Efficient psychology (a study of the artificial)

1. SUMMARY AND OVERVIEW ... 1

2. ARTIFICIAL AND NATURAL ... 2

2.1 THE ARTIFICIAL MIND ... 2
2.2 BEYOND HERBERT SIMON ... 5
2.3 THE FEATURES OF APPROPRIATE RESEARCH ... 5
2.4 WASTED RESEARCH EFFORT ... 6
2.4.1 To understand the mind ... 6
2.4.2 To improve the mind ... 7

3. USEFUL RESEARCH STRATEGIES ... 8

3.1 SIMULATION ... 8
3.2 THE ENGINEERING APPROACH ... 9
3.3 LEARNING FROM TOP PERFORMERS ... 10
3.4 QUICKLY IDENTIFYING METHOD USERS ... 11

4. RESEARCH DECISIONS ... 12

4.1 THEORIZING AND DESIGNING ... 12
4.1.1 Ideal performance ... 12
4.1.2 Individual differences ... 13
4.1.3 Causal models ... 14
4.1.4 Neural activity ... 14
4.2 EXPERIMENTAL TASKS ... 15
4.3 DATA CAPTURE ... 16
4.4 INTERVENTIONS ... 17
4.4.1 Changing the task environment ... 17
4.4.2 Changing skills ... 17
4.5 USE OF INTROSPECTION ... 19
4.6 SUBJECTS ... 19
4.6.1 Expert subjects ... 19
4.6.2 High-performing subjects ... 20
4.6.3 Self-experimentation ... 20
4.7 DATA SUMMARY AND ANALYSIS ... 20
4.8 APPLICATION OF FINDINGS ... 23

5. RESEARCH DESIGN CHECKLIST ... 23

6. SCOPE FOR IMPROVEMENT ... 24

6.1 AN INDIVIDUAL CHOICE ... 24
6.2 DETAILED EXAMPLES ... 24
6.2.1 Immediate and delayed rewards ... 24
6.2.2 Anchoring ... 26
6.2.3 In- and out-group commenting ... 27
6.3 SELF-TEST ... 28

7. CONCLUSIONS ... 29

8. REFERENCES ... 29

9. APPENDIX A: SELF-TEST SUGGESTED ANSWERS .. 32

1. Summary and overview

This article explains how psychological research, especially by academics, can become more productive and valuable in society by responding to the largely artificial nature of the mind. There is an opportunity for society as a whole, but also for individual researchers and research groups who want their work to be more highly valued, influential, and supported.

The article begins by explaining what it means to say that the mind is 'artificial', contrasting 'artificial' with 'natural'. The distinction is roughly the same as that drawn by Herbert Simon.

Some general suggestions are made about research methods appropriate for studying natural and artificial processes in psychology, arguing that they are profoundly different. These proposals may be persuasive for you on their own but are intended as an overview of the implications of the following sections.

Those following sections describe research strategies for artificial psychology and then discuss research decisions in detail. This goes beyond the proposals of Herbert Simon during the 20th century.

Many examples are given to illustrate techniques for researching artificial psychology. It is much easier to appreciate the scope for change and improvement if you are familiar with a wide range of more appropriate research tactics.

The practical implications for research design are summarised through a checklist that can be applied to ideas for psychological studies.

The final section explores the scope for improvement, arguing that this can be considered a matter of individual choice for researchers.

Almost all the examples given show appropriate, efficient, useful research and this may create the impression that such research is typical of psychology. Sadly, it is not. Finding these examples took a lot of effort and involved discarding many, many more examples of inappropriate, often useless research.

2. Artificial and natural

2.1 The artificial mind

The observation that our thinking is artificial (in the sense of 'man-made' rather than 'fake') can be understood through some simple examples:

> E.g. Imagine that a psychologist sets out to understand how humans solve problems and decides to focus on how we solve quadratic equations. Obviously, most people solve quadratic equations the way they have been taught. People can explain their methods. They may know of alternative methods and have chosen their favourites. Different people use different methods, though there are only a few commonly used approaches. People can change their methods by learning a new one and adopting it if they prefer it. In short, the way we solve quadratic equations is something that has been designed and adopted by each person and they can change their methods if they want to.

With this research example the artificiality of the mental processes being studied is obvious to everyone. It contrasts with research problems looking at natural systems.

> E.g. Imagine that a chemist is studying combustion of hydrogen gas in air. She does this by setting light to various samples of hydrogen gas in air under different conditions. Although the process responds to the conditions, it does so repeatably. All samples of pure hydrogen respond in the same way. They do not learn, do not have different methods, and do not try something different from time to time to see if it works better. Mathematical descriptions of the process fit it closely and, when used to make predictions in somewhat different circumstances, they predict behaviour quite well.

This is an equally clear-cut example of studying a natural system. Other natural systems include reactions between acids and metals, the currents of the oceans, and the eye of a spider.

Inanimate systems are not all natural. Humans have designed thousands of inanimate systems from chairs to space rockets. Research can be done into these too, as artificial systems.

> E.g. Imagine that an engineer is trying to design a water turbine. Water flows into one end of a large tube, turns the turbine, and flows out of the other end. The turbine itself has a variety of fins of different shapes, sizes, and positions around a central shaft. The engineer wants to design a better turbine so she systematically varies elements of its design (e.g. the positions and shapes of fins) and records data on the performance of each variation. She summarises these with charts and mathematical models and then looks for ways to improve

performance. Is it best to run at high revolutions per second or low? Are many small fins better than a few big ones? Should the fins be thick or thin?

This kind of study is sometimes called 'engineering research' or 'applied research'.

Much human thinking is not so blatantly artificial as solving quadratic equations. A person may be unable to explain how they think something through, may have forgotten the choices they made while learning, or were never aware of them.

Nevertheless, if it is now possible for a person to learn a new method then the thinking is artificial. This is my proposal for the best test of whether a piece of thinking is artificial. This is what matters for practical purposes.

Our mental processes typically reflect both the subject matter and the characteristics of our minds. The contribution of each differs between tasks. When there is a significant contribution from the characteristics of our minds then psychology is relevant.

E.g. The way people solve quadratic equations is not simply the result of the logic of quadratic equations. The characteristics of the mind are relevant too, which makes this a legitimate topic for psychology.

- An electronic computer might well solve them using iterative numerical methods but humans do not. Such methods require calculations that computers can do quickly but we do very, very slowly. Also, our symbolic ways allow us to write exact answers.
- In equations where the coefficient of the squared term is not 1 the alternative methods include some that are mentally hard work and others that involve more writing but are easier to think through.
- Also, the amount of algebra we do in between lines written down is carefully controlled by skilled solvers to give a good combination of progress with low mental strain and risk of error.

E.g. How do humans design bridges? Again, a lot of knowledge about bridges, land, materials, forces, and so on is relevant. However, the processes by which people design bridges are not just a reflection of this external reality. Design engineers have different approaches to this kind of problem reflecting mental characteristics too. There is also innovation in these approaches, with many books, articles, and videos produced by people promoting different approaches to engineering design.

If you want to develop better mental processes for accomplishing tasks then it should help to know about the strengths and weaknesses of the underlying machinery of the mind.

Conversely, if you are interested in understanding the underlying limitations of the mind then it should help to know what mental processes a person is using and try changing those to see where limits impinge.

The opportunities for improving thinking increase when external aids are used. Humans have been doing this for a very long time. Examples include marks to help with counting, reading and writing generally, the abacus, mechanical calculators, electronic calculators, and computers. When external aids are used, the way a task is approached can be designed to get around some of the limitations of unaided human thinking.

E.g. One way to combine several factors to reach an overall assessment is to reduce each factor to a number, multiply each by a weight, and add the results. This is easiest with an electronic calculator or computer. The number for each factor might be the result of gut feeling or something more factual. This calculation method is called a linear model and will often combine information better than human judgement, even if the weights are poorly chosen and the linear model type is not ideal (Dawes, 1979).

It is also crucial that humans have been working on their methods of thinking for a very long time and can share their discoveries through writing, recordings, and direct teaching. This expertise is a valuable part of our culture.

Languages and notations are important parts of that culture, in part because they can profoundly affect how we think and the performance we can achieve.

E.g. Algebra today is easier than it was in the time when mathematicians wrote their equations as sentences using words. Our compact notation makes more efficient thinking possible. Despite the advances, modern mathematical notation still contains many ambiguities and even errors. When computer scientists tried to use mathematics to specify computer systems in the 1980s, they had to develop a more rigorous style of mathematical notation to do it. The specification language Z is an excellent example of this and a powerful tool for thinking. For more detail on Z see Spivey (1992).

E.g. Computer programming languages have usually been developed to be easy to use, expressive, suited to likely uses, and easy to compile or interpret efficiently. New languages continue to be developed and offered as people search for advantages. There are also some languages developed for amusement that aim to be as difficult and frustrating to use as possible.

E.g. So-called 'natural' languages (e.g. Latin, English, Spanish) are really the result of many choices by people down the centuries. Some languages are harder to learn and to read than others. For example, English is harder because of its huge vocabulary and the loose relationship between written symbols and sounds. This is the result of centuries of invasions and migrations, making 'English' an amalgam of earlier languages.

Some deliberate attempts to change 'natural' languages have had profound effects. For example, Samuel Johnson's 'Dictionary of the English Language', published in 1755, was the first full featured dictionary and, along with subsequent dictionaries designed along the same lines, helped to standardize spelling of English. Dictionary editors continue to make choices about which words to recognize and which spellings to promote. People have invented words for new technologies (e.g. television, radio, laser). Political movements have tried to make some words unsayable, replacing them with alternatives.

In Russia, just after the revolution in 1917, the government imposed new, simplified writing standards for the Russian language. Similarly, the government of China has been trying to promote a defined standard Chinese language across the country for decades.

Examples of types of human thinking that are artificial in the sense described above and within the scope of psychology include types of problem solving, design, writing, reading, memorizing, communicating, negotiating, sense making, decision making, planning, diagnosing, exploring, deducing, learning and performing perceptual motor skills, and developing interpersonal relationships. These contribute to applications such as:

- teaching and learning in schools, colleges, and other contexts;
- helping people overcome depression, addiction, procrastination, compulsions, trauma, and other mental problems not entirely caused by organic disease;
- improving performance in business; and
- improving influence and arriving at better decisions for individuals, families, groups, organizations, and whole societies.

Topics that are natural in psychology include such things as:

- The biology of the eye and individual brain cells.
- Mental conditions with a biological cause, such as a genetic problem, injury, or infection.
- Learning in babies (though the way adults try to teach them is artificial).

2.2 Beyond Herbert Simon

The great psychologist and economist, Herbert Simon, explained the idea of artificial systems and the need to study them in a different way. He wrote about this in his book 'The sciences of the artificial' (Simon, 3rd edition, 1996) and it was the basis of his proposal for 'cognitive science' (Simon, 1980).

Simon pointed out how the artificial aspects of thought made it more difficult to study the mind scientifically[1]. Uncovering the underlying limitations and invariant patterns was harder, he observed, when so much could change as people adapted to their environment.

Simon's suggested approach to research is briefly described in a later section of this article. His proposals inspired work but did not revolutionise psychology. This was perhaps because his suggested research method was only one possibility and not as productive as he had imagined. The intensive studies of small numbers of subjects that Simon often preferred probably seemed laborious to many researchers, while the benefits were perhaps unclear to them.

2.3 The features of appropriate research

Natural and artificial systems require profoundly different research methods.

To give a sense of this and as a preview of what is to come, here are some characteristics of the methods that are appropriate in each case. (More rationale for these characteristics is provided in later sections.)

For a natural system it is usually appropriate to:

- Focus on describing, explaining, and predicting behaviour.
- Make no attempt to influence the way subjects in studies think, other than by changing conditions.
- Expect people to think the same, typically. (You might go as far as treating the system as adaptive and considering individual differences as

[1] Simon pointed out that economies were also artificial. The way they operate depends on human invention and on the way individual actors make their decisions. We might also observe that the earth's climate is increasingly artificial and that humans have been slow to realize it.

you might for studying a species of plant or fungus, but you would not expect to see strategy switching, insights, or individuals using different methods.)

- Make no attempt to identify differences in thinking methods between people and assume that effects seen are common to everyone.
- Average the behaviour of lots of people to find a common pattern.
- Expect the basic features of behaviour to be the same in any task, whether useful in practice or not.
- Use any experimental task that seems convenient, even if it is not a task faced outside the laboratory.
- Test behaviour once or for a short time, not expecting qualitative changes over time.
- Pay no attention to introspection.

In short, stand back, expect homogeneity in mental processes (between people, tasks, and over time), and do not use introspection.

However, for a mental system that is artificial it is appropriate to:

- Focus more on improving performance by improving mental processes rather than just describing, explaining, and predicting behaviour.
- Often try to influence the way people think by going further than just changing conditions.
- Expect variations in mental processes between people and over time, unless they are controlled.
- Try to characterize performance (including interesting effects) while using specific thinking methods (rather than assuming that performance is the same for everyone regardless of the method they are using).
- Avoid averaging the behaviour of lots of people.
- Expect performance to be highly task specific.
- Focus on tasks that people want to do better.
- Study thinking and behaviour over time and in detail, expecting changes.
- Use introspection, though cautiously.

In short, try for improvement in thinking, expect variety in thinking processes (between people, between tasks, and over time), and be willing to use introspection.

2.4 Wasted research effort

Research is likely to be less efficient and sometimes useless when inappropriate methods are used. My observation is that a frequent mistake in psychology is to use natural science methods for artificial systems. This way, progress in understanding the mind and improving it is usually slow.

2.4.1 To understand the mind

Time can be wasted trying to uncover the invariant properties of the mind because they are obscured by the shifting, unknown, artificial elements. Our artificial thinking processes are such powerful determinants of our behaviour that, unless we know what those processes are by careful study or training, we will struggle to see what is invariant.

The artificial elements can shift for several reasons:

- Different subjects may have different preferred methods for doing a task, or might think of different methods when faced with a new task.
- They may experiment with different methods during the study.
- They may act in accordance with their own psychological theory. For

example, when a subject answers a battery of questions about themselves and their behaviour, they may give answers that are consistent with their theories about themselves or human nature rather than answering accurately.

- They may act to please the experimenter by providing data that fit what the experimenter seems to expect or want. Less often, they may do the opposite to displease the experimenter. Experimenter effects like this are well known in psychology.

Because of the obscuring effect of shifting, artificial elements, 'applied' research that tries to control the mental processes of subjects can be more efficient than 'pure' research at uncovering the empirical facts that 'pure' research aims for.

Time can also be wasted trying to establish, experimentally or by regression, which factors drive behaviour in a task or how information is being processed. If it is possible to ask people how they learned to do it and what they are trying to do then we can save a lot of time by doing so. Studies to see which factors really drive behaviour then just confirm what is already reasonably expected.

Acting on the idea that people are a natural system and more similar to each other than they really are can lead to more wasted effort. Average behaviour may not characterize any individual, even approximately. Important clusters of similar behaviour can be missed.

> E.g. Imagine that a researcher wants to understand how consumers in a market evaluate a particular type of product. Which factors are most important? If all customers are considered to think the same then they should all prefer the same product. However, once we understand that consumers are not the same but they often fall into clusters the situation is clearer. One example of market segmentation using cluster analysis (out of many) is Morton, Anable, and Nelson (2017).

2.4.2 To improve the mind

Perhaps most importantly, one consequence of treating the artificial mind as a purely natural system is that many opportunities to design improvements will be missed.

> E.g. In the early 1960s, Milgram and colleagues carried out a series of studies to explore how far authority could push people to behave badly, even to the point of murder. Participants thought they were giving electric shocks to another participant to motivate them to learn, but the severity of the shocks increased progressively to the point where they appeared to be causing great pain and risking death. The studies were harrowing for participants[2] although no real pain was inflicted and there was never any risk of death.
>
> In total, over many variations of the study, 720 people went through this horrible experience. Much was learned about conditions that increased or decreased the proportion of people who turned the voltage up to maximum instead of refusing to carry on (Milgram, 1974).
>
> However, none of the studies as far as I can determine tested ways to prepare participants that might make them more resistant to authority in this situation. What about teaching people to check when someone seems to be

[2] Apparently more harrowing than Milgram let on, according to Perry (2013a and b). In particular, most people left the laboratory thinking they had really given electric shocks.

in pain? What about inoculation against the arguments typically used by officials to motivate bad behaviour? What about teaching the law on causing pain and risk of death to people, even when it seems to be sanctioned by an official? What about teaching how to blow the whistle when you find something unlawful happening? We will probably never know if any of these works because these ideas were not tested and it is almost impossible to get ethical approval for this type of study in a modern university.

Further waste can arise from studying performance in useless tasks. Although there is a slight chance of establishing some characteristics of brain performance that are also true in useful tasks, it will almost always be more useful to study people doing useful tasks. People are likely to be more willing to participate as subjects. Even if generalization proves unreliable, at least the research provides information about that particular task, which is useful.

E.g. Hermann Ebbinghaus (1880 and 1885) did valuable and famous work by conducting a long series of tedious experiments on himself. He was dedicated enough to memorise many lists of nonsense syllables in order and record his learning and forgetting performance in detail. However, most learning that people do is not learning lists of nonsense syllables in order. Material like this is so hard to learn that we write it down. We are much more likely to be learning material that is meaningful, connected, patterned, and sensible. With this material we can use mental processes to create memories that are more durable and reliable, and we can respond to different uses of the learned information. Consequently, although Ebbinghaus's data show what happens when nonsense syllables must be learned, they give us no reliable idea of how much effort is needed to learn more typical material, how long it lasts before being forgotten, or what methods tend to improve memory.

Time can also be wasted studying the behaviour of incompetent people when our objective is to help almost everyone progress far beyond that level. Instead, we should study how to help people improve and understand the performance of more refined methods of thinking.

3. Useful research strategies

Tackling psychology as a study of the artificial suggests research strategies likely to produce useful insights and findings.

The 'pure' science strategy of trying to understand a system so completely that it then becomes possible to do useful things is unlikely to work well with artificial systems. Consequently, most of the strategies suggested below are more like 'applied' science strategies but they can still produce useful insights and generalizations.

3.1 Simulation

Herbert Simon's proposal was to try to identify and quantify the invariant parameters (usually limits) of human cognition, characterize the smallest units of thought, and create simulations of larger mental processes.

The idea was to use the simulations to understand and predict human thought but trying to improve it would also be possible via this route.

Simon's approach was attempted by many researchers and their simulations eventually became sophisticated and impressive.

E.g. Richman, Staszewski, and Simon (1995) reported the results of a simulation of digit span memory using EPAM IV, the fourth version of a mind simulator first developed in 1959.

However, the work had little impact and interest seems to have fizzled out. Modern work in 'artificial intelligence' is more concerned with developing computer software that does useful tasks effectively.

Knowing about the fundamental parameters and most basic processes of human information processing was not as useful as hoped. There is a large gap between this knowledge and using it to predict or design human thinking. The task was rather like predicting the behaviour of an electronic computer from knowledge of its most basic functions. Application software stands between those basic functions and the overall information processing behaviour. It is important and complex. Appropriate research strategies are needed to tackle it that do more to adapt to the challenge of working with an artificial system like the mind.

Nevertheless, simulation can be useful. Even without much understanding of mental limitations, simulation can test methods of thinking to establish their likely efficiency in practice.

E.g. Durbach, et al (2020) tested heuristic methods for choosing portfolios of projects using simulations and looked at human behaviour to see if people already used those heuristics to some extent. The simulation was able to compare choices made using the easy-to-apply heuristic with ideal choices and found that, with some heuristics, there was little difference.

Although Durbach et al did not try to teach their simulated methods to humans, Pande, Papamichail, and Kawalek (2021) had subjects follow simple algorithms that had previously been explored to see how well subjects would comply. The only efforts made to get subjects to follow the instructions were to have them read the instructions and attempt a small number of practice problems. People who did very badly in the practice problems were eliminated from the study and the overall failure rate of those remaining was 25%. This seems a high rate of failure but the level of training was minimal. Further work could raise this and measure the decision performance achieved in realistic problems.

3.2 The engineering approach

This involves studying the performance characteristics of promising designs to understand what drives performance and improve the designs. The approach was illustrated by the example given earlier of studying a water turbine[3].

E.g. An engineering approach to studying the way children learn to spell words in the English language would involve designing a particular method for the children to use, training them to use it, monitoring to check use, and measuring the performance achieved when children use variations of that method. (Many young people in their early teens and even younger are capable of participating in this way.)

An example of a study with many of these elements is Arnbak and Elbro (2000), who helped struggling readers by making them more aware of how words are built from smaller units ('morphological awareness training'). A review of many studies like this is

[3] In 'The sciences of the artificial' Simon gave the example of developing time-sharing computers. People just built them to see how they would perform and improved the designs in stages from this experience.

provided by Bowers, Kirby, and Deacon (2010).

E.g. Teasdale et al (2002) gave Mindfulness Based Cognitive Therapy to patients who had recovered from depression. This involved them thinking about their thoughts in a particular way so that worries seemed less likely to be true concerns and less important overall. This reduced the proportion of patients who relapsed into depression. Farb et al (2018) used a similar approach to compare the impact of Mindfulness Based Cognitive Therapy with Cognitive Therapy and found them equally helpful in reducing relapse into depression. Their effectiveness seemed to be related to the effect on the patient's thinking about their own thinking.

Performance can be characterized using mathematical models, tables of data, and narratives that describe the effort needed to implement and use methods of thinking, and the results they achieve, including patterns of errors.

These characterizations can be used in various ways:

- Look at what factors drive performance and use that knowledge to design better methods.
- Use knowledge of the performance at small tasks to design larger methods and predict their performance.
- Use the quantifications for planning (e.g. to predict how much learning time will be needed for particular students to reach a particular level of proficiency).

3.3 Learning from top performers

This strategy is to study many people doing the same task and identify the methods used by the people who perform best. Alternatively, the methods of acknowledged top performers can be compared with those of more typical people. A follow up is to teach those methods to other people to see if they improve performance.

E.g. Thorndyke and Stasz (1980) used this strategy to study skills for learning maps. Their 8 subjects had to learn two maps in detail for recall tests that involved drawing the maps from memory and answering questions. Three of the subjects were highly experienced map users while five were students with no special map experience. The ability to learn the maps was unrelated to experience but strongly related to the number of uses of effective strategies for learning. Teaching poor map learners the effective strategies in a second experiment improved their performance.

Research done in this way can be interesting to many people, leading to big-selling books and commercially successful consulting businesses.

E.g. Rackham has studied expert performers to understand what works in negotiations (Rackham and Carlisle, 1978a, b) and in sales meetings (e.g. Rackham, 1988).

E.g. Kepner and Tregoe (1965) examined the thinking processes of successful managers to distil processes they thought worth teaching to others.

E.g. Data from the Good Judgement Project's geopolitical forecasting competitions have been used to study the thinking patterns of the most successful forecasters. An example is Karvetski et al (2021), who examined the rationales given by good forecasters for clues to their thinking patterns.

If discoveries from this kind of research are published, they can become widely known and affect the processes used by many people. This is usually helpful but there can be negative effects, especially if knowledge is misused.

E.g. In the late 20th century many studies looked at human social behaviour as a skill. They identified what was normal behaviour, especially non-verbal behaviour, and what was abnormal. This was quickly picked up by authors of popular books and incorporated into training for any group of people with an interest in getting what they wanted from other people (e.g. sales people, interviewers, people advising job hunters, therapists). Social skills training was also offered to people with problems such as social anxiety and shyness.

Soon behaviours identified by this research were common knowledge. If a person used one of these behaviours then they might be deliberate and other people would notice and know that. Postural mirroring, repeating a person's name, saying 'you' a lot, smiling and nodding, eye gaze behaviours, open postures, and particular types of handshake became recognizable motifs of the 'people person'. What had once been an unnoticed cultural tendency had become, with some people, a deliberate and annoying battleground.

3.4 Quickly identifying method users

Sometimes it is helpful to understand the beliefs and skills of individuals so that they can be supported more effectively (e.g. when teaching, when designing software tools, when helping a customer make a purchase decision). A research strategy based on this is to create and test ways to quickly identify an individual's beliefs and skills.

E.g. Some people trying to learn a particular skill may get stuck because they are using a method that does not work. Quickly identifying the method they are using, recognizing the problem, and coaching them to change might allow them to make progress.

When I was a child, I failed to learn to speak French (as a second language) because I tried to translate by mapping French and English words. This method does not work. It is necessary to map sentences in one language to meaning, and then meaning to sentences in the other language.

When I was an even younger child, I struggled at first to learn to read. At that time my school was using 'flashcards' to show us words without explaining that individual letters were associated with sounds. I didn't get it. Fortunately, an older teacher spotted my struggle, took me aside, and explained the link between letters and sounds. I quickly caught up and got ahead but what might have happened without that teacher?

E.g. Reisen, Hoffrage, and Mast (2008) reported a test of the performance of a software tool to identify the decision strategies used by consumers. The tool was called InterActive Process Tracing (IAPT) and combined three process tracing tools: Active Information Search, Mouseman, and retrospective verbal protocol. They studied people choosing between alternative mobile phones and tested the extent of agreement between subject choices and choices made by the system using the identified strategy. For an understanding of some of the consumer decision strategies that have

> been investigated over the years, see Richarme (2005).

4. Research decisions

The insight that the mind is largely artificial suggests some more detailed points about research decisions.

4.1 Theorizing and designing

The usual overall objective for much research in psychology should be to design and test better methods, often methods for thinking. It should not be to describe, explain, and predict behaviour as if it is a natural system.

4.1.1 Ideal performance

Whether studying the mind as a natural (but evolved) system or as an artificial system, it is often helpful to consider what ideal, rational performance would look like. However, thinking about ideal performance is more important for researchers who recognize the mind as artificial.

Humans often behave in adaptive ways, so thinking about ideal reasoning can lead to good predictions. However, once differences between ideal and actual performance have been found, researchers seeing the mind as natural typically focus on models that try to account for the defects.

In contrast, with the mind recognized as artificial, ideal thinking is always helpful to consider. It helps us recognize opportunities for improvement and devise improved ways to think.

> E.g. One theory of human thinking is that we are 'risk averse'. This means that (1) for each potential course of action under consideration in a decision, our minds compute a quantity we can call 'risk' and (2) we are less attracted to courses of action with a higher risk quantity associated. This theory explains why we are not attracted to a bet with a 50:50 chance of winning or losing the same amount of money. However, it is far from the only explanation.
>
> For more than a century an alternative theory has been that a rational person will tend to link value to money in a non-linear way such that each additional unit of money adds a slightly decreasing value. For example, losing £10 is more important to a poor person than to a billionaire. This makes rational sense but does not explain all the nuances of human judgement and decision-making seen in experiments. So, the 'risk aversion theory' has persisted (among others).
>
> However, diminishing utility of money is not the only rational reason why we might behave in a way that seems 'risk averse'. Another is that, if we have more information and so less uncertainty, it is often possible to think of and follow a more advantageous course of action. It is not that we are averse to risk; we prefer knowing more when we can use that knowledge.
>
> Also, one effect of unexpected events is that they can cause us to adjust our lifestyles. A windfall might allow us to redecorate the kitchen or join a club. Losing a job might force us to sell the car or take the kids out of private schooling. All lifestyle changes involve a certain amount of effort and stress but, typically, forced lifestyle change downwards is worse than upwards because it is under greater time pressure. It is not that we have a mindless preference for the status quo; we have a sensible appreciation of the costs of change, even when that change is for the better.

If we have reserves (e.g. money saved) then a setback such as losing a job can often be ridden out with no need to change lifestyle significantly. So, having reserves should remove or reduce the anticipated impact of change.

Preferring more knowledge when it can be used and taking the cost of change into account, bearing in mind reserves, are effects that would be seen in rational behaviour. When these are tested in experiments with humans we can expect that, if the issues are sufficiently obvious in the task, then (1) most subjects will respond rationally, but (2) many subjects will respond in an imperfectly rational way, with some appearing completely confused. (See Delquié, 2008 for an example.) Beyond that, teaching subjects the logic involved and helping them develop their skills would almost certainly lead to more rational behaviour.

4.1.2 Individual differences

Knowing human thinking is artificial should also increase our attention to individual differences and change how we detect and explain them.

We must be careful not to assume that differences arise from fixed characteristics of the person when in fact they arise from characteristics that change, especially beliefs and skills, which usually reflect experiences and efforts at self-improvement. This can be an easy mistake to make because, even in simple behaviours, large individual differences in ability can arise from seemingly tiny differences in skill.

E.g. Boot, Becic, and Kramer (2009) challenged subjects with four different visual scanning tasks, two of which were best done without eye movements and two of which were best done with eye movements. They noted the performance of their subjects and whether they moved their eyes. Many people used the wrong strategy, seeming to use a favourite strategy by default for all tasks. However, when given feedback on their performance (but not eye movements), subjects often switched to using the better method.

If we make the mistake of assuming differences arise from fixed characteristics of the person when they do not, we may fail to think of beneficial changes that could be made (e.g. through getting better evidence, through changing skills).

We should also be careful not to assume that differences between individuals will manifest themselves in every situation when in fact they are situation specific.

A lot of theorizing and research based on trait theories of personality has made both these mistakes.

E.g. Boyce, Wood, and Powdthavee (2013) discussed past ideas about personality being fixed and reported a study that powerfully challenges them. They used a longitudinal data set covering 8,625 Australians. For 2005 and 2009 the data recorded each person's 'personality' on the 'Big Five' personality traits, some facts about their economic situation (e.g. income), and their life satisfaction. All these were found by asking lots of questions, then computing scores from the answers.

Both personality scores and economic facts statistically predicted life satisfaction ratings at the same date, with personality being slightly more important. This was consistent with previous studies.

The new discovery was that personality scores changed more than

> economic facts and changes in personality predicted changes in life satisfaction better than did changes in economic facts. This implies that the changes in personality scores were not just measurement errors. If they had been then they would not have predicted life satisfaction.

4.1.3 Causal models

Recognizing the mind as largely artificial suggests that psychologists should focus on testing the effects of different types of variable and might get cleaner data by controlling mental processes.

One inspiration for approaching the mind as a natural system may have been the way physics has succeeded. In physics, much has been achieved by defining variables operationally and using experiments to see and quantify the causal relationships between them.

However, in psychology it is rarely possible to be so precise because data from experiments are much less consistent than typical data in most branches of physics. So psychologists more often try to establish just if a causal relationship exists at all (or at least a statistical relationship) and whether one variable moves with another or in opposition to it. Factorial designs are common for this, treating different groups of people in different ways to see how they respond. The precise quantity of changes is not usually used later; the focus is on whether the change is statistically significant (i.e. a real, reliable effect or just a coincidence).

It is a slow and uncertain programme of research, often with much effort for meagre gains.

However, if we design and implement the mental process a person is using then we should already know a lot about how differences in the environment will be responded to. Research can then focus on exploring the impact of variables of more interest to the mental process designer. For example, what can overload the person using a mental process? What makes the task easier? Can the process be tweaked to relieve the strain?

Another opportunity for improvement is through cleaner data. A major reason for the 'noisiness' of data in psychological studies is that the subjects' mental processes are usually left uncontrolled. Mental processes vary between individuals and over time, even when conditions are not changed, as people try different approaches and learn, making data seem messy.

> E.g. I was a psychology undergraduate in the early 1980s and remember that when memory experiments involved remembering lists of words there was always a risk that a subject would use Visual Associative Mnemonics (VAMs) to help their performance. When they did, their results would usually be much better than normal, leaving the experimenter with the choice of either removing that subject's results from the study or letting averaging across all subjects swamp the results of the few who used VAMs.

By identifying or controlling mental processes it may be possible to remove some of that noise and fit models more precisely to data. Reliance on significance testing should reduce. Models would more often be useful for predicting results in other situations.

4.1.4 Neural activity

The thoughts we have, and the skills we build and choose to use, affect the activity of our brains and the way they grow. The artificial elements drive the development of neural tissue, as well as being driven by it. When comparing brain activity with thinking we need to be careful about the direction of causation.

While obvious brain damage is an explanation for thinking problems, the presence or lack of activity in apparently healthy brain regions is unlikely to be an explanation of thinking.

> E.g. Imagine that fMRI scanning has revealed that particular regions of the cerebral cortex are active when a person performs a particular skilled task, provided they have successfully learned the skill. However, some people have struggled and failed to learn that skill and their brains do not show the same pattern of activity when they try (and fail) to perform the task. Does the lack of brain activity explain their failure to acquire the skill? Probably not. More likely their inability to find an effective way to learn the skill is the reason why their brain has not developed the connections needed to perform the skill and so is not active when performance is attempted.
>
> Richards and Berninger (2008) found that a specific brain connectivity abnormality in dyslexic children disappeared after successful instruction in a particular phoneme task. Their study contrasts with the more usual tendency to assume that correlated brain abnormalities are the cause of struggling to learn to read.

4.2 Experimental tasks

If psychology is to generate useful insights and findings it must focus on tasks that people want to do better. Finding ways to perform useless tasks is largely a waste of effort.

The history of psychology provides many examples of useless tasks used extensively in experiments. Some examples:

- The Stroop task, which involves reading the names of colours written in different colours. (Over 700 studies using this task have been published.)
- Learning to recall lists of nonsense syllables or random words in order.
- Pressing buttons as quickly as possible in response to shapes, lights, or words appearing on a computer screen, where the required reactions are arbitrary ones specified purely for the experiment.
- Answering questions about your beliefs or behaviours by expressing a level of agreement with scores of vaguely worded statements.

For contrast, tasks that many people would like to do better include:

- Learning the number facts needed for school arithmetic.
- Learning to spell.
- Learning a new language.
- Making friends.
- Deciding what to do each day.
- Diagnosing diseases.

Useful tasks like these can yield as much insight into common features of human thinking as useless tasks. In addition, they are worth studying in themselves, lead to generalizations about methods that work across many tasks, and are more attractive to people who might be recruited as experimental subjects.

> E.g. Roediger and Karpicke (2006) reviewed work on the effect of testing recall on the strength of memories. They pointed out that very few studies of the effect had used materials from real educational settings. They wrote that this was probably why the findings were not in use in education at the time of their paper. Later in their paper they showed the results of studies of

their own using educationally relevant materials. This work has now gripped teachers and been popularized in the book 'Make it stick' by McDaniel, Brown, and Roediger (2014). Using realistic tasks has helped to shift findings from the laboratory to the real world.

4.3 Data capture

Many experimental tasks involve some thinking, some outward behaviour, and a particular behaviour that is the central focus of the study. Many studies that treat the mind as a natural system record just the central behaviour, such as a button press, the decision to approach or avoid another person, or the decision to comply with a request or not.

Since the mind is artificial and we often want to know what thinking methods a person is using, it is helpful to go much further. Every clue to thinking can be captured, including notes made, words spoken, eye movements, pauses, and errors.

E.g. Studying the way people do mathematical problems is easier because solutions are usually written. People are also conscious of much of their thinking while solving mathematical problems and can be trained to speak their thoughts so that they can be recorded and analysed.

E.g. Rosengrant, Van Heuvelen, and Etkina (2009) studied students' use of free body diagrams when solving problems about forces over a two year period. They examined the written answers of students and interviewed them about why they used the diagrams. They found that efforts to teach students to use these diagrams had been successful with many. Students used the diagrams even when the problem did not explicitly request them and got the right answer more often as a result.

E.g. Cho (2014) studied academic Internet reading skills in a way that illustrates detailed data gathering. He selected 7 high school students for their unusually strong reading and verbalization abilities and intensively trained them to make sure they understood the experimental task and how to provide as much information as possible for the study.

The task was to write a question for a discussion on a controversial topic. The requirements were quite detailed. In the first 45-minute session, the students had to select three relevant, informative websites to use as sources for developing the question. In the second 45-minute session they had to read from the selected websites and develop their question.

All screen activity was recorded by the computer. They were required to talk almost continually to show what they were thinking throughout. If they were silent for 3 seconds then they were prompted to say what they were thinking. If they took an action and were silent then they were prompted to speak. They were required to point with the mouse at any text they were reading.

The goal is to identify (1) the skills of each experimental subject and (2) changes over time. Ideally, this involves capturing details of each individual so that as much of their thinking as possible can be tracked and characterized, then persisting with this as the skill is learned over a long period.

E.g. Chase and Ericsson (1981) reported their study of a single subject who spent over 250 hours practising to develop his digit span (a test of short term memory). He changed it from the usual span of around 7 digits to 80. The experimenters recorded detail of his progress day by day and learned – often by listening to the subject – how his methods changed over time as he made discoveries.

E.g. One of my roles is as a private tutor helping young people learn mathematics for secondary school. Sitting with one person and working with them very closely I can see their every hesitation and facial twitch, and what they write, and hear them speak as they grapple with a problem I understand well. With all this detailed information coming at me I can sometimes see exactly when they are stuck and on what, which helps me help them.

4.4 Interventions

4.4.1 Changing the task environment

Following a natural science approach, we would focus on changing only the task and its cues, regarding the thinking of the experimental subjects as natural and out of reach.

This may be a sensible approach if the people concerned are uncooperative or uninterested in improving.

E.g. Hornsey at al (2021) used an advanced statistical method to uncover small clusters of highly vaccine-skeptical people. The value of this work is in finding the characteristics of these people so that they can be tackled more effectively by public health programmes (for COVID-19 in particular).

A different type of study is needed to find ways to alter the thinking of these people, perhaps by helping them improve their ability to evaluate medical information.

4.4.2 Changing skills

Recognizing that human thinking is artificial we can also try changing the way people think and act to see what difference that makes. This can be done by talking to them, giving them written instructions, teaching them, training them minutely, and monitoring them very closely during studies to check that they are thinking in the way the experiment requires. It can also be done using details of experimental tasks.

E.g. In 1975, Craik and Tulving reviewed recent research on how the way people process material affects their memory of it. In these studies people were typically asked to make some judgement about the meaning of a stimulus and were later asked to remember the stimulus (usually without having been warned to expect a memory test). The tasks people performed were, in effect, mental methods for memory formation requested by the experimenters and followed by subjects. Craik and Tulving then reported 10 new experiments further testing what they called 'depth of processing' as a theory of memory performance. Subjects were asked to respond 'yes' or 'no' by pressing buttons shortly after being asked a question and then shown a word to which the question referred. Questions typically concerned the meaning of the word or its appearance.

E.g. Bower (1970) reviewed many studies of visual associative mnemonics (VAMs), which are techniques for memorising that involve linking pairs of ideas by forming mental images that combine them. This mental process has been known for over 1,000 years and is particularly useful for remembering random words in order. In effect, the mnemonics impose meaning (albeit rather arbitrary) on material that otherwise would be random and hard to learn. Experiments either used people who had already learned to use the mnemonics or were taught to as part of the study. The improvement in memory performance when VAMs are used is often huge.

E.g. Bower and Reitman's (1972) study explored in detail the performance characteristics of visual associative mnemonics, focusing on interference between memories when the mnemonics were used in different ways. This is an example of prescribing a mental process for subjects to follow and characterizing performance resulting.

The application of some interventions is harder than just changing a person's environment, but still practical if the people implementing the methods can make persistent, intelligent efforts to improve. These people include the person whose thinking is improving, their parents, teachers, coaches, tutors, and therapists.

E.g. Mathematical techniques are perhaps the most familiar examples of thinking methods that are developed and taught to people who then use them in their lives. Typically, untutored guesswork in matters of shape and quantity is replaced by thinking that is far more precise and reliable, though often slower until automated. A massive effort is made to teach these thinking skills. In the UK they are taught on most school days to all children between the ages of 4 and 16, with many continuing their studies for years after that.

Using computer tools to support learning could also allow individuals to put hundreds of hours of effort into mastering complex skills, with programmed help on strategy.

Some studies today test very brief, superficial interventions. This is because they are looking for something that can be done by a professional person (e.g. teacher, doctor, psychologist) with only a short time in contact with the person whose thinking is to be influenced.

E.g. Gollwitzer and others have conducted many studies to explore the effects of 'implementation intentions'. Subjects are asked to form intentions to perform specific actions in specific situations. The subjects presumably think a sentence to themselves in the required format (condition followed by action), meaning to carry out the action when the condition arises. Using this mental process has been found to increase the proportion of people who remember to do an academic assignment, avoid unhealthy snacks, speak up against bad behaviour, and perform many other desirable behaviours. The main benefit seems to be in remembering to act. Forming the intention in the correct format is important to its effectiveness.

E.g. Another brief intervention to help people with turning good intentions into good behaviour is 'mental contrasting'. The subject is required to think about how they would ideally like to behave in future, then how they have been behaving in reality or about problems they need to overcome. This usually encourages thinking about the specific actions they can take to reduce the difference. This method is often successful in helping people and takes very little time. Gabriele Oettingen has written a book on the subject, having proposed it in Oettingen (2000). Typical studies are discussed in Oettingen et al (2009).

Carrying out studies without testing ways to achieve better performance can be wasteful, as illustrated earlier by the example of Milgram's experiments on obedience to authority.

4.5 Use of introspection

Cautious use should be made of introspection[4], ideally in conjunction with methods that record detail about behaviour.

> E.g. Newell (1967) gave a person a tricky mathematical problem[5] and recorded what they wrote and said while working on it. The 'think aloud protocol' became a commonly used method, although it can be time consuming to analyse thoroughly.

4.6 Subjects

Subjects for psychological studies are usually selected to be representative of people generally, or representative of people with a particular challenge (e.g. learning difficulties, depression).

(The fact that many of the subjects in psychological studies are undergraduates, often psychology undergraduates, is a matter of convenience.)

However, if the study requires people to think in particular ways and perhaps also talk about their thinking then a different approach becomes attractive.

4.6.1 Expert subjects

Research would be facilitated by developing a set of subjects skilled at:

- learning specified mental processes;
- executing them during studies; and
- speaking their thoughts accurately, promptly, and with minimal interference.

This group might be selected for their initial ability to do these things and given special training, encouragement, and rewards.

The subjects would participate in more than one study. Ideally, they would become regular contributors. People would be motivated to volunteer as subjects by the opportunity to develop useful skills with expert help and the encouragement of consistent monitoring.

These subjects would be easier to work with in studies and, over time, experimenters would learn about the subjects' abilities and mental processes. This would make identifying processes used easier.

Psychologists usually prefer that subjects do not know what the study is for until after they have participated. This reduces the risk of subjects providing the behaviour they think the experimenter is hoping to see. This risk must be reduced in other ways if expert experimental subjects are used and they are using specified mental processes. They are much more likely to know or guess what the study is for.

This will often be simple. Subjects will usually be trying to carry out variations on promising mental processes and the objective is to see how the processes perform. If neither the experimenter nor subject has any preference between variations of the process being tested then everyone is interested in finding out which is best, not in rigging the test. Alternatively, the subject could be kept ignorant of competing processes and encouraged to do the best they can with the process they have learned.

[4] During arguments about the use of 'think aloud' protocols some psychologists distinguished between introspection and verbalizing the contents of working memory while thinking. In this article I just mean anything of this general type but recognize that the true 'think aloud' method gives more reliable results. Details can be found in Ericsson and Simon (1980, 1984, and 1998).

[5] In the sum DONALD + GERALD = ROBERT each letter represents a digit. The task is to work out which digit goes with each letter, given that D=5.

4.6.2 High-performing subjects

If the research strategy is to learn the processes used by top performers then subjects who are top performers are needed. If they are also expert experimental subjects then that is helpful but a different kind of expertise.

4.6.3 Self-experimentation

Experimenters probably should not try to get other people to adopt a mental process that they themselves have not successfully used beforehand.

More generally, self-experimentation can be valuable and worth publishing. Experimenters are well placed to act as subjects in their own studies when it is important to control mental processes and introspection is used. They may also be willing to put up with more discomfort, boredom, and lost time than ordinary subjects.

Although using just one subject usually provides only a little information, self-experimentation should be used occasionally, especially to generate ideas worthy of a larger study.

There is a risk of bias (e.g. seeing what you hoped to see) but this can be managed in various ways. Studies can be replicated by others. Experimenters can aim to test alternative processes against each other to see which works best without having a favourite beforehand. (This makes the study more worthwhile than one that just confirms something you knew already.)

Self-experimentation makes it possible to explore effects with less risk of subjects failing to follow required mental processes.

> E.g. When people are asked to learn verbal material in experiments they often repeat the material silently to themselves (though it can emerge as audible muttering). This is called 'rehearsal'. But suppose we wanted to compare the effect of spending 10 seconds silently rehearsing each item with 10 seconds of mental rest? The usual way to stop people rehearsing is to give them another task that blocks rehearsal, but that would not be mental rest. We need subjects to understand the experiment, know how to rest for a few seconds, and do so during just the relevant trials. Anyone can learn to do this, but the experimenter is likely to be the first person to be capable.
>
> E.g. In the story 'A scandal in Bohemia' Sir Arthur Conan Doyle drew attention to the difference between merely seeing something and observing it. Subjectively the difference is clear but psychologists have not explored it as far as I can tell. What they have done is to establish that verbalizing a fact about something to be remembered improves memory. They have debated whether this is because an additional memory 'trace' is created by doing this but have not tested whether making the observation without speaking also improves memory.
>
> My own self-experimentation with non-fiction text, pictures, and furniture has shown me that deliberately observing specific facts enables me to recall them all (usually) later the same day. If I look at the material for the same time but without deliberately observing specifics, I can still recall some specifics later but far fewer of them. This is potentially important for learning skills but needs to be tested by more people and more rigorously across more materials and circumstances.

4.7 Data summary and analysis

If the mind was a natural system and everyone was at least similar, it would make some sense to average the performances of multiple people to reveal common patterns of behaviour.

However, recognizing the mind as an artificial system where people have alternative methods and change them from time to time it makes more sense to treat each person as individual and changing until proven otherwise.

In psychological research, however, it has been very common to use statistical techniques that pool the behaviour of many people to make one view in which exceptions to an 'average' tendency are obscured. These include:

- Averaging curves.
- Averaging performance scores.
- Factorial experiments where group averages are compared.
- Regressions of various types, including multi-level modelling.

One commonly used form of averaging is to average the performance of many people to create an average curve.

> E.g. On average across many people and many attempts at learning, learning and improvement progress steadily.
>
> However, when looked at individually and in detail the progress is often far from steady (e.g. Donner and Hardy, 2015). Sometimes people plateau before making a breakthrough and racing ahead for a period. Sometimes they get worse before getting better again.
>
> People progress at different rates and with different patterns of fast and slow progress. This is probably partly because of using different methods of learning.
>
> The patterns shown by individuals and the thinking methods that contribute to them are more important and more interesting than the average learning curve, which does not represent the performance of any individual. The averaged curves obscure the impact of breakthroughs in learning.

Siegler (1987) provided a useful analysis of the problems of averaging across differing strategies and a demonstration using children's strategies for adding small numbers. This showed variations between subjects, between superficially similar tasks for the same subject, and even variations within subjects for the same task. His paper also references several other studies showing variations in strategy that would be misrepresented by average performance.

If we must average at all, a better approach is to cluster subjects statistically to see if there are some whose performance is so similar that they seem to be using the same method of thinking. Averaging performance within a cluster might be helpful.

If it is possible to identify the strategy used each time (e.g. from thinking aloud or written workings) then this is another way to group by strategy.

A different form of averaging is to put people into experimental and control groups then look for differences on average between the groups.

> E.g. On the basis of averages across groups, people are irrational in many experimental tasks. However, in each such task it is usually found that some people within the groups perform rationally. The differences between people and the thinking that produce them are more interesting than the average of irrationality.

Another very common form of averaging in psychology involves giving people a long questionnaire then turning their answers into a list of summary scores. These scores are typically considered to be measures of their personality (e.g. 'agreeableness', 'need for achievement', 'conservatism'). The questions are put into groups on the basis that the answers given in a past study were correlated with each other but less so with answers to questions in other groups. The scores for each question are usually averaged across each group.

Using sophisticated statistical methods does not make an approach scientific, and this approach has some serious problems.

The correlations between answers often come from the fact that the same question is asked multiple times using different words.

This also creates the potentially false impression that all relevant traits have been identified. A typical attempt to demonstrate that all relevant traits have been identified relies on thinking up lots of questions, putting them in a study, and using cluster analysis to find the main clusters. The problem is that the clusters you see still depend on the questions you asked. If you ask the same question multiple times using different words you will get a cluster[6].

Another problem is that the questions asked are often poorly defined. The subjects are not being asked for factual answers; they are being given an opportunity to let the experimenter know what kind of people they are.

> E.g. You might be asked to state your level of agreement with the following statement: 'I am easily annoyed'. How easily is 'easily'? How 'annoyed' is 'annoyed'? If a person has unusually annoying co-workers, they might have the impression that they are easily annoyed because they are often annoyed. Would this be fair? What is the difference between 'strongly agree' with this statement and 'agree'[7]?
>
> This is very different from a question designed to elicit facts. An example of a better-defined factual question might be: 'In the past 6 months have you become angry enough to shout loudly at someone or try to hurt them physically?'

Averaging scores across many poorly defined questions does not make the answers more meaningful. It is little comfort that these scores are mostly stable over time (i.e. people give mostly similar answers over time) and correlate with parental characteristics, risk of psychopathologies, and future behaviour (but only weakly). Statistical correlation is not evidence of accurate measurement and only the weakest evidence of validity.

Responding to the artificial nature of mind, we should ask well defined questions about beliefs and methods. This will help to understand why individuals behave as they do and what might cause them to behave differently (e.g. development of a new skill, revision of a false belief).

> E.g. Many people think we each have an individual 'attitude to risk' that plays an important role in our decisions. Supposedly, some people are more 'risk averse' than others and some are even 'risk seekers'. Nobody has yet

[6] You can also ask the opposite question and reverse score it, again boosting the appearance of a cluster.

[7] In the early 1980s when I was a psychology undergraduate, we were shown a test of 'conservatism' that asked subjects if they 'believe in beatniks'. The term 'beatnik' was already obsolete by that time but the phrase 'believe in' was even more difficult. Was there ever doubt as to the existence of people who were beatniks? Was it their approach to life that was being asked about? How did this ever get into the questionnaire?

found a satisfactory way to show that this is the case and there are good reasons for thinking it cannot be done and the theory is probably wrong.

Schoemaker (1993) explained the problem in detail. There are many other factors that drive our behaviour and revealing an inherent risk attitude requires factoring out all the other drivers in some way. For example, a person might do a dangerous sport because they are good at it, do not see it as dangerous, and get paid a lot to do it. Others may see them as risk seekers but the dangerous sport player takes all possible safety precautions, indicating an aversion to risk.

The simplest mistake, made often by researchers, is to ask people questions about how often they do risky things and then calculate a score said to be their risk attitude. It is only a measure of their risk behaviour. In practice people tend to be different in their risk taking between different domains and this often reflects their skills. For example, people who have excellent interpersonal skills are more likely to initiate conversations to resolve sensitive problems.

4.8 Application of findings

When research results are published in an academic journal the job is not done. Until people benefit in some way from the research it has yet to be useful.

One way to drive a programme of research is to find people who want to improve and work with them. This is a good way to get experimental subjects, fund research, and have an immediate useful impact.

E.g. Rackham's research into sales behaviour was the basis of a successful consulting business: Huthwaite International. By 1988, Rackham was able to claim that their advice was based on observation and study of more than 35,000 sales calls by 10,000 people selling high-value products or services in twenty-three countries over 12 years. The content of their books and courses was, and is, heavily based on this research.

5. Research design checklist

If you are involved in generating psychological research ideas then try using this checklist to evaluate possibilities.

1. Artificial or natural?

Is the system to be studied artificial or natural? Could a person learn to act differently if they put a lot of time and effort in and had help? Is there any prospect of this happening?

2. Useful task

In your study, would you ask subjects to do a useful task? That is, something important that some people do often outside the laboratory and might want to do better? The 'task' need not be work. For example, it might be to play a game, stay relaxed, or make a new friend.

It is not enough to have a task that is the same as something real in principle only. The task must be realistic.

3. Individual mental processes

Does your research method recognize that individuals each have their own mental processes for carrying out the task, and that these may change during the study?

You might try to identify their mental processes (e.g. to find the skills used by top performers) or control them (e.g. to record the performance characteristics of a promising approach).

4. Individual, detailed recording

Would you record details of individual performance that provide insights into their thinking, not just their final chosen action? Would this include their thoughts, perhaps verbalized at the time, or using interviews later?

5. Analysis that captures individual detail

Would your method for analysing your data recognize individual differences in mental process and potential changes over time?

6. End users

Would you have people who would be interested in using your discoveries? For example, people wanting to perform the useful task better.

6. Scope for improvement

6.1 An individual choice

My impression over the years and while researching for this paper is that the scope for improvement is massive. Most studies reported over the past several decades and even today use natural-science methods to study artificial mind phenomena. Finding positive examples of efficient research for this paper was difficult because they are so rare. Even studies that demonstrated good research choices were often let down by poor choices on other elements.

> E.g. Chase and Ericsson (1981) was given earlier as a good example of detailed recording of data but the task used was a useless one: digit span. Yoon, Ericsson, and Donatelli (2018) reported a follow up with Dario Donatelli, who, 30 years earlier, had increased his digit span from 8 to 106 through around 800 hours of practice. The purpose of the follow up was to see what had happened to his skill after 30 years of not using it (because it was useless).

Correcting this problem of using natural-science methods for the largely artificial mind should lead to much more efficient psychology, both in improved understanding and improved technology of mind.

However, the scope for improvement seems to vary by specialism. To really understand the overall situation now and perhaps pick up any trends would require a time-consuming study. Fortunately, such a study is unnecessary.

What matters is whether there is scope for improvement in the research ideas of individual psychologists. If you are involved in psychological research then you can look at studies published in your areas of interest, and your own studies, and decide if there is scope for your thinking to change. To help you, this paper provides the checklist (above), some detailed examples (next), and a self-test (below).

6.2 Detailed examples

The following examples look in detail at recently published studies to show some of the problems that need to be addressed. The studies were not chosen because they are particularly poor examples. They are not. They were chosen because they were understandable, published recently in a journal that anyone can read, and detailed data from the study were available easily to download and analyse.

6.2.1 Immediate and delayed rewards

Chen et al (2021) had volunteers answer questions about their preferences in three different hypothetical situations. Half the subjects were made to feel mild pain while the control group subjects were pain free. In all there were three experiments, each using a different kind of pain. The first two used physical pain from a Shiatsu mat and cold water respectively. The third showed a documentary video of a very sad situation for a woman.

In each hypothetical situation, subjects were asked 27 questions (making 81 questions in total per person). All questions asked whether the subject would prefer a smaller immediate reward or a larger delayed reward. The sizes of each reward and the length of the delay were varied.

The hypothetical situations were:

- **Money:** Would you prefer X Yuan now or Y Yuan after N days?
- **Vacation:** Imagine that you have won a prize at work and can choose, would you prefer X days of holiday now or Y days of holiday after a delay of N days?
- **Health:** Imagining that you are currently unwell, would you prefer to return to good health for X days now or Y days after a delay of N days?

The decision-making involved is artificial because a person can learn to tackle these in different ways. In the Money situation there are already well-known ways to calculate the best decision.

The tasks are not directly useful. While weighing immediate and delayed rewards is important, these particular decisions are not ones we need to make in real life. The money situation is closest to real life but even here we are usually faced with the choice of paying for an expensive item immediately or paying later but with a slightly higher price.

Ideal rational decision-making helps understand this task. With the money situation, we can compare the effective annual rate of interest implied by the larger delayed reward to the rate the person is currently paying or receiving. A person who is a net saver earning 2% per annum at best will not be attracted to any of the smaller immediate rewards in this study because the interest rate of the larger delayed reward is almost 6%, even in the worst case. However, a person who is repeatedly taking out payday loans with an effective annual interest rate of over 1,000% (because of fees and interest) would gain from taking the immediate payment and using it to avoid payday loans in about half of the choices offered by the study.

The study did not ask subjects about their personal financial situation, so we do not know what would have been rational for each subject. However, we do know that some combinations of preferences are irrational. It would not be rational to prefer an immediate reward in one question but a delayed reward in another where the implied interest rate is lower.

Only 7.8% of subjects across all experiments and conditions gave completely consistent preferences for all three situations.

This is typical of preference studies like this that I have seen, asking for large numbers of difficult judgements. However, once performance across all subjects is averaged the high level of irrationality is hidden. Collectively, the subjects in this study made choices that were, on average, consistent even though they might not have been ideal.

The study report gives no indication of whether any subjects did explicit calculations, used an electronic gadget to do calculations, or kept notes to avoid giving inconsistent answers[8]. We do not know how long they took to give answers, or whether they could see all questions as they were answering (which might have

[8] Unlikely under these uncomfortable circumstances, though some subjects might still have been able to do calculations.

helped them be more strategic and efficient). Consequently, we know nothing about how subjects tackled this task other than that most did so irrationally.

As a result, we learn nothing about the performance achieved by any specific method and what we do learn mostly concerns incompetent performance in the task. We would hope to educate people to avoid this level of incompetence, so the performance most worth knowing is that of more competent methods.

6.2.2 Anchoring

Bahnik (2021) explored the causes of an effect called 'anchoring'. In the study people were asked to estimate numbers most people do not know, such as the height of the Eiffel tower. However, they were first asked to say if the height was more or less than a number generated at random (and which they knew to be random). The anchoring effect is the well-known tendency for the estimates to be biased towards the number earlier generated at random and used in the comparison. Anchoring (of small average magnitude) was found in this study and traced to the number used for the comparison and not other numbers used more recently in other comparisons.

The behaviour studied is again artificial because different methods of making estimates can be learned. The task is a realistic and useful one because estimating uncertain quantities is useful. The danger of anchoring is real since sometimes (e.g. in a meeting at work) we might be asked for a comparison before being asked for an estimate. When that happens, strong reasoners will want to know how to make their estimates with the least bias.

The statistical method used for this research and published was a complex set of regressions designed to reveal the influence of potentially biasing numbers. However, the raw data were also made available and reveal interesting individual differences and inconsistencies. The data include times taken for the initial comparison and later estimate.

An inconsistent pair of answers occurs where the subject says a quantity is higher than the random comparator, but then gives an estimate that is lower than the comparator (or says 'lower' but then estimates higher). Overall, only 33% of subjects gave completely consistent answers across all 13 quantities they were asked to estimate. Being consistent was associated with spending more time on the first judgement (the comparison).

However, spending more time on the initial comparison was also associated with the comparator being close in value to the subject's later estimate. It could be that people spend more time trying to decide on the comparison when it is a close one, or that when people spend longer on the comparison it anchors their estimate more.

Although recording time taken provides some insight into the reasons for individual differences in correlations between estimates and comparators, it is not enough. A person might deliberately ignore the comparator and try to form an estimate of the uncertain quantity before making the required comparison. Alternatively, they might dwell on the comparator and not try to make an independent estimate first.

A study providing cleaner data and more insight would be one where subjects are trained to (1) use particular strategies (independent estimate first or not) and (2) spend a controlled amount of time on each judgement. From a practical point of view, it would be useful to know if deliberately forming an estimate before making the comparison (even though the comparison number has been mentioned) helps reduce anchoring.

6.2.3 In- and out-group commenting

McCrea, Erion, and Thürmer (2021) had subjects assess, in various ways, comments supposedly made by people from three countries about their experience of teamwork with people from those three countries. The comments to be assessed were a paragraph long and were strongly negative generalizations about people from the countries involved.

In addition to the assessments, subjects were asked to decide how much reward the commenter should get (measured in lottery tickets) for each comment. The rewards were for the 'quality' of the comments and, complicating matters further, the instructions said that the subject would get a higher reward if they gave a higher reward to the commenter. This was to promote 'fairness' according to the instructions, though it was more likely to promote biased assessments.

The statistical analysis published used averages, ANOVA, and regression, which are all techniques that tend to make individual differences less prominent.

On average, subjects gave higher rewards when the commenter was denigrating the people of their own country than when they denigrated the people of another country. The angrier they were about the comments, the lower the rewards they gave. The more they thought the commenter was acting out of good motives, the higher the reward.

On average the results appear to reflect plausible tendencies. However, the raw data were also made available and when individual ratings and rewards are inspected, new questions arise that cast doubt on the meaning of the findings.

Rewards given varied greatly between individuals, even for the same comment text and country pair. In every such case, some subjects gave the maximum reward while others gave the minimum, with every other possibility also being chosen by someone.

One obvious potential explanation for this is differing interpretations of the instructions. Rewards were to be given for the 'quality' of comments. To some people that might have meant the social acceptability of the comments, to others the factual accuracy of the comments, and to others the detail and clarity of the comments. Each basis gives a different reward. The comments were negative over-generalizations about the people of a country, so they were unacceptable on social and factual grounds, deserving minimal reward. However, they were unusually long and revealing for comments to surveys, so a good source of insight and deserving a high reward, perhaps the maximum.

Subjects may have come up with other interpretations, varied between these suggestions, or ignored the instructions and tried to maximize their own rewards by giving maximum rewards to commenters. The most often chosen reward was the mid-point between the maximum and minimum rewards. The next most common choice was to give maximum rewards to all comments (which 48 subjects out of the 1,014 did).

Consequently, it is a practical impossibility to gain any insights from the reward data. The study needed to give clearer instructions and check they were followed.

The experimental task in this study was tantalizingly close to being useful. It could have been used to find out how to teach people to identify unfair over-generalizations about demographic groups. These are a frequent component of racism and sexism, for example. Alternatively, it could have explored the skill of giving ratings of text that are objective in a situation where the initial instructions are inadequate. This is often needed by auditors, doctors, and clinical psychologists.

Instead, the study looked at ratings without giving adequate instructions or giving subjects the opportunity to seek clarification. The objective of the study was to explore why a commonly observed bias occurs (on average). In this the study was unconvincing due to the interference generated by ambiguous instructions and failure to check that subjects had even understood the subtle difference in reward information.

Although the study involved some direct questions about social norms for commenting on national characteristics, it did not ask direct questions about the motives for punishing criticism. A missed opportunity.

This study was not selected as an example because it is a poor study. It was just the first social psychology study I found in a reputable journal that I could understand and for which detailed data were available to download. In most respects it is technically an excellent study reflecting the scholarship and advanced scientific skills of the experimenters.

What has undermined it, as with the other two detailed examples given above, is the effect of the inappropriate experimenting methods discussed earlier in this paper. This starts with the choice of a useless task and continues to the inattention to individual differences in data analysis.

6.3 Self-test

Here are some brief descriptions of research, usually based on published studies. Is the method appropriate for what is being studied?

Suggested answers are in Appendix A.

Study 1: Used data on twins raised together and apart to look at the extent to which autism is hereditary.

Study 2: Explored whether people rate faces as more 'trustworthy' when they appear with a group of other faces or alone. Subjects were shown photographs of faces with and without surrounding faces and asked to rate trustworthiness on a scale of 0 to 10. No guidance was given on how to think when doing this task.

Study 3: Elderly subjects were trained to learn foreign language vocabulary more efficiently by thinking about how easily they recalled translations correctly and spending more time on items they found harder. This was found to bring their performance closer to that of younger subjects, who more often spent longer on harder words without training.

Study 4: The aim was to create a questionnaire that assessed coping strategies, giving overall scores on 6 dimensions. Questions asked how often subjects used each of 30 coping strategies (e.g. trying to solve the main problem, drinking alcohol). There were 5 questions for each of the 6 dimensions. All the usual psychometric methods and criteria were used to try to create dimensions in which answers to the questions correlated with each other and people gave the same answers each time they completed the questionnaire.

Study 5: People were shown pictures of themselves and others that had been digitally tweaked to make them look fatter or thinner. They were then asked to adjust an image to show the correct body size. The expectation was that the attempt to show correct body size would be biased by the previous pictures.

Study 6: Adolescents with learning difficulties were individually taught a strategy to help them identify words (especially longer, less familiar words).

Their ability to identify words and understand sentences and passages was tested before and after training, and there was a control group that did the same tests but received no training.

Study 7: Subjects were shown words with two meanings, one of which was related to movement, and asked to say what the words meant. In some trials subjects were asked to move their hands while doing the task to see if this would influence their choice of meaning.

Study 8: Subjects were required to study a difficult text giving definitions in discrete mathematics but to take frequent rests according to a repeating schedule for one hour. The schedules used were (a) work 5 minutes and rest for 1, (b) work 10 minutes and rest for 2, and (c) work for 16 minutes and rest for 4. Their feelings of strain and focus were recorded every few minutes along with their scores in a test at the end of the study session and a week later.

Study 9: A study tried to reduce the pace of cognitive decline in dementia by providing the patients with extra stimulation in the form of social events and group activities. A control group was used.

Study 10: To understand how people select mates, the results of a simulation of mate selection were compared to real life couples formed before the study began.

7. Conclusions

Human thinking is largely artificial, in the sense that we can redesign and change it. Tackling psychology as if the mind is just a natural system to be observed, understood, and predicted is a mistake and a route to inefficient research.

Instead, there are several ways to research the mind as an artificial system that are efficient in producing understanding and useful results.

8. References

Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian Journal of Educational Research*, 44(3), 229-251.

Bahník, Š. (2021). Anchoring without scale distortion. *Judgment and Decision Making*, 16(1), 131. Available (with downloadable data) at: http://journal.sjdm.org/vol16.1.html

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision*, 9(3), 7-7. Available at: https://jov.arvojournals.org/article.aspx?articleid=2193543

Bower, G. H. (1970). Analysis of a mnemonic device: Modern psychology uncovers the powerful components of an ancient system for improving memory. *American Scientist*, 58(5), 496-510.

Bower, G. H., & Reitman, J. S. (1972). Mnemonic elaboration in multilist learning. *Journal of verbal learning and verbal behavior*, 11(4), 478-485.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of educational research*, 80(2), 144-179.

Boyce, C. J., Wood, A. M., & Powdthavee, N. (2013). Is personality fixed? Personality changes as much as 'variable' economic factors and more strongly predicts changes to life satisfaction. *Social indicators research*, 111(1), 287-305.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. *Cognitive skills and their acquisition*, 141-189.

Chen, J., He, Z., Sun, Y., Xia, W., Liao, Z., & Yuan, J. (2021). Does pain make people short-sighted? The impacts of physical and psychological pains on intertemporal choice. *Judgment & Decision Making*, 16(4). Available (with downloadable data) at: http://journal.sjdm.org/vol16.4.html

Cho, B. Y. (2014). Competent adolescent readers' use of Internet reading strategies: A think-aloud study. *Cognition and Instruction*, 32(3), 253-289.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, 104(3), 268.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist*, 34(7), 571.

Delquié, P. (2008). Valuing information and options: an experimental study. *Journal of Behavioral Decision Making*, 21(1), 91-109.

Donner, Y., & Hardy, J. L. (2015). Piecewise power laws in individual learning curves. *Psychonomic Bulletin & Review*, 22(5), 1308-1319. Available at: https://link.springer.com/article/10.3758/s13423-015-0811-x

Durbach, I. N., Algorta, S., Kantu, D. K., Katsikopoulos, K. V., & Şimşek, Ö. (2020). Fast and frugal heuristics for portfolio decisions with positive project interactions. *Decision Support Systems*, 138, 113399.

Ebbinghaus, H. (1880). Urmanuskript "Ueber das Gedächtniß". Passau: Passavia Universitätsverlag.

Ebbinghaus, H. (1885). Über das Gedächtnis. Leipzig: Dunker.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. the MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3), 178-186.

Farb, N., Anderson, A., Ravindran, A., Hawley, L., Irving, J., Mancuso, E., Gulamani, T., Williams, G., Ferguson, A. & Segal, Z. V. (2018). Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. *Journal of consulting and clinical psychology*, 86(2), 200.

Hornsey, M. J., Edwards, M., Lobera, J., Díaz-Catalán, C., & Barlow, F. K. (2021). Resolving the small-pockets problem helps clarify the role of education and political ideology in shaping vaccine scepticism. *British Journal of Psychology,* 112(4).

Karvetski, C. W., Meinel, C., Maxwell, D. T., Lu, Y., Mellers, B. A., & Tetlock, P. E. (2021). What do forecasting rationales reveal about thinking patterns of top geopolitical forecasters? *International Journal of Forecasting*.

Kepner, C. H., & Tregoe, B. B. (1965). *The Rational Manager: A Systemic Approach to Problem Solving and Decision Making*. McGraw-Hill, New York.

McDaniel, M. A., Brown, P. C., & Roediger III, H. L. (2014). *Make It Stick: The Science of Successful Learning.* Cambridge, MA Harvard Univ Pr.

McCrea, S. M., Erion, C. J., & Thürmer, J. L. (2021). Why punish critical outgroup commenters? Social identity, general norms, and retribution. *British Journal of Social Psychology*.

Milgram, S. (1974), *Obedience to Authority: An Experimental View*, London: Tavistock Publications.

Morton, C., Anable, J., & Nelson, J. D. (2017). Consumer structure in the emerging market for electric vehicles: Identifying market segments using cluster analysis. *International Journal of Sustainable Transportation*, 11(6), 443-459. Available at: https://eprints.whiterose.ac.uk/109577/1/Morton%20et%20al%20-%202016%20%5Bintegrated%20paper%5D.pdf

Newell, A. (1967). Studies in problem solving; subject 3 on the crypt-arithmetic task Donald+ Gerald= Robert. Available at: https://kilthub.cmu.edu/articles/Studies_in_problem_solving_subject_3_on_the_crypt-arithmetic_task_Donald_Gerald_Robert/6609998/files/12102137.pdf

Oettingen, G. (2000). Expectancy effects on behavior depend on self-regulatory thought. *Social Cognition*, 18(2), 101-129.

Oettingen, G., Mayer, D., Timur Sevincer, A., Stephens, E. J., Pak, H. J., & Hagenah, M. (2009). Mental contrasting and goal commitment: The mediating role of energization. *Personality and Social Psychology Bulletin*, 35(5), 608-622.

Pande, S. M., Papamichail, K. N., & Kawalek, P. (2021). Compatibility effects in the prescriptive application of psychological heuristics: Inhibition, Integration and Selection. *European Journal of Operational Research*, 295(3), 982-995.

Perry, G. (2013a). Deception and illusion in Milgram's accounts of the obedience experiments. *Theoretical & Applied Ethics*, 2(2), 79-92.

Perry, G. (2013b). Behind the shock machine: The untold story of the notorious Milgram psychology experiments. New Press, The.

Rackham, N., & Carlisle, J. (1978a). The effective negotiator—Part I: The behaviour of successful negotiators. *Journal of European Industrial Training*, 2(6), 6-11.

Rackham, N., & Carlisle, J. (1978b). The effective negotiator—Part 2: planning for negotiations. *Journal of European Industrial Training*, 2(7), 2-5.

Rackham, N. (1988). *Account Strategy for Major Sales*. Aldershot: Gower.

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and decision making*, 3(8), 641-658. Available at: https://serval.unil.ch/resource/serval:BIB_209010488EFA.P001/REF.pdf

Richards, T. L., & Berninger, V. W. (2008). Abnormal fMRI connectivity in children with dyslexia during a phoneme task: Before but not after treatment. *Journal of neurolinguistics*, 21(4), 294-304.

Richarme, M. (2005). Consumer decision-making models, strategies, and theories, oh my! *Decision Analyst*, 604, 76011-3100.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological review*, 102(2), 305.

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, 1(3), 181-210. Available at: https://linguistics.ucla.edu/people/hayes/T eaching/papers/2006_Roediger_Karpicke_Review.pdf

Rosengrant, D., Van Heuvelen, A., & Etkina, E. (2009). Do students use and understand free-body diagrams? *Physical Review Special Topics-Physics Education Research*, 5(1), 010108.

Schoemaker, P. J. (1993). Determinants of risk-taking: Behavioral and economic views. *Journal of Risk and Uncertainty*, 6(1), 49-73.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, 116(3), 250.

Simon, H. A. (1980). Cognitive science: The newest science of the artificial. *Cognitive science*, 4(1), 33-46.

Simon, H. A. (1996). *The sciences of the artificial (3rd edition)*. MIT press.

Spivey, J. M., & Abrial, J. R. (1992). *The Z notation* (Vol. 29). Hemel Hempstead: Prentice Hall. Available at: https://spivey.oriel.ox.ac.uk/wiki/files/zrm/zrm.pdf

Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S., & Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: empirical evidence. *Journal of consulting and clinical psychology*, 70(2), 275.

Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive psychology*, 12(1), 137-175. Available at: https://static.trogu.com/documents/articles/palgrave/references/thorndyke_stasz_maps_1980.pdf

Yoon, J. S., Anders Ericsson, K., & Donatelli, D. (2018). Effects of 30 years of disuse on exceptional memory performance. *Cognitive Science*, 42, 884-903. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12562

9. Appendix A: Self-test suggested answers

Study 1: Since autism is very likely a condition with an organic cause, this is a natural system and the approach is reasonable.

Study 2: Is it even a good idea to try to judge a person's trustworthiness from the appearance of their face? This is not just a useless task; it may be worse than useless. Better to look more generally at how people should think about the trustworthiness of other people.

Study 3: A useful task that many people want to do better, and a suitable method. The subjects were trained to act more effectively and the effects were recorded. This could be done down to the level of individual subjects, looking at the improvements achieved by each person in their study methods and results.

Study 4: Not an appropriate method. This is treating coping skills as a fixed characteristic of the person. Helping people learn to cope more effectively by using coping skills better would be a much better study topic. That would call for a questionnaire that carefully checked understanding of which coping skills are useful in different situations. For example, ignoring a problem is often a bad approach but can be useful to achieve calm before then trying to solve it.

Study 5: A useless task. Today if they want to show what a person looks like, most people get their phone out and take a picture.

Study 6: A worthwhile task and an opportunity to develop something helpful through this appropriate study. The more detail that can be recorded and analysed about individual progress the better. Future studies might test revisions of the method for identifying words and perhaps integrate it better with methods for understanding sentences and passages.

Study 7: A useless task. Both meanings of the word should be in play until relevant information resolves the ambiguity. To ask people to make a choice without that information is pointless, like this study.

Study 8: A useful task for mathematics students and alternative work-rest schedules are a popular topic in study advice. This is a broadly appropriate method. It would probably over-complicate the study if the subjects also had to study using particular mental processes, but a later experiment might check this because better study processes can be less tiring.

Study 9: This was looking at a problem that most likely has an organic cause, so a natural science approach was reasonable. The subjects (patients) were given a different environment but not asked to think differently. It might be worth asking them to think in particular ways, but this gets harder as dementia progresses.

Study 10: What could be more natural than choosing someone to love? And yet this is artificial behaviour. Many people would benefit from choosing more wisely. Treating this as a natural system and just trying to describe and predict it is missing important opportunities. The larger skill is to attract, select, and shape a mate.