another to make acceptable choices, even though in learning one acquires little information that one did not already have.
- We articulate many of the considerations we find relevant in threshold terms. Is the prospect of finding relevant enough to justify the expense? Does the design rule out enough alternatives? Have we collected enough evidence that we can now devote our resources to other tasks?

These have immediate explanations on the picture I am suggesting. But they become miraculous if we do not see them in terms of a specific acquired virtue. Acquiring the virtue is a central and indispensable part of any scientist's training.

And acquiring the corresponding virtues in social interaction and in learning from others is essential to success in those areas. In social life one learns to probe, show emotions and provoke reactions, in ways that will lead others to reveal their emotions, intentions, and opinions. You frown when you want the other to explain more fully. In one's education one learns who to turn to for explanations, and how to do it effectively. You search out someone who understands why some customers hate some websites. Some people do some forms of each of these better than others, and everyone gets at least a little better at it with practice.

5 **theory to the rescue?**

There is an objection that will occur to anyone with experience of planning and carrying out experiments in contemporary science. We have a complex and developed *theory* of experimental design. It mixes common sense and sophisticated unintuitive statistics, and is treated with respect from field botany to theoretical physics. But if we can choose our experiments on the basis of a theory, virtues are not needed. Knowledge, intelligence, and careful rule-following will be enough.

The theory of experimental design, from Fisher to the present (Fisher 1935, Cochrane and Cox 1950) provides quantitative measures of the tests that experiments provide of hypotheses and estimates. (Careful phrasing is needed here, as the relation of these measures to familiar notions of evidence, probability, or support is controversial.) Armed with these measures, we can say in advance how stringent a test a given experimental design will give a hypothesis, and how one design compares to alternative designs. So we can give advice about how suitable a proposed experiment is for given purposes: about which factors should be randomized, how many trials should be run and how large samples should be, and so on. The theory explains how factorial and sequential designs can be efficient in unexpected ways, though comparison of their results with those of more traditional designs is subtle and unobvious. All this advice is based not on the operation of any carefully acquired virtue, but the direct application of an explicit theory.

Some of the information provided by the theory of experimental design is not at all obvious. I am thinking in particular of conclusions about number of trials and size of sample needed to get confidence limits within given bounds, which do not fall into an easily intuited pattern (Cochrane and Cox pp. 23-29). If without aid of theory an experimentalist had an intuitive grasp of what was needed here, she would indeed possess a rare and delicate virtue. And indeed the complexity and subtlety of the theory is an argument that in non-scientific contexts the art is based on a delicate projection from similar cases rather than on an application of explicit principles. I do not think it is impossible that someone designing scientific experiments might have an intuitive grasp of the force of sample sizes and numbers of trials, but it must be rare. So let us disregard the possibility and assume that when we want such things in precise form they can only be known by derivation from an unobvious theory.

Virtues are still required. The most obvious one addresses the question "given that we can do experiments E or F providing tests of stringency s and t of theories A and B, with which should we proceed?" Suppose you can try an experiment which, if it succeeds, will provide a weak test of an ambitious version of your theory, or a different experiment which if it succeeds will provide a stronger test of a special case of the theory. Neither may succeed, and no theory of experimental design will provide you with probabilities for the underlying facts. The prudent course might be to test the special case first, but success there might not be dramatic enough to get you funding to do the more ambitious test. So how much more valuable, scientifically, is the more ambitious claim? And how reliable is your hunch that your experiment will confirm it? For that matter how much more believable will the claim be if a test of this kind is passed? None of these questions are answered by any theory of experimental design, and all of them require a very delicate mixture of sensitivities.

Good use of the theory can be of immense help to experimentalists. It is not an easy theory, and most practicing experimentalists have more sense than to trust their grasp of it. Instead, they consult statisticians, or colleagues in their disciplines who specialise in the topic. So part of the experiment-shopping virtue is substituted for by a different complex of virtues, which also have an informal experimental component, the virtues of knowing when one's own ignorance suggests taking advice, who to consult, and how to adapt what they say to your actual situation. Statisticians and experimentalists tend to speak rather different languages and have rather different concerns, so that the adaptation is often not trivial. No wonder that many departments of psychology or zoology include a colleague whose special role is to bridge the gap between theory and practical sense.

6 **continuity**

I have been arguing that there are special skills of acting to know, of interacting with the environment in such a way that one acquires the knowledge one needs, and that these serve a distinctive purpose which justifies us in gathering them together as virtues of experimentation. In particular, skills of assessing whether an experiment is worth performing are worthy of attention, and separating out as a distinctive virtue. The skills involved in achieving these ends in social life, everyday practical activity, and scientific disciplines are varied though overlapping. But the virtues they constitute when performed successfully can be drawn together as an identifiable contribution to our capacity to know and accomplish. They are part of a neglected but vital area of human capacity, the ability to do the right thing in order to know an interesting thing.

BIBLIOGRAPHY

Achinstein, Peter (2001) *The book of evidence*. Oxford University Press.

Chang, Ruth (1996) *Incommensurability, Incomparability, and Practical Reason.* Harvard University Press

Cochran, William, and Gertrude Cox (1950) Experimental Designs, second edition. Wiley.

Fairweather, Abrol (2012) 'Duhem-Quine virtue epistemology.' *Synthese* **184**, pp. 1-20

Fantl, Jeremy and Matthew McGrath (2010) ‘Pragmatic Encroachment,’ in D. Pritchard and S Bernecker, eds. *The Routledge Companion to Epistemology*, Routledge, pp. 558-568

Fisher, Ronald A (1935) *The Design of Experiments.* Oliver and Boyd

Gigerenzer, Gerd, Zeno Swijtink, Lorraine Daston (1990) *The empire of chance: how probability changed science and everyday life*. Cambridge University Press

Hacking, Ian (1983) *Representing and Intervening.* Cambridge University Press

Harman, Gilbert (1973) *Thought.* Princeton University Press

Kvanvig, Jonathan (2003) *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press

Morton, Adam (1990) *Disasters and Dilemmas: strategies for real-life decision-making.* Blackwell

Morton, Adam (2004) 'Epistemic virtues, metavirtues, and computational complexity,' *Nous*,**38**, 3, pp. 481-502

Morton, Adam (2012a) Accomplishing Accomplishment' *Acta Analytica.* **27**, 1, pp.1-8

Morton, Adam (2012b) *Bounded Thinking: Intellectual Virtues for Limited Agents.* Oxford University Press

Radder, Hans (1996) *In and About the World: Philosophical Studies of Science and Technology.* State University of New York Press

Radder, Hans (ed) (2003) *The Philosophy of Scientific Experimentation.* University of Pittsburgh Press

Sosa, Ernest (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I.* Oxford University Press

Zagzebski,Linda (1996) *Virtues of the mind*. Cambridge University Press